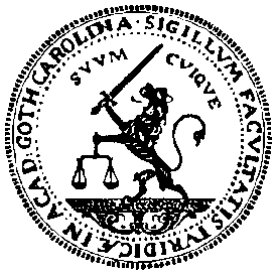

FACULTY OF LAW

LUND UNIVERSITY

Suana Tafić

Cracking the AI-gma Code?

The Interpretability and Explainability of AI in light of Forst's Right to Justification

JURM02 Graduate thesis

Graduate thesis, Master of Laws program

30 higher education credits

Supervisor: Eduardo Gill-Pedro

Semester: VT 2023

Table of contents

Summary

In April 2018 the European Commission took their first step into establishing the European AI Strategy, which aims to make Europe the world-leading centre for AI while ensuring fundamental rights and freedoms. In April 2021 the proposal for a regulation on artificial intelligence, referred to as the AI Act, was announced. The proposed Regulation, which has yet to gain status as secondary law, uses a risk-based approach and particularly emphasises the importance of transparency and trustworthiness in the context of AI. Furthermore, the proposal addresses how opacity can unfavourably affect various fundamental rights that are enshrined in the Charter of Fundamental Rights of the European Union.

The efficiency of the current EU framework has been questioned from various stakeholders who dispute whether the framework can address and safeguard the AI-induced fundamental rights in an adequate manner. The proposal sets forth that AI systems should be developed in a way which allows humans to understand their actions. Therefore, the thesis intends to investigate whether the measures outlined in the proposed AI Act, which aim to safeguard transparency through interpretability and explainability, adequately uphold the right to justification. Moreover, the thesis aims to clarify what the right to justification entails, how interpretability and explainability has been defined in EU law and what obligations the proposed Regulation imposes in respect to interpretability and explainability of AI systems.

The thesis establishes that human rights require us to have our moral right to justification respected. The principle of justification enables a power to demand justification and challenge false legitimations. If interpretability and explainability are not ensured, the right to demand justification and challenge decisions that AI systems make is violated. Thus, the thesis argues that the measures in the proposal, which aim to safeguard transparency through interpretability and explainability, do not sufficiently uphold the right to justification. The EU is showing a clear commitment to fundamental rights by incorporating these measures in the proposal, however this initiative does not mean that the measures are adequately defined and clear.

Sammanfattning

I april 2018 tog EU-kommissionen sitt första steg mot att upprätta den europeiska strategin för artificiell intelligens, vars syfte är att göra Europa till det världsledande centret för artificiell intelligens samtidigt som grundläggande rättigheter och friheter garanteras. I april 2021 presenterade Kommissionen sitt förslag till en förordning om harmoniserade regler för artificiell intelligens, den så kallade AI-förordningen. Förordningen, som ännu inte utgör sekundärrätt, använder en riskbaserad strategi och betonar särskilt vikten av transparens och tillit för AI-system. Vidare betonar förslaget hur opacitet kan ha en negativ inverkan på rättigheterna i Europeiska unionens stadga om de grundläggande rättigheterna.

Efterlevnaden av EU:s nuvarande regelverk har betvivlats av flera akademiker och forskare som ifrågasatt huruvida ramverket kan hantera och skydda de grundläggande rättigheter som påverkas av artificiell intelligens på ett adekvat sätt. AI-förordningen anger att AI-system bör utvecklas på ett sådant sätt som gör det möjligt för människor att förstå deras beslut. Därmed är syftet med uppsatsen att undersöka om de föreskrivna bestämmelserna i AI-förordningen, som syftar till att garantera transparens genom tolkningsbarhet och förklarbarhet, upprätthåller rätten till rättfärdigande på ett adekvat sätt. Uppsatsen syftar dessutom till att klargöra vad rätten till rättfärdigande innebär, hur tolkningsbarhet och förklarbarhet har definierats i EU-rätten och vilka skyldigheter som den föreslagna förordningen ålägger AI-system vad gäller tolkningsbarhet och förklarbarhet.

I uppsatsen fastställs det att mänskliga rättigheter kräver att vår moraliska rätt till rättfärdigande blir respekterad. Om tolkningsbarhet och förklarbarhet inte garanteras, kränks rätten att kräva rättfärdigande och ifrågasätta de beslut som AI-system tar. I uppsatsen konstateras därför att åtgärderna i förslaget, som syftar till att garantera transparens genom tolkningsbarhet och förklarbarhet, inte upprätthåller rätten till rättfärdigande i tillräcklig utsträckning. EU visar ett tydligt engagemang för de grundläggande rättigheterna genom att införliva

dessa åtgärder i förslaget. Detta innebär emellertid inte att bestämmelserna är tillräckligt definierade och tydliga.

Förord

Hur ska jag kunna beskriva mina fem år i Lund? Sanningen är att ord aldrig kommer att räcka till för att kunna beskriva mitt liv i denna fantastiska stad. När jag kom in i Pufendorfsalen hösten 2018 hade jag aldrig kunnat gissa vad de följande fem åren skulle innebära. Det är med stort vemod men också med stor tacksamhet som jag ser tillbaka på mitt liv här i Lund. Jag skulle därför vilja rikta några särskilda tack.

Först och främst vill jag tacka min handledare Eduardo, som från början varit ett fantastiskt bollplank. Utan dig hade jag svävat iväg i EU-rättens djungel.

Jag vill tillägna ett särskilt stort tack till min fantastiska familj. Mamma, pappa och Aldin. Tack för att ni har orkat med mitt eviga pluggande och för er obegränsade kärlek. Ett särskilt tack till dig mamma, för att du alltid lyft luren så fort jag behövt dig. Mamma och pappa, tack för era uppoffringar och för att ni ständigt påmint mig om att ta vara på de fantastiska möjligheter som finns i vårt Sverige. Volim vas najviše! Arnela, baka i dido, hvala i vama za podršku, volim vas!

Ett väldigt emotionellt tack till mitt Lundagäng. Jag är så obeskrivligt tacksam för varenda en av er. Vi har blomstrat i varandras sällskap och ni har lärt mig så, så mycket. Skratt, tårar, resor, tenta- och uppsatsplugg på studiecentrum, hjärtekross, förälskelser, diskussioner och allt däremellan har präglat våra år tillsammans. Jag är så ledsen över att vi inte kommer kunna dela vardagen på samma sätt som vi fått göra under alla dessa år, men jag är också så tacksam över att veta att detta inte är slutet utan bara början. Ni är the real deal mina gäris!

Jag vill även tacka de personer som började som kollegor på Nätis och som till slut blev nära vänner. Att komma in till kontoret och få dela både bra och dåliga stunder med er har skänkt mig en enorm glädje och trygghet. Ni är fantastiska och jag kommer hålla hårt i er.

Ett stort tack till Hana, Mirnes och Meliha för de fantastiska vänner ni är. Hana och Mirnes, utan er hade jag aldrig vågat söka mig till vare sig Lund eller juristprogrammet. Min Hana, ett särskilt stort tack till dig för att du har trott på mig sedan vårt allra första möte. Ingen av oss hade kunnat ana att ett extrajobb skulle resultera i ett så starkt systraskap.

Ett kärleksfullt tack till Jorge. Du är mitt lugn i livets utmaningar. Gracias por amarme de la manera en que lo haces.

Slutligen, tack Lund för allt. Och tack till mig själv för att jag aldrig gav upp, trots allt som livet slängt på mig!

Lund, 16 juni 2023

Suana Tafić

Abbreviations

AI	Artificial Intelligence
AI Act	Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts
ALTAI	The Assessment list for Trustworthy AI
CFR	The Charter of Fundamental Rights of the European Union
CJEU	The Court of Justice of the European Union
Commission	The European Commission
ECHR	The Convention for the Protection of Human Rights and Fundamental Freedoms
EDPB	The European Data Protection Board
EDPS	The European Data Protection Supervisor
EPRS	The European Parliamentary Research Service
EU	The European Union
FAccT	ACM Conference on Fairness, Accountability, and Transparency
FATE	Fairness, Accountability, Transparency and Ethics in AI
GDPR	The General Data Protection Regulation
Guidelines	Ethics Guidelines for Trustworthy AI
HLEG	High-Level Expert Group on Artificial Intelligence
OECD	The Organisation for Economic Cooperation and Development
Parliament	The European Parliament
TEU	The Treaty on European Union
TFEU	The Treaty of the Functioning of the European Union
UDHR	The Universal Declaration of Human Rights
UNGP	United Nations Guiding Principles on Business and Human Rights
XAI	Explainable AI

1 Introduction

1.1 Background

’Like the steam engine or electricity in the past, AI is transforming our world, our society and our industry.’[1]

Artificial Intelligence (AI) is rapidly becoming a pervasive aspect of the present and will be the technological leader of the future.[2] It will not only make our lives easier, for instance by improving healthcare and predicting environmental and climate change,[3] but it will also entail several potential risks, such as discrimination, opaque decision-making and intrusion in our private lives.[4] The advances in AI have resulted in increasing challenges within various areas, a process which could have serious implications for the citizens and organisations of the European Union (EU).[5]

On the 25th of April 2018 the European Commission (the Commission) published their Communication “Artificial Intelligence for Europe”, taking their first step into establishing the European AI Strategy. The strategy aims to make Europe the world-leading centre for AI while ensuring fundamental rights and freedoms.[6] In February 2020 the Commission published a White Paper on AI and in April 2021 the proposal for a regulation on AI, referred to as the AI Act, was announced.[7] In December 2022 the AI Act progressed

[1] European Commission, ‘Communication from the Commission - Artificial Intelligence for Europe’ COM(2018) 237 final, 25 April 2018, 1.

[2] Stuart Russell, *Human Compatible: Artificial Intelligence and the problem of control* (Viking, 2019) preface.

[3] European Commission, ‘Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions Empty – Building Trust in Human-Centric Artificial Intelligence’, COM(2019) 168 final, 8 April 2019, 1.

[4] European Commission, ‘White Paper on Artificial Intelligence – A European Approach to excellence and trust’, COM(2020) 65 final, 19 February 2020, 1.

[5] Francesco Molinari and others, ‘AI Watch. Beyond pilots: sustainable implementation of AI in public services’ (Publications Office of the European Union 2021) JRC 126665, EUR 30868 EN, 1 <https://www.standict.eu/node/5035> accessed 14 april 2023.

[6] COM(2018) 237 final.

[7] European Commission, ‘A European approach to Artificial Intelligence’ (2023), <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> accessed 14 February 2023.

towards becoming law when the European Council (the Council) adopted its general approach on the proposal.[8] What follows now is interinstitutional negotiations after the European Parliament (the Parliament) has finalised its common position on the matter.[9]

When algorithms are used to make impactful decisions, it is only natural that humans have a desire to understand how the decisions made by AI have come about.[10] The Commission has continuously emphasised the necessity of trust and accountability around the development and use of AI and underlined that the citizens and businesses of EU must have confidence in the technology they interact with.[11] Moreover, scholars within the field have expressed an explicit need of a human rights perspective in the context of trustworthy AI.[12] Human rights extend not only to states but also to organisations and companies and those employed by them, such as those employed within the technology sector.[13] By employing a human rights framework, it is possible to clarify who has responsibilities to do what in certain situations.[14] This thesis therefore draws upon the concept of the right to justification, as proposed by the German philosopher Rainer Forst. What does the right to justification de facto mean and how can it be used to analyse the specific characteristics of AI, such as opacity and complexity? The AI Act seeks to ensure a high level of protection for the fundamental rights enshrined in the Charter of Fundamental Rights of the European Union (CFR) and aims

[8] European Council, 'Artificial Intelligence Act: Council Calls for Promoting Safe AI That Respects Fundamental Rights' (2022) <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/> accessed 10 May 2023.

[9] European Parliament, 'AI Act: A Step Closer to the First Rules on Artificial Intelligence' (2023) <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> accessed 3 May 2023.

[10] Mario Günther and Atoosa Kasirzadeh, 'Algorithmic and human decision making: for a double standard of transparency' (2021) 37 *AI & Soc*, 1 <https://link.springer.com/article/10.1007/s00146-021-01200-5> accessed 29 March 2023.

[11] COM(2018) 237 final, 14.

[12] Jessica Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI' (2020) Research Publication No. 2020-1 Berkman Klein Center for Internet & Society, 8-9 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482> accessed 15 March 2023.

[13] Vinodkumar Prabhakaran and others, 'A Human Rights-Based Approach to Responsible AI' (2022) 5 <https://arxiv.org/abs/2210.02667> accessed 15 March 2023.

[14] ibid.

to, by using a clearly defined risk-based approach, address numerous sources of risks.[15] In order to strengthen the trust for AI and safeguard fundamental rights, AI systems must be developed in a way which allows humans to understand their actions.[16] Existing EU legislation, such as the General Data Protection Regulation (GDPR), has been interpreted as including a right to explainability, a concept the AI Act alludes to as well.[17]

The EU's vision to ensure and scale trustworthy and transparent AI does however not come without challenges.[18] The proposed AI Act has since its announcement been heavily discussed not only within the European law community but also globally. The efficiency of the current EU framework has been questioned from various stakeholders who dispute whether the framework can address and safeguard the AI-induced fundamental rights in an adequate manner.[19] Thus, the question remains, do the measures outlined in the AI Act, which aim to safeguard transparency through interpretability and explainability, adequately uphold the right to justification?

1.2 Purpose and research question

The primary purpose of this thesis is to examine whether the proposed AI Act sufficiently defines and ensures trustworthy AI, with particular focus on interpretability and explainability. The thesis aims to analyse the proposed Regulation of these rights from a human rights perspective, specifically *the right to justification*. Furthermore, the aim is to highlight the challenges of ensuring transparency, through interpretability and explainability, in the proposal in relation to the right to justification.

---

[15] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' COM(2021) 206 final, 21 April 2021, 11.

[16] Ibid.

[17] Sebastian Bordt and others, 'Post-Hoc Explanations Fail to Achieve Their Purpose in Adversarial Contexts' (2022) 2022 ACM Conference on Fairness, Accountability, and Transparency, 1 <https://arxiv.org/pdf/2201.10295.pdf> accessed 1 March 2023.

[18] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (2019) 4 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 1 February 2020.

[19] European Digital Rights and others, 'An EU Artificial Intelligence Act for Fundamental Rights – A Civil Society Statement' (2021) <https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf> accessed 20 February 2023.

The thesis answers the following research question:

Do the measures outlined in the AI Act, which aim to safeguard transparency through interpretability and explainability, adequately uphold the right to justification?

To achieve this, the thesis answers the following sub-questions:

1. What is the right to justification?
2. How has interpretability and explainability been defined in EU law?
3. What kind of obligations does the proposed AI Act impose in respect to interpretability and explainability of AI systems?

1.3 Methodology, materials and perspective

The interpretation of EU law is crucial to the thesis in the light of its aim and purpose. Taking this into account, an EU law approach is mainly applied.[20] The EU method entails that EU law provisions must be taken in the light of their purpose and therefore not be interpreted independently from their context.[21] The Court of Justice of the European Union (CJEU) does not interpret provisions solely on the basis of its wording but also considers the context and the purpose of the provision.[22] The case law of the CJEU is of great importance for the interpretation of EU law and should be seen as a complement to the legislation where there are ambiguities or gaps.[23]

Furthermore, as the aim of the paper is partly to describe and outline existing law the thesis also applies legal dogmatics. Besides outlining how trustworthy AI, specifically transparency, interpretability and explainability, has been defined and interpreted within the EU, i.e., argumenting de lege lata, the thesis will include a critical assessment of how interpretability and explainability could be interpreted and further developed in the proposal, i.e.,

[20] Jörgen Hettne & Ida Otken Eriksson, *EU-rättslig metod: teori och genomslag i svensk rättstillämpning* (2nd edn, Norstedts Juridik 2011) 36.
[21] ibid.
[22] ibid.
[23] ibid 59.

argumenting de lege feranda.[24] Thus, this chosen method can be characterised as critical doctrinal.[25] By applying a critical perspective, it is possible to identify potential deficiencies of the proposal and what consequences this could possibly have for the safeguarding of interpretability and explainability.

To establish the applicable law and the definition of interpretability and explainability as well as relate to the chosen methods, the legal sources used in the thesis is furthermost European Union law and international treaties and regulations. EU law is organised hierarchically with two different types of law; primary law and secondary law.[26] Primary law takes precedence over secondary law and consists of treaties such as the Treaty of the Functioning of the European Union (TFEU), the CFR and the Treaty on European Union (TEU). General principles of EU law are as well part of primary legislation.[27] Secondary law is constituted by various types of law which can either be binding or non-binding (the latter referred to as soft law). Regulations, directives, and decisions are binding whereas recommendations, opinions and White Papers are non-binding legislation.[28]

Due to the wide scope of the judicial area, several important sources have been selected and analysed. Most of the material is acquired from all categories within the EU's hierarchy of norms, however, the starting point is secondary law, the proposal as well as relevant opinions and drafts leading up to the proposed Regulation. Other essential sources used are guidelines from the Commission such as 'Ethics Guidelines for Trustworthy AI'[29] as well as 'Communication on Artificial Intelligence for Europe'[30] and 'White Paper on Artificial intelligence: a European approach to excellence and trust'.[31]

[24] Jan Kleineman, 'Rättsdogmatisk metod' in Maria Nääv and Mauro Zamboni (eds.) *Juridisk metodlära* (2nd edn, Studentlitteratur 2018) 36–38.

[25] ibid 40.

[26] European Commission, 'Types of EU law' <https://commission.europa.eu/law/law-making-process/types-eu-law_en> accessed 27 January 2023.

[27] EUR-Lex,'EU hierarchy of norms'<https://eur-lex.europa.eu/EN/legal-content/glossary/european-union-eu-hierarchy-of-norms.html> accessed 27 January 2023.

[28] Hettne and Otken Eriksson (n 20) 47.

[29] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 4.

[30] COM(2018) 237 final.

[31] COM(2020) 65 final.

Secondary law and general principles of the EU form the foundation of the thesis and have therefore been studied to enable a greater understanding and analysis of the current framework and legal situation. Additionally, the thesis has also been based on legal literature and articles concerning transparency and trustworthy AI. The human rights perspective used in the thesis is based on the article 'The Justification of Human Rights and the Basic Right to Justification: A Reflexive Approach' and on the book *Justice, Democracy and The Right to Justification* where Professor Rainer Forst discusses his constructivist theory of justice as the right to justification.

Furthermore, considering the subject of the thesis is of largely technical character the thesis intends to explain technical standards and terminology, which will complement the legal sources and facilitate a greater understanding of the complexity of AI. As the area of study is of contemporary character there is a lack of relevant case law which could give indications or directions on the interpretation of the proposed law. Thus, an inevitable limitation of the thesis is that the AI Act is only a proposal and has yet to be adopted. This will naturally create hypothetical arguments in several cases.

1.4 Outline

The thesis is divided into five chapters. Chapter two begins with presenting human rights and the theory of the right to justification according to German philosopher Rainer Forst. Moreover, the chapter then continues with presenting the Union's fundamental rights framework. The third chapter presents the proposed AI Act and describes trustworthy AI and its connection to fundamental rights. The underlying interests of regulating AI within the EU are described as well as the definition of AI. Chapter four presents the provisions concerning transparency, interpretability and explainability in the proposal as well as discusses the meaning of the concepts. The last chapter answers the research question and the sub-questions by discussing whether the measures, which aim to safeguard interpretability and explainability, adequately respect the right to justification. Furthermore, the concluding chapter summarizes the main points of the thesis.

1.5 Delimitations

Seeing as the aim of the thesis is to assess how interpretability and explainability are ensured in the AI Act it is not of relevance to analyse provisions which do not regulate this specific aspect of transparency of the proposed legislation. Consequently, seeing as the thesis will not investigate other provisions in the AI Act several questions will be excluded from the scope of the thesis. The discussion of fundamental rights in this thesis will be presented in a general manner, wherein certain fundamental rights may receive more emphasis than others. Moreover, the thesis does not focus on a specific type of AI, such as machine learning, but instead focuses on AI systems in a more general way. Since the thesis is limited to EU law, national law will not be analysed.

The concept of AI liability, as enshrined in the AI Liability Directive, could potentially play an essential role in determining the legal responsibilities in situations where individuals contest decisions that they consider to be unexplainable or uninterpretable. However, it is important to note that conducting further examination of this act's applicability in effectively addressing these potential challenges falls outside the scope of this thesis and thus rests with the legislature.

Furthermore, the thesis will not focus on the GDPR, the Data Protection Directive or other legislation concerned with AI. Some of this legislation will however be touched upon briefly. With the proposal being a moving target and still being debated in Parliament, changes to the Regulation will inevitably occur over the duration of the writing process. The thesis will therefore adapt, among other things, the definition of AI accordingly.

2 Human rights and the Right to Justification

2.1 About the chapter

The following chapter will present the meaning of human rights and the right to justification from the perspective of German philosopher Rainer Forst as well as show the link between human rights and the field of AI. Moreover, it will also demonstrate how fundamental rights are regulated and safeguarded within the EU, with special focus on human dignity and democracy. The purpose of the chapter is to explain the normative framework on which the thesis will be based on. Thus, this chapter will form the basis for the perspective used throughout the whole thesis. The chapter will through Forst's perspective and theory provide a normative framework for the practice of human rights law in Europe, which will guide the normative evaluation of the obligations of interpretability and explainability in the following chapters.

2.2 What are human rights?

Human rights is a complex phenomenon, consisting of several different aspects.[32] Forst's perspective on human rights is based on the notion that human rights have three different lives, the first being a moral one which expresses crucial human claims and concerns that must not be ignored or violated anywhere in the world. Human rights also have a legal life, as they are enshrined in lists of basic rights and national constitutions as well as in treaties, international declarations, and covenants.[33] Besson shares this notion, establishing that the law does not create universal moral rights, but instead turns pre-existing universal moral rights into human rights, de facto making them human rights.[34] Thirdly, as human rights express standards of

[32] Rainer Forst, 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (2010) 120(4) *Ethics*, 711.

[33] ibid.

[34] Samantha Besson, 'Human rights and democracy in a global context: Decoupling and recoupling' (2011) 4(1) *Ethics & Global Politics*, 26.

basic political legitimacy, they also have a political life.[35] Apart from these fundamental aspects, human rights also have a historical existence, a matter which is highly disputed in regards to when the idea of human rights was materialized for the first time and what the notion de facto means.[36] Considering the different aspects of human rights, it is clear that human rights are a recurrent topic in the political sphere, both nationally and transnationally. The topic gives rise to several questions regarding the fulfilment and violation of human rights and how the latter can be avoided or sanctioned.[37]

The aspects mentioned above are all fundamental and must be integrated in the correct way for a complete philosophical account of human rights. When doing so it is crucial to assure that the central social aspect of human rights is not overlooked, namely that when and where human rights have been claimed, it has been because the individuals concerned suffered from and protested against forms of exploitation and/or oppression that they believe neglected their dignity as human beings.[38] Human rights emphasise standards of treatment that no individual could justifiably deny to others and should therefore, be protected in a legitimate social order.[39] This reflexively implies that the right to justification, namely the claim human beings have 'to be respected as autonomous agents who have the right to not be subjected to certain actions or institutional norms that cannot be adequately justified to them', underlies all human rights.[40]

According to Forst the reflexive argument has three dimensions, the first being that human rights have a common ground in one basic moral right, namely the right to justification. Second, the political and legal function of human rights is to make the right to justification socially effective, both in a substantive and procedural manner. The substantive aspect entails developing

[35] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 711.
[36] ibid 712.
[37] ibid 711.
[38] ibid 712.
[39] ibid.
[40] Ibid.

and defining rights that express adequate forms of mutual respect the violation of which cannot be sufficiently justified between equal and free individuals. The procedural aspect consists of the crucial condition that no one should be subjected to a set of duties and rights which have been determined and in which one cannot participate in as an autonomous agent of justification. Therefore, human rights do not only safeguard the agency and autonomy of individuals, but they also express their autonomy in a political manner.[41] Thirdly, the reflexive argument Forst puts forward claims that by grounding human rights in this way it is not accessible to the charge of ethnocentrism haunting many justifications of human rights. The charge itself requires a right to adequate justifications that do not exclude the individuals affected.[42] Altogether, Fort's reflexive approach construes the very notion of justification in a normative manner as a basic concept of practical reason and as a practice that implies the moral right to justification. More importantly, this approach interprets the right to adequate justification as a practice that grounds human rights based on the right to justification.[43]

To understand the deeper normative grammar of human rights it is necessary, according to Forst, to keep the historical dimension of human rights in mind. Human rights first appeared as 'God-given' or 'natural' rights in early modern social conflicts and quite often revolutions.[44] The language of these rights was a politically and socially emancipatory language that was aimed against a feudal social order and an absolute monarchy that demanded 'divine' rights for itself. Forst highlights that many human rights views have a tendency to neglect the fundamental message of human rights, which is the claim to be a fully integrated member of society who is free from political or arbitrary social domination and who 'counts' and is recognised as someone with dignity and the effective right to justification.[45] Thus, the right to justification implies that there can be no legitimate political or social order that cannot be

[41] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 712.
[42] ibid.
[43] ibid.
[44] ibid 716.
[45] ibid 717.

adequately justified to those who are subject to it. The notion and right to participate in the political structures that determine rights and duties is not merely a right established in the contemporary human rights context, but can be traced back to seventeenth-century England when the Levellers argued to be independent political and social agents, free from feudal domination.[46]

Forst believes a brief historical reflection is also necessary when dealing with the issues of normative substance, legal function and moral justification of human rights. By using a distinction between ethics and morality, as developed by Dworkin and Habermas, Forst considers that a notion of human rights must have an independent and adequate moral substance and justification which does not rely on an idea of the good.[47] Moreover, Forst constructs the moral basis for human rights as 'the respect for the human person as an autonomous agent who possesses a right to justification'.[48] This means the right to, as an agent, demand acceptable reasons for actions that claim to be morally justified and for any political or social structure or law that claims to be binding upon an individual.[49] Based on a fundamental moral requirement of respect, human rights ensure that everyone has an equal status in the political and social spheres.[50] The primary goal and function of human rights is to secure, guarantee and express each individual's status as an equal given their right to justification.[51]

Furthermore, Forst argues that a moral justification for human rights must not only be universally valid, but it must also be reflexive. The meaning behind the reflexivity is the very notion of justification itself being redefined with respect to its practical and normative connotations.[52] The reflexive argument sets forward that since 'any moral justification of the rights of human beings must be able to redeem discursively the claim to general and reciprocal validity raised by such rights, then such a justification presupposes the right

[46] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 717.
[47] ibid 718-719.
[48] ibid 719.
[49] ibid.
[50] ibid.
[51] ibid.
[52] ibid.

to justification of those whose rights are in question'.[53] These individuals have a qualified right to veto any justification which does not pass the criteria of reciprocity and generality and which can be criticized as arbitrary or paternalistic. Reciprocity entails that no individual may make a normative claim, such as a rights claim, that they deny to others. This is referred to as reciprocity of content. Moreover, another aspect of reciprocity is reciprocity of reasons which means that no individual may project one's own values, interests, perspectives or needs onto others in a way where one claims to speak in their 'true' interests or in the name of some truth beyond mutual justification.[54] The criterion of generality means that all affected parties must be able to share the reasons supporting general normative validity, given their legitimate and reciprocal claims and interests.[55]

Forst also believes that the notion of human dignity and personhood is central in human rights discourse.[56] According to Forst, the notion of dignity does not have a metaphysical or ethical meaning. Dignity means that 'a person is to be respected as someone who is worthy of being given adequate reasons for actions or norms that affect him or her in a relevant way'.[57] As each person is an authority in the space of reasons, dignity is a relational term which can only be ascertained by way of discursive justification.[58]

As mentioned in chapter 2.1 the essential issue at stake is where to detect the normative 'anchor' of a notion of human rights. Forst follows the criteria of reciprocity and generality and puts forward that according to these criteria mutual justifiability is what confers normative weight to essential rights claims. For mutual justification to be regarded as a morally binding procedure, the rights claim to be a subject of justification must be prioritised and seen as independently morally valid. Thus, Forst advocates for a view where there is no 'derivation' of certain rights from basic interests in pursing

[53] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 719.
[54] ibid 719-720.
[55] ibid 720.
[56] ibid 723.
[57] ibid 734.
[58] ibid.

the good. Instead, human rights are the result of a discursive, intersubjective construction of rights claims 'that cannot be reciprocally and generally denied between persons who respect one another's right to justification'.[59] This type of respect is owed in a deontological way, something which is crucial to carry the weight of what is meant by human rights.[60] Individuals who have the status of normative agency have a human right to particular forms of respect seeing as one cannot reasonably justify a denial of their basic claims.[61]

Furthermore, human rights and democracy have an important connection. Forst holds that the normative grammar of human rights, both systematically and historically, requests for an understanding of basic rights to democratic participation.[62] An understanding which is shared by Besson who emphasises the mutual and close relationship between democracy and human rights.[63] She believes that in order for human rights to be democratically legitimate, human rights ought to be the result of a process in which human rights-holders are able to be the authors of their own rights.[64] Forst believes human rights should be understood as rights that end political oppression and the imposition of social status which strips one of one's freedom and access to social means crucial to being an individual of equal standing. By grounding human rights on the right to justification, the political and social meaning of human rights is captured and in opposition to earlier and modern forms of social exclusion. Inclusion is about being considered 'as an agent worthy of effective political justification, of giving and receiving reasons in the political realm'.[65] Seeing as every rights claim must be reciprocally and generally justifiable to be binding, it is these criteria which determine its content. The normative rights claim is not determined by the ethical judgment about the value of a practice but is instead determined by a claim concerning a legal and social standing

[59] i Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 722.
[60] ibid 723.
[61] ibid 724.
[62] ibid 725.
[63] Besson, 'Human rights and democracy in a global context: Decoupling and recoupling' (n 34) 40.
[64] ibid 30.
[65] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 725.

that cannot reasonably be denied to citizens who are recognized as social equals.[66]

The first question of human rights is therefore not about limiting sovereignty from the outside, but it is about the fundamental prerequisites of the possibility of establishing legitimate political authority. The question of legitimate intervention is however not an easy task, seeing as numerous factors need to be considered. International law and a politics of intervention must follow a certain logic of human rights, thus refraining from putting the cart before the horse.[67] Human rights serve primarily to ground internal legitimacy and not to limit internal sovereignty or autonomy. Claiming external respect depends on the internal respect which is based on justified acceptance. This does not however mean that one can infer the legitimacy of intervention directly from an absence of internal acceptance. According to Forst, 'violations of human rights place the internal legitimacy of a social and political structure in question, but they do not automatically dissolve the independent standing of that state in the international arena'.[68] Thus, human rights provide grounds for constructing a basic political and social structure in the right way; where the primary perspective of human rights is from the inside and not that of the outsider, who witnesses a political structure and asks whether there are reasons for intervention.[69] The political dimension of the right to justification is thus especially important. Not only does the moral justification for human rights have to be reflexive, but human rights also have a reflexive nature: they are basic rights to participate in the processes that give citizens' fundamental rights a clear and enforceable form. These rights are of a higher order as they are rights not to be subjected to legal norms or social institutions that cannot be adequately justified to those concerned by them. The ultimate aim with human rights, ideally speaking, is 'a fully justified basic structure'.[70]

[66] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 726.
[67] ibid.
[68] ibid 727.
[69] ibid.
[70] ibid 736.

Forts also highlights that human rights are a fundamental part of social and political justice, however they are merely a part.[71]. According to Forst it is the state's responsibility and task to assure human rights as well as protect citizens from having their human rights violated by private actors, for instance large companies. Failing to do so, either due to the state disregarding the real possibility to act or because the state is too weak, constitutes inadequate protection of human rights. This despite the violation not being the work of the state but of other agents. Although the main addressee of claims to protect human rights is the state it does not mean that it is the only agent who can violate these rights.[72] Human rights also apply to organisations and companies and those employed by them, such as individuals working in the technology sector. Using human rights as a framework helps to clarify who has moral responsibilities to do what in certain situations.[73] Technology companies have significant responsibilities within the system of duty-bearers, something which is emphasised in the United Nations Guiding Principles on Business and Human Rights (UNGP).[74] The UNGP sets down the specific responsibilities businesses have in relation to the respect of human rights, emphasising the necessity of identifying, preventing, and reducing prominent risks to human rights.[75] Those who create new technologies, such as AI systems, must make an informed effort to understand the significance technologies will have for rights holders. Thus, it is fundamental for tech businesses, including private actors developing, deploying and using AI, to move beyond good intentions and focus on enabling measures which will uphold human rights and the right to justification through evaluation, reviews, and various assessments.[76] Human rights-based considerations can provide several valuable functions in the context of AI, one being the understanding of how ethical principles guide the development and deployment of AI

[71] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 737.

[72] ibid.

[73] Prabhakaran and others, 'A Human Rights-Based Approach to Responsible AI' (n 13) 5.

[74] ibid 6.

[75] ibid.

[76] ibid.

systems as well as how these principles generate different responsibilities for the actors that constitute various parts of the ecosystem of AI.[77]

2.3 Justice and the Right to Justification

Forst does not only discuss the meaning of human rights but also questions how justice is generally viewed and understood. He presents and discusses two pictures of justice: the first residing in the moral idea that human beings 'should not lack certain goods that are necessary for a good life or one befitting a human being' and the second viewing human beings as beings 'whose dignity consists in not being subject to domination'.[78] Forst underlines the importance of both ideas but emphasises that in order to understand the grammar of justice the second idea must be the central image of justice.[79]

The first question of justice is the question of power.[80] Justice has its 'proper place where the central justifications for a social basic structure must be provided and the institutional ground rules are laid down which determine social life from the bottom up'.[81] Allocating goods is not only about legitimate distribution, but it also concerns how the goods come to be in the first place and how this distribution is made. Allocative-distributive focused theories view justice from the perspective of the recipient, leading to a lack of consciousness of power and a lack of emphasis of the political question of how the allocation of goods and the structures of production are decided.[82] The principle of justification enables a power to demand justification and challenge false legitimations. It also allows individuals to be regarded as independent agents of justice, ensuring that their dignity and autonomy is not violated by merely viewing them as recipients of redistributive measures.[83]

[77] Prabhakaran and others, 'A Human Rights-Based Approach to Responsible AI' (n 13) 11.

[78] Rainer Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (1st edn, Bloomsbury Academic 2014) 24-25.

[79] ibid 25.

[80] ibid ix.

[81] ibid 22.

[82] ibid 21.

[83] ibid 22.

Moreover, the belief that every person shall get the goods they deserve leads to either comparisons between individuals sets of goods or gives rise to the issue of whether individuals have 'enough' of essential goods. This type of recipient-oriented point of view does nonetheless have value according to Forst, seeing as distributive justice is concerned with the goods individuals can claim in an appropriate way.[84] Nevertheless, this picture of justice does also overlook fundamental and crucial aspects of justice. Goods-focused views disregard the question of how the goods come into existence, which results in issues of just organisation and production. Furthermore, this type of perspective on justice downplays the political question of who establishes the distribution and the structures of production and in what ways.[85]The political point of justice must be recognized, and one must liberate oneself from a one-sided, goods-fixated picture of justice. Instead, justice must be based on intersubjective relations and structures and not putatively objective states of the provision of well-being or of goods. By considering the question of justifiability of social relations and how much justification power groups or individuals have in a political background, and thus expanding the goods-fixated views of justice, a radical and critical conception of justice can be developed. This can in turn get to the roots of the relations of injustice.[86] The basic question of justice is, according to Forst's theory, how you are treated and not what you have.[87]

Furthermore, the concept of justice has a core meaning that is different from the concept of arbitrariness. Arbitrariness entails that people, or a part of the community (such as a class), can dominate over others without reason, which is rationalized as an immutable fate. Justice, however, means that people have equal rights and are treated fairly. It is a task which must be carried out by humans aiming at non-domination and not by Gods who are aiming at a world without historical or natural contingency.[88] Arbitrariness as domination is a

[84] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 4.

[85] ibid.

[86] ibid 6.

[87] ibid.

[88] ibid.

human vice and means of injustice. The term 'domination' is important in the context of justice seeing as it means that people are ruling without good reasons and without any legitimate structures of justification in place.[89] Thus, domination is to be understood as rule without justification.[90] A just social order is an order where individuals have equal rights and where they can give their consent, not only their counterfactual consent, but also a consent which is based on institutionalized justification procedures. Forst argues for the supreme principle of general and reciprocal justification, which sets forth that every claim for rights, liberties and goods must be justified in a general and reciprocal manner as a way of preventing one side projecting its reasons onto the other. A justification must take place in a discursive manner.[91]

Furthermore, the impulse that opposes injustice is not primarily motivated by a desire of wanting something, or more of it, but by a desire to be free from harassment, domination or being overruled in one's claim to the basic right to justification.[92] This moral right to justification asserts that political and social relations which cannot be sufficiently justified towards those involved should not exist, as mentioned in chapter 2.2. Forst argues that this political essence of justice is obscured and suppressed by the recipient-focused perception of the principle suum cuique. Justice rests on the idea that each individual should be respected in one's dignity, offering and demanding justifications.[93] Justice is always about what human beings owe to one another given the relations between them. Thus, justice is to be seen as a relational matter, where the relations between individuals are fundamentally in need of justification.[94] Forst emphasises the vast difference between someone who lacks certain goods for any reason, for instance due to a natural catastrophe, and someone who is deprived of specific opportunities and goods in an unjust way without justification.[95] The primary victim of injustice is not the person who lacks

[89] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 7.
[90] ibid 20.
[91] ibid 21.
[92] ibid 8.
[93] ibid.
[94] ibid 10-11.
[95] ibid 11.

certain goods but rather the one who does not count in the decision-making process of producing and allocating goods.[96] The decision of how goods shall be produced and distributed should involve all human beings, allowing them to enjoy equal rights in the order of justification.[97] Social and political justice is about ensuring that the political and social systems in which we live are just. For this to be possible, every individual must be involved in the political and social process and the result of these processes must be just in a way where they can be accepted by all, enabling non-domination.[98] The essential demand of justice is that all individuals should be supplied with equal rights in the political and social context and based on this, claim certain goods.[99]

Moreover, institutions play an essential role in the matter of justice and the right to justification. Institutions are the primary objects of assessment in the matter of social injustice. They represent fundamental expressions of social life and are guarantors for the realization of equal respect. However, how the institutions work can cause a violation of the principles of equal respect, both to the outcomes but also to the processes. When dealing with a result as an outcome from an institution there are essential aspects which are important to how the result came about, such as who participated in the decision, which factors were decisive and what interests were considered.[100] Justice is at its core about who determines the structure of society and its institutional workings and with what justification. The principles of participation, non-domination and equal respect must be prioritised within the framework of a society.[101] What is crucial is that institutions operate in a manner which aligns with generally justified principles and that this does not include any social privileges. Further, the principles must not lead to certain groups being largely excluded from the system of cooperation.[102] Forst argues that 'it is one thing to argue for a better distribution and realization of basic capabilities by

[96] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 8.
[97] ibid 11.
[98] ibid 12.
[99] ibid 13.
[100] ibid 16.
[101] ibid 17.
[102] ibid 20.

way of a theory of social development and progress, yet it is another thing to argue for a comprehensive conception of social and political justice'.[103] Thus, the most crucial of all principles of distribution is the one which establishes *who* has the authority to receive a certain good in the first place.[104]

2.4 Fundamental rights within the EU

Member States have a duty to protect human rights, seeing as EU fundamental rights do not only bind institutions but also Member States when they implement EU law.[105] The point of protecting fundamental rights in EU law is the need to maintain the unity, primacy and effectiveness of Union law, as expressly affirmed in the case of Siragusa.[106] The case establishes that 'the reason for pursuing that objective [of protecting fundamental rights in EU law] is the need to avoid a situation in which the level of protection of fundamental rights varies according to the national law involved in such a way as to undermine the unity, primacy and effectiveness of EU law'.[107] Thus, fundamental rights became part of Union law not because of the aim to ensure the protection of such rights per se, but rather due to the need to protect the unity, primacy and efficacy of EU law.[108] When Member States protect their own national fundamental rights it can interfere with the primacy, uniformity, and efficacy of Union law.[109] Therefore, to ensure that these are protected, EU takes over the safeguarding of human rights, not as national rights, but as EU fundamental rights.[110] The obligations imposed on Member States by the Union must treat all Member States equally. Thus, the EU fundamental rights which shape and condition the obligations must also treat the States of the Union in an equal manner. The rights must be uniform EU fundamental rights, and not Member States' fundamental rights.[111] This

[103] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 17.
[104] ibid.
[105] See Article 52 CFR.
[106] Eduardo Gill-Pedro, *EU law, Fundamental Rights and National Democracy* (1st edn, Routledge 2018) 115.
[107] Case C-206/13 *Cruciano Siragusa v Regione Sicilia* [2014] EU:C:2014:126, para. 32.
[108] Gill-Pedro, *EU law, Fundamental Rights and National Democracy* (n 106) 117.
[109] ibid 118.
[110] ibid.
[111] ibid 121.

means that EU measures which harmonize the obligations of Member States, and of private actors within the Member States, to protect EU fundamental rights, must also comply with the right to justification, seeing as this right is the normative foundation of national human rights, as presented in chapter 2.2.

It is important to note that the Union does not have competence to impose obligations on Member States in pursuance of protecting human rights.[112] The EU does however have several specific human rights competences which aim to respect and protect human rights within the scope of its other competences, such as non-discrimination as well as the protection of human rights for asylum seekers.[113] This competence is referred to as indirect human rights competence.[114] The EU does not have the capability to legislate and adopt measures which would respect and promote fundamental rights outside the scope of its other competences in the treaties. This constraint is based on Member States' wish to retain their own human rights competence. The resistance to a centralised EU human rights competence has been expressed, among other things, when Member States have denied the existence of such competence in the so-called standstill clauses, including Article 51 paragraph 2 of the CFR.[115] The resistance against EU law becoming a direct source of human rights was further emphasised by the fact that the United Kingdom, Czech Republic and Poland opted out of the CFR when it became binding in 2009.[116]

In light of this, it is clear that fundamental rights are an essential component of the EU.[117] While the protection of fundamental rights holds a great importance in the contemporary debate within the Union,[118] human rights

[112] Opinion 2/94 of the Court on Accession by the Community to the European Convention for the Protection of Human Rights and Fundamental Freedoms, [1996] ECR 1996 I-01759, EU:C:1996:140, para. 23.

[113] Besson, 'Human rights and democracy in a global context: Decoupling and recoupling' (n 34) 34.

[114] ibid 34-35.

[115] ibid.

[116] ibid 37.

[117] Giacomo di Federico, *The EU Charter of Fundamental Rights: from declaration to binding instrument* (1st edn, Springer 2011) V.

[118] ibid 4.

have not always a pressing concern in the EU legislation.[119] As the previous treaties, such as the European Convention on Human Rights (ECHR) and the Universal Declaration of Human Rights (UDHR), focused on human rights, the EU founding treaties did not refer to human rights or bind the institutions and Member States of the EU to human rights duties, apart from the principle of non-discrimination and the equality between women and men.[120] Prior to the implementation of the CFR fundamental rights were recognised the status of 'general principles of the law' by the CJEU.[121] The acknowledgement by the Court led to fundamental rights enjoying a minimal protection in the case law of the CJEU.[122] Thus, the sources of EU fundamental rights were, and broadly still are, indirectly derived as general principles of EU law from Member States' international human rights duties as well as national constitutional traditions.[123]

EU law has increasingly had a direct effect on fundamental individual interests where individuals within the EU have been recognised as direct fundamental rights bearers as well as EU citizens under EU law.[124] The EU fundamental rights framework ought to be described as an attempt to harmonise the fundamental rights norms in Europe. What they all have in common is their understanding of rights being moral standards on how we treat one another. One of the most important treaties within the Union is the Lisbon Treaty which was signed in 2007 and is based upon the TEU and the TFEU.[125] The Treaty, which entered into force in December 2009, introduced several fundamental changes to the human rights protection in the EU, the most significant being the amendments to Article 6 of the TEU which

[119] Sionaidh Douglas-Scott, 'The European Union and Human Rights after the Treaty of Lisbon' (2011) 11(4) *Human Rights Law Review*, 647 <https://doi.org/10.1093/hrlr/ngr038> accessed 17 April 2023.

[120] Besson, 'Human rights and democracy in a global context: Decoupling and recoupling' (n 34) 34.

[121] Douglas-Scott, 'The European Union and Human Rights after the Treaty of Lisbon' (n 119) 648.

[122] di Federico, *The EU Charter of Fundamental Rights: from declaration to binding instrument* (n 117) 5.

[123] Besson, 'Human rights and democracy in a global context: Decoupling and recoupling' (n 34) 34.

[124] ibid 33.

[125] di Federico, *The EU Charter of Fundamental Rights: from declaration to binding instrument* (n 117) V.

recognises the rights, freedoms, and principles of the CFR.[126] Moreover, Article 2 of TEU as amended establishes that the Union is 'founded on the values for human dignity, freedom, democracy, equality, the rule of law and respect for human rights'. The amendments enhanced provisions with the aim to strengthen the protection of fundamental rights in the EU and actualised the binding effect of the CFR, which in turn led to the CFR gaining primary EU law status.[127] The Charter has since then become the primary source of human rights within the Union and has on several occasions been referred to by the CJEU, which has increased the Court's profile within the field of human rights in the EU.[128] The protection of fundamental rights in the EU has evolved in an ad hoc manner, leading to the importance of the CFR becoming somewhat of an identifier and road map of EU rights.[129] The second recital of the preamble gives the following precision:

> Conscious of its spiritual and moral heritage, the Union is founded on the indivisible, universal values of human dignity, freedom, equality and solidarity; it is based on the principles of democracy and the rule of law. It places the individual at the heart of its activities, by establishing the citizenship of the Union and by creating an area of freedom, security and justice.

The importance of dignity is further emphasised in Article 1 of the CFR, while Articles 2 and 3 of the CFR establish and highlight the right to life and the right to the integrity of the person. Thus, the preamble implies an understanding of the legal order of the Union where the individual is accorded a greatly important status. Similar to how Forst believes that the notion of human dignity is central in human rights discourse,[130] the preamble of the Charter establishes that dignity and universal human rights are to be respected in a legitimate and political order. The articles established in the

[126] Douglas-Scott, 'The European Union and Human Rights after the Treaty of Lisbon' (n 119) 645.
[127] ibid.
[128] ibid.
[129] ibid 649.
[130] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 723.

Charter are based on the understanding of how human beings are to be treated, which aligns with Forst's basic question of justice, namely the treatment of independent agents with a right to dignity and autonomy.[131]

Moreover, pursuant to Article 51 (1) of the CFR, all EU institutions, bodies, offices, and agencies must respect the rights established in the Charter. Additionally, the Article also applies to Member States when implementing Union law. Article 8 of the CFR establishes that 'everyone has the right to the protection of personal data concerning him or her'. Article 7 corresponds to those rights guaranteed by Article 8 of the ECHR and establishes everyone's right to have their private and family life, home and communications respected. The protection of these specific rights is one of the essential aims adopted by the EU legislature seeing as these two rights are especially important in the context of AI.[132]

Furthermore, it is important to note that EU fundamental rights law is inspired by several essential external instruments, one being the ECHR, which entered into force 1953. The ECHR is a fundamental human rights treaty which gained status as secondary EU law and is today one of the most fundamental documents concerning human rights.[133] Similar to the CFR, the concept of human dignity forms the fundamental basis and essence of the Convention.[134] Article 6(3) of the consolidated TEU states that the Union shall accede to the ECHR and that the fundamental rights as assured in the Convention and as they result from the constitutional traditions common to the Member States, shall constitute general principles of EU law. The Article displays that there is a clear commitment of the Union to respect fundamental rights in the way

[131] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 6.

[132] European Data Protection Board and European Data Protection Supervisor (EDPB and EDPS) 'Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' (2021) 2 <https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf> accessed 27 March.

[133] Manfred A. Dauses, *The Protection of Fundamental Rights in the Legal Order of the European Union* (1st edn, Frankfurt am Main: Lang 2014) 34.

[134]European Court of Human Rights, 'Guide on Article 8 of the European Convention on Human Rights' (2022), 26 <https://www.echr.coe.int/documents/guide_art_8_eng.pdf> accessed 23 March 2023.

that it reflects national fundamental rights in the Member States. The rights in the ECHR do not solely reflect the commitments between the different Member States but they also display the commitments that each Member State has to every individual within its own jurisdiction.[135] All Member States are signatories to the ECHR and provide some level of constitutional protection of fundamental rights, meaning they are required to comply with their national protection of human rights.[136] As mentioned in chapter 2.3, human rights and justice is about how we treat each another.[137] The right to justification ought to be upheld by the Member States seeing as this right underlies national human rights. As previously stated, the reasons for protecting the fundamental rights in EU law is the aim to preserve the primacy, unity, and efficacy within the Union.[138] Thus, the fundamental rights within the Union are intended to harmonise national human rights. These latter rights would however not be construed in light of the framework and goals of EU law as they are not EU fundamental rights, which could undermine the unity, precedence, and efficacy of EU law.[139] As national law, including human rights, are given a status as general principles of EU law national human rights law must be treated as legal norms that reflect moral concerns about how people are treated, leading to the necessity to respect and uphold the right to justification. As Article 2 of the TEU sets forth, one of the key values the Union is based on is the respect for human rights. Thus, when legislating and imposing duties on Member States, the EU must ensure that all fundamental rights are respected.

Furthermore, although the EU has not yet acceded to the ECHR, the ECHR was and continues to be an important inspiration for the human rights framework within the EU, as shown in Article 53(3) of the CFR. The Article sets forth that the meaning and scope of rights in the CFR, which correspond to the rights guaranteed in the ECHR, shall be the same as those laid down by

[135] Gill-Pedro, *EU law, Fundamental Rights and National Democracy* (n 106) 70-71.
[136] ibid 121.
[137] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 6.
[138] Gill-Pedro, *EU law, Fundamental Rights and National Democracy* (n 106) 118.
[139] ibid 121.

the Convention. Thus, the Union is showing its commitment to the ECHR by ensuring that the rights in the Charter are given the same meaning and scope as the rights of the ECHR.

Another important instrument that has inspired the Union's law and stresses the political meaning of human rights by also putting human dignity at its centre, is the UDHR.[140] The modern notion of human rights emerged broadly after the Second World War as a reaction to the moral traumas and the crimes committed in the Holocaust.[141] Thus, the Declaration was deeply influenced by the most extreme and cruel forms of tyranny, resulting in the adoption of the UDHR in 1948.[142] The Declaration reiterates the connection between being a participant in political affairs and being safe from arbitrary and unjust rule. The social and international order which is set forward in the UDHR is to be one in which no set of legally binding rights is decided without the participation of those who are the subjects of the rights established in the declaration.[143] Seeing as there is no universal ratification of the human rights treaties, the UDHR has been embraced and used by the EU to set standards in international agreements, internal legislation and to guide the Union's external policy.[144] Moreover, the Declaration is a useful reference for the way in which the Union perceives fundamental rights.[145] For instance, Article 1 of the UDHR sets forward that all humans are born free and equal in dignity and rights, a right which has inspired Article 1 of the CFR and is shared by the Union. The CJEU established in a judgement that a fundamental right to human dignity is part of Union law.[146]

---

[140] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 718.

[141] Prabhakaran and others, 'A Human Rights-Based Approach to Responsible AI' (n 13) 2-3.

[142] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 718.

[143] ibid.

[144] Ionel Zamfir, 'At a Glance - the Universal Declaration of Human Rights and Its Relevance for the European Union ' (European Parliamentary Research Service 2018) 1 <https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/628295/EPRS_ATA(2018)628295_EN.pdf> accessed 15 April 2023.

[145] ibid 2.

[146] Case C-377/98 *Netherlands v European Parliament and Council* [2001] ECR I-7079, paras 70 — 77.

2.5 Conclusion

In conclusion, Forst establishes that human rights are 'those basic rights without which the status of a being with a right to justification is not socially secured'.[147] Human rights have three different lives: moral, legal and political.[148] Forst also sets forward that the right to justification underlies all human rights[149] and that 'each member of a context of justice has a fundamental right to justification'.[150] Moreover, human rights are meant to ensure that no individual is being treated in a way that could not be justified to them as a person equal to others.[151] These rights include the fundamental political, personal, and social rights required to establish the social structure of justification. Furthermore, they also entail substantive rights which no one can reasonably deny to others without disregarding the requirements of reciprocity and generality.[152] Human rights protect and materialize the status of human beings as autonomous social beings. Thus, through procedures of reciprocal and general justification, claims based on human interests can be transformed into rights claims.[153] Forst's central idea is that the purpose of human rights is that individuals have the fundamental right to live in a society where they, as social and political agents, can decide which rights they have to recognize and which they can claim. Thus, human rights underline and highlight the autonomous agency. They have a reflexive nature, meaning they are rights that protect against a multitude of social harms, especially the harm of not being part of the political determination.[154] The normative ground of human rights is essentially the fundamental claim to be respected as an agent who has a right to justification.[155]

---

[147] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 737.

[148] ibid 711.

[149] ibid 712.

[150] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 21.

[151] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 712.

[152] ibid 737.

[153] ibid 736.

[154] ibid.

[155] ibid 739.

Within the EU, the CFR plays a fundamental role in safeguarding human rights, such as human dignity and the important status of the individual. Although the EU has not acceded to the ECHR the Union is committed to the Convention as it has been an essential inspiration for the CFR. The Charter and the Convention both assure that human rights are to be respected, creating a legitimate political order in which human rights and the right to justification are protected. Seeing as the Union must ensure that fundamental rights are respected when legislating and imposing duties on Member States, the AI Act and other harmonising measures of the Union must also impose a standard of human rights protection. However, this protection must also respect the right to justification seeing as this moral right is what underlies human rights in the Member States. The principle of justification enables a power to demand justification and challenge false legitimations. It also allows individuals to be regarded as independent agents of justice and democracy, ensuring that their dignity and autonomy is not violated by merely viewing them as recipients of redistributive measures.[156] A just and democratic social order is an order where individuals have equal rights and where they can give their consent, not only their counterfactual consent but also a consent which is based on institutionalized justification procedures.[157] The moral right to justification asserts that political and social relations which cannot be sufficiently justified towards those involved should not exist.[158] Thus, the AI Act must ensure that the provisions respect human dignity and the right to justification.

[156] Forst, *Justice, democracy and the right to justification Rainer Forst in dialogue* (n 78) 22.

[157] ibid 20-21.

[158] ibid 10-11.

3 The proposed AI Act and trustworthy AI

3.1 About the chapter

The following chapter will present the proposed AI Act as well as trustworthy AI and its connection to human rights. The purpose of the chapter is to provide an overall insight into the Act, which then forms the basis for the discussion that follows in subsequent chapters. However, to understand the purpose of the AI Act an account of the work that has led to the proposal is necessary. Thus, the chapter begins with an overview of the discussions which have followed within the EU as well as globally regarding the emergence of AI. Furthermore, the chapter will thereafter describe the Union's work with AI from recent years, demonstrating what has led to today's proposal. This is later followed by a presentation of the Act itself, with primary focus on the definition of AI as well as the objectives and purpose of the proposal. Lastly, the chapter concludes with a conclusion that makes links to the previous chapter.

3.2 The emergence of AI

The origins of the field of AI can be traced back to the 1950s, when British computer scientist Alan Turing asked himself whether machines can think.[159] The scientific community does not agree on a single definition of AI, and the term 'AI' is frequently used as a blanket term to refer to a variety of computer applications built using various techniques which display capabilities associated to human intelligence.[160] In recent years, AI has come to refer to a

[159] Gonçalo Carriço, 'The EU and Artificial Intelligence: A Human-Centred Perspective' (2018) 17:1 *European View*, 1 <https://journals.sagepub.com/doi/full/10.1177/1781685818764821> accessed 8 May 2023.

[160] Tambiama Madiega, 'Briefing EU Legislation Artificial Intelligence Act' (European Parliamentary Research Service 2022) 4 <https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf> accessed 13 February 2023.

machine or system that is able to solve issues that human beings usually solve with natural intelligence.[161]

Over the last decades, AI-automated reasoning, learning and decision-making have become prevalent parts of our daily lives. AI provides hundreds of millions of people with search results, book recommendations, optimized GPS routes, language translations and numerous companies have and are working on developing self-driving cars.[162] In addition to AI's daily role in our lives, AI is also playing a significant role in the field of medicine and science, for instance identifying rare but calamitous side effects of medications. The contributions of AI do not only have a big influence today but will make even more profound contributions in the future.[163] AI systems are believed to bring monumental benefits to society in general but also to the organisations and companies who use them. Increases in job creation, growth and innovation are only a few examples of the impact AI will have on European Industry.[164]

Nonetheless, the development of AI has generated discussions regarding the risks with AI systems.[165] The growth of digital technology, including AI, has been discussed in the EU since 2015 when the Union presented their Digital Single Market Strategy. The strategy aims towards strengthening the internal market through improving access to digital technology as well as facilitating technological development and expanding growth and innovation.[166] Several

[161] Carriço, 'The EU and Artificial Intelligence: A Human-Centred Perspective' (n 159) 31.

[162] Thomas G Dietterich and Eric J Horvitz, 'Rise of concerns about AI: reflections and directions' (2015) 58(10) *Communications of the ACM*, 38 <https://mags.acm.org/communications/october_2015/?folio=38&&pg=40#pg40> accessed 14 April 2023.

[163] ibid.

[164] James Eager, 'Opportunities of Artificial Intelligence, Study for the committee on Industry, Research and Energy, Policy Department for Economic, Scientific and Quality of Life Policies' (2020) PE 652 713, 35 <http://www.europarl.europa.eu/RegData/etudes/STUD/2020/652713/IPOL_STU(2020)652713_EN.pdf> accessed 27 February 2023.

[165] Dietterich and Horvitz, 'Rise of concerns about AI: reflections and directions' (n 162) 39.

[166] European Commission, 'Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions – A Digital Single Market Strategy for Europe' COM(2015) 192 final, 6 May 2015, 3ff.

institutions within the Union have expressed concerns regarding the potential risks with the growing reliance on AI systems, particularly the threat to fundamental rights and democracy.[167] Parliament has raised concerns regarding ethical issues and the enjoyment of several fundamental rights, such as the right to privacy, the right to life and the freedom of expression.[168] Furthermore, organisations with a connection to technology have in the past years either endorsed or authored a set of ethical principles for AI.[169] Different actors, such as companies, professional associations and civil society, have authored different principles but with common key themes, such as transparency and the promotion of human values.[170] Thus, the work and discussions regarding AI and its potential challenges concern several actors. One essential example from the world of economics is the work of the Organisation for Economic Cooperation and Development (OECD). In May 2019 the OECD member countries, as well as six other countries, approved the 'OECD Council Recommendation on Artificial Intelligence', where five principles of AI were established.[171] Despite the principles not being legally binding the existing principles have been proven to be highly influential in setting an international standard as well as facilitating the design of national legislations.[172] AI initiatives have not only been taken from governments and stakeholders at a national level but the initiatives have also been made on an international level.[173] Several actors, including the EU, have expressed apprehension regarding the development of AI which has resulted in a pressure to regulate AI. Some Member States have already considered

[167] Mihalis Kritikos, 'Artificial Intelligence ante portas: Legal & ethical reflections' (European Parliamentary Research Service 2019) PE 634.427, 4 <https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/634427/EPRS_BRI(2019)634427_EN.pdf> accessed 27 March 2023.

[168] ibid.

[169] Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI' (n 12) 4.

[170] ibid 4-5.

[171] Organisation for Economic Cooperation and Development (OECD), 'Putting the OECD AI Principles into Practice: Progress and Future Perspectives - OECD.AI' (2021) <https://oecd.ai/en/mcm> accessed 3 May 2023.

[172] OECD, 'Forty-Two Countries Adopt New OECD Principles on Artificial Intelligence - OECD' (2019) <https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm> accessed 4 May 2023.

[173] OECD, 'Recommendation of the Council on Artificial Intelligence' (2022) OECD/LEGAL/0449, 3 < https://legalinstruments.oecd.org/api/print?ids=648&lang=en> accessed 4 May 2023.

national rules to ensure that AI is regulated in a safe way, ensuring that it adheres to fundamental rights obligations.[174] However, if Member States start introducing independent national regulations, the internal market will be fragmented and there will be a consequential diminishment of legal certainty on how existing national rules will apply to the AI systems in the EU. Thus, the Union found that these essential issues could best be solved through the Union harmonising legislation.[175] The AI Act will therefore not only affect national regulations in the Member States of the Union but will as 'the first initiative, worldwide, that provides a legal framework for AI' have an effect globally as well.[176] The implementation of the GDPR resulted in a 'Brussels Effect' abroad, an effect that some scholars believe will occur with the proposed AI Act as well.[177]

3.3 Background

As seen above, AI has been a controversial phenomenon for several years. The process leading up to the final proposal has been long running and has been shaped through varies initiatives and documents. The Commission has been facilitating cooperation on AI across the EU with the aim to expand the Union's competitiveness and safeguard trust based on Union values for several years.[178] In May 2015 the Digital Single Market strategy was introduced.[179] Two years later, the Commission published a mid-term review on the implementation of the Digital Single Market Strategy.[180] The review underlined the vast possibilities with AI, labelling it a 'key driver for

[174] COM(2021) 206 final, 6.

[175] ibid.

[176] EDPS, 'Artificial Intelligence Act: A Welcomed Initiative, but Ban on Remote Biometric Identification in Public Space Is Necessary' (2021) <https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative_en> accessed 3 May 2023.

[177] Charlotte Siegmann and Markus Anderljung, 'The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market' (2022) 3 <https://arxiv.org/pdf/2208.12645.pdf> accessed 9 May 2023.

[178] European Commission, 'Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence' (2021) <https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682> accessed 8 May 2023.

[179] COM(2015) 192 final, 1.

[180] European Commission, 'Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions on the Mid-Term Review on the implementation of the Digital Single Market Strategy – A Connected Digital Single Market for All' COM(2017) 228 final, 10 May 2017.

future economic and productivity growth'.[181] The Commission also emphasised the importance of the Union being in a leading position in the development of AI platforms, technologies and applications as well as the importance of ensuring that humans and robots can interact in the safest and best possible way.[182]

In 2018, the Union legally recognised the need for modern protections of technology by putting Europe's new data privacy and security law, the GDPR, into effect.[183] The requirement of providing data subjects with 'meaningful information about the logic involved' in a decision-making process that is automated was firstly introduced by the establishment of the GDPR.[184] The same year, the Commission published its Communication 'Artificial Intelligence for Europe', which included an initiative to implement a solid European framework.[185] The Communication emphasises the numerous benefits and new opportunities that arise with AI, from treating chronic diseases to anticipating cybersecurity threats.[186] It is necessary for the Union to have a harmonised strategy where the EU and the Member States as well as private and public actors cooperate with each other.[187] Moreover, the Communication stresses the importance of the Union ensuring that the future AI framework promotes innovation as well as respects the values and fundamental rights of the Union, especially the ethical principles of accountability and transparency.[188]

In December 2018, the Commission published its Coordinated Plan on AI, which included strengthening the cooperation between the Commission and the private sector to expand research and innovation, support the deployment

[181] European Commission, 'Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions on the Mid-Term Review on the implementation of the Digital Single Market Strategy – A Connected Digital Single Market for All' COM(2017) 228 final, 10 May 2017, 21ff.

[182] ibid 21-22.

[183] GDPR.eu, 'What Is GDPR, the EU's New Data Protection Law? - GDPR.eu' (2023) <https://gdpr.eu/what-is-gdpr/> accessed 10 May 2023.

[184] Kritikos, 'Artificial Intelligence ante portas: Legal & ethical reflections' (n 167) 2-3.

[185] COM(2018) 237 final, 1.

[186] ibid.

[187] ibid 3.

[188] ibid 2.

of AI as well as increase private investments. These three elements are essential to ensure that the Union does not fall behind in the technological development.[189] In order for AI to be used to its full potential, society needs to have confidence in the new technology. Human beings must understand how AI makes decisions. A human-centric approach and ethics-by-design principles is therefore required.[190] Trust is however only achieved when technology is safe, ethical, predictable and respects fundamental rights. For this reason, the Commission set up the expert group High-Level Expert Group on Artificial Intelligence (HLEG) tasked with developing a framework of ethical guidelines for AI.[191] Due to the increasing concern, as established in chapter 3.2, the Union brought together representatives from academia, the industry and civil society to ensure consensus.[192] Although the Guidelines are not binding and thus do not create any new legal obligations, many existing provisions of Union law already reflect several of the key requirements presented in the Guidelines, such as safety and personal data protection rules.[193] Furthermore, the Commission set out some guidelines that should be met for AI to be assessed as trustworthy and thus also ethical.[194]

In February 2020 the Commission published three important Communications, the first concerning Europe's digital future,[195] the second consisting of a European strategy for data[196] and the last being the White Paper on AI.[197] The purpose of the White Paper is to set out policy options

[189] European Commission, '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' COM(2018) 795 final, 7 December 2018, 3.

[190] ibid 8.

[191] ibid.

[192] Eduardo Gill-Pedro, 'The Most Important Legislation Facing Humanity? The Proposed EU Regulation on Artificial Intelligence' (2021) *Nordic Journal of European Law*, 4:1, V.

[193] COM(2019) 168, 3-4.

[194] ibid 3ff.

[195] European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Shaping Europe's digital future' COM(2020) 67 final, 19 February 2020, 2.

[196] European Commission, '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' COM(2020) 66 final, 19 February 2020, 1.

[197] COM(2020) 65 final.

concerning the promotion of the uptake of AI and the risks associated with certain uses of the new technology.[198] The Commission stresses the importance of making sure the Communication on Building Trust in Human-Centric AI and the Guidelines by the HLEG are considered, thus facilitating the process of trustworthy AI.[199]

Months after the publication of the White Paper, various stakeholders were given the opportunity to comment on the content of the White Paper. The majority emphasised that the measures presented in the Paper were crucial, however some believed that certain clarifications were needed, including the definitions of AI and the different risks of AI systems.[200] The Council later invited the Commission to provide a clear, objective definition of high-risk AI systems and emphasised that the Union needs to be a global leader in the development of trustworthy, ethical, and secure AI.[201]

Following up to previous work the Commission launched and presented its AI package in April 2021. This included its Communication on fostering a European approach to AI, a review of the Coordinated Plan on Artificial Intelligence and most importantly, the proposed AI Act. Moreover, a relevant impact assessment was also presented.[202] The proposed Regulation immediately prompted widespread reactions from several stakeholders and rapidly became an extremely significant event within the European Community as well as globally.[203] Being the world's first legal framework for AI, the proposal is expected to be a landmark piece of legislation.[204]

---

[198] COM(2020) 65 final, 1.

[199] ibid 2-3.

[200] COM(2021) 206 final, 8.

[201] European Council, 'Special meeting of the European Council (1 and 2 October 2020) – Conclusions' (2020) EUCO 13/20, 6 <https://www.consilium.europa.eu/media/45910/021020-euco-final-conclusions.pdf> accessed 10 May 2023.

[202]European Commission, 'A European approach to Artificial Intelligence' (2023), <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> accessed 14 February 2023.

[203]Luca Bertuzzi, 'EU Launches AI Blueprint in Bid to Become World Leader' (April 2021) <https://www.euractiv.com/section/digital-single-market/news/commission-launches-ai-package-proposal/> accessed 28 April 2023.

[204]European Parliament, 'AI Act: A Step Closer to the First Rules on Artificial Intelligence' (n 9).

Several documents were handed in after the publication of the proposed Regulation, such as draft reports, amendments and opinions.[205] This led to the Council adopting its general approach on the proposal in December 2022.[206] At the time of this thesis being written, the proposal and its provisions are being discussed and negotiated in Parliament. The discussions concerning the Regulation have taken longer than expected in consequence of political infighting in the Parliament. Parliament is however expected to finalise its position by May to quickly enter into negotiations with the Commission and the Council in the so-called trialogues.[207] EU lawmakers in the leading Parliament committees were scheduled to vote on the agreement on the proposed AI Act on 26 April. However, at the time of writing, the voting has been postponed.[208]

As a result of the European strategy for data in 2020, the Union's work with regulating AI continues. In 2022, a proposal for an AI Liability Directive was announced[209] as well as the Data Governance Act.[210] Following this, the Data Act was presented. The Act complements the Data Governance Act by clarifying who can create value from various data and under which conditions this can be done.[211] Thus, revealing the main assumptions and understanding

[205] Luca Bertuzzi, 'AI Regulation Filled with Thousands of Amendments in the European Parliament' (June 2022) <https://www.euractiv.com/section/digital/news/ai-regulation-filled-with-thousands-of-amendments-in-the-european-parliament/> accessed 3 May 2023.

[206] European Council, 'Artificial Intelligence Act: Council Calls for Promoting Safe AI That Respects Fundamental Rights' (n 8).

[207] Luca Bertuzzi, 'AI Act: European Parliament Headed for Key Committee Vote at End of April' (March 2023) <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-european-parliament-headed-for-key-committee-vote-at-end-of-april/> accessed 2 April 2023.

[208] Luca Bertuzzi, 'AI Act: MEPs Close in on Rules for General Purpose AI, Foundation Models' (April 2023) <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-meps-close-in-on-rules-for-general-purpose-ai-foundation-models/?utm_source=substack&utm_medium=email> accessed 2 May 2023.

[209] European Commission, 'Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM(2022) 496 final, 28 September 2022.

[210] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) [2022] OJ L152/1.

[211] European Commission, 'Data Act' (2022) <https://digital-strategy.ec.europa.eu/en/policies/data-act> accessed 8 May 2023.

behind AI-based decision-making is a great juridical concern that exists within several domains of the digital economy.[212]

3.4 The definition of AI

As presented in chapter 3.2 there is not one uniform definition of AI. However, within the EU the HLEG on AI proposed a baseline definition of AI which has had an increased use in the scientific literature within the field. The Commission emphasises that there is a strong need for a clear definition of the notion of an AI system as such a definition is essential for the distribution of obligations under the new AI framework. A clear definition also helps to ensure legal certainty.[213]

As the proposed AI Act is a moving target, the provisions of the Act are also continuously under discussion. In March 2023 the Parliament reached a political agreement to adopt an AI definition similar to the one used by the OECD. The agreed definition states that an AI system is a 'machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate output such as predictions, recommendations, or decisions influencing physical or virtual environments'.[214] Using the OECD's definition as inspiration is based on an effort to ensure legal certainty, harmonisation and wide acceptance.[215]

The OECD defines AI system as:

> …a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. When

[212] Kritikos, 'Artificial Intelligence ante portas: Legal & ethical reflections' (n 167) 2.
[213] Madiega, 'Briefing EU Legislation Artificial Intelligence Act' (n 160) 4.
[214] Luca Bertuzzi, 'EU Lawmakers Set to Settle on OECD Definition for Artificial Intelligence' (March 2023) <https://www.euractiv.com/section/artificial-intelligence/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/?utm_source=substack&utm_medium=email> accessed 10 March 2023.
[215] Ibid.

> applied, AI has seven different use cases, also known as patterns, that can coexist in parallel within the same AI system.[216]

The final draft of the proposed AI Act defines AI system in Article 3(1) as:

> …a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge-based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts;

3.4.1 High-risk AI versus low-risk AI

The AI Act follows a risk-based approach which means that AI systems are categorised in accordance with each systems' concrete level of risk.[217] The proposal distinguishes between systems which pose an unacceptable risk, high risk, limited risk and low or minimal risk. This approach allows AI systems to be regulated only to the extent it is strictly necessary, ensuring that the specific level of risk is sufficiently addressed.[218]

Article 5 of the proposed Act explicitly bans certain AI practises which are an evident threat to people's livelihoods, safety and rights due to the 'unacceptable' risk these systems create. Examples of such, according to the Article, are AI systems that exploit specific vulnerable groups and systems which are used by public authorises for social scoring purposes.

Article 6 regulates high-risk AI systems which can negatively affect people's fundamental rights or safety. The Article differentiates between two groups of high-risk systems: systems which are to be used in certain items that pose

[216] OECD, 'Artificial Intelligence & Responsible Business Conduct' (2019) 1 <https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf> accessed 2 March 2022.

[217] Madiega, 'Briefing EU Legislation Artificial Intelligence Act' (n 160) 5.

[218] ibid.

a high risk of harm to health and safety or fundamental rights, and AI systems which are deployed in eight specific areas in Annex III of the proposal. Examples of such are remote biometric identification systems and AI systems which are designed to be used by a judicial authority or on their behalf to interpret the law or facts to apply the law to a set of facts.[219] By way of a delegated act, the Commission would have the authority to update the eight areas identified in Annex III as necessary.[220]

All high-risk AI systems would be subjected to a different set of rules, which would entail an obligatory ex-ante conformity assessment. The providers of these systems will be required to register their systems in an EU-wide database before they are on the market or put into service. AI products and services which are governed by current product safety legislation will be considered to fall under the existing third-party conformity frameworks (such as medical devices). The providers of AI systems that are not governed by any EU legislation will have to conduct their own self-assessment to show they comply with the requirements for high-risk AI systems. By doing so they can use CE marking.[221] AI systems that are intended for biometric identification would however be obliged to conduct a conformity assessment by a notified body.[222] Furthermore, high-risk systems would also have to comply with several requirements particularly on risk management, technical robustness, transparency and human oversight (Articles 8-15 of the proposal). If a provider is established outside the Union, they will have to assign an authorised representative in the EU, ensuring that the conformity assessment is done.[223] Moreover, the provider will in all cases have to establish a post-market monitoring system and take corrective actions as necessary.[224]

[219] European Council, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach' (2022) 14954/22, 200-201.

[220] Madiega, 'Briefing EU Legislation Artificial Intelligence Act' (n 160) 6.

[221] ibid.

[222] ibid.

[223] COM(2021) 206 final, art 39.

[224] ibid art 61.

Lastly, AI systems which present a 'limited risk', for instance AI systems that interact with humans (such as chatbots), and systems that manipulate or generate, video, audio, or image content, so called deepfakes, would be subject to a limited amount of transparency obligations found in Title IV of the proposal.[225] AI systems which present a low or minimal risk are not obliged to conform to any additional legal requirements. However, it is encouraged that the providers of non-high-risk AI systems voluntarily apply the compulsory requirements for high-risk systems.[226] Article 69 sets forth codes of conduct for voluntary application of certain measures.

3.5 The purpose and the objectives of the proposal

The purpose of the proposed Regulation is to 'improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, marketing and use of artificial intelligence in conformity with Union values'.[227] As the development and use of AI often takes place across national borders, differing national regulation of AI would be an issue as they would fragment the internal market and reduce legal certainty. There is therefore a need for a consistent and high level of protection, hence why the Commission's proposal takes the form of a regulation which will be binding and directly applicable in all EU Member States.[228]

The general objective of the proposed AI Act is to guarantee 'the proper functioning of the single market by creating the conditions for the development and use of trustworthy artificial intelligence in the Union'.[229] Moreover, the proposed Regulation is a legal instrument that is intended to give a high level of protection of fundamental rights and public interests.[230] The primary objective of the proposal is to assure 'the proper functioning of the internal market by setting harmonised rules in particular on the development, placing on the Union market and the use of products and

[225] Madiega, 'Briefing EU Legislation Artificial Intelligence Act' (n 160) 6.
[226] ibid 7.
[227] COM(2021) 206 final, recital 1.
[228] COM(2021) 206 final, recital 2. Also see art. 288 TFEU regarding regulations being binding and having direct effect in all Member States.
[229] COM(2021) 206 final, 99.
[230] ibid recital 1, 7.

services making use of AI technologies or provided as stand-alone AI systems'.[231] The proposal sets out to achieve a set of four specific objectives. The first is to ensure that AI systems which are placed and used in the EU market are safe and respect existing Union law, such as Union values and fundamental rights. The second is to guarantee legal certainty to provide greater opportunities of innovation and investment. Furthermore, the proposed Regulation sets to improve the control and effective compliance of Union law on fundamental rights and safety requirements applicable to AI systems. Lastly, the proposal aims to make the development of a single market for lawful, safe, and trustworthy AI systems easier and counter market fragmentation.[232]

Furthermore, one of the key ideas of the proposal is to introduce uniform and binding rules for all Member States, whilst also leaving room for national measures and actions. For this reason, the Regulation is supplemented by annexes which can be updated as and when necessary.[233]

3.6 Trustworthy AI

As mentioned in chapter 3.2, trustworthiness has been continuously referred to in the preparatory work of the AI Act. When humans can understand how AI systems function, their trust for the designers and developers of the systems significantly improve.[234] Users may opt out of the requirement for full and open access to the data set and underlying algorithms if they are given clear explanations of the system by a reliable and qualified entity or expert.[235] To avoid unwanted consequences AI systems and the human beings behind them must essentially be worthy of trust. [236] Furthermore, the Commission and the Parliament have emphasised the importance of trustworthiness within

[231] COM(2021) 206 final, 6.
[232] ibid 3.
[233] ibid 7.
[234] Donghee Shin and Yong Jin Park, 'Role of Fairness, Accountability, and Transparency in Algorithmic Affordance' (2019) 98 *Computers in Human Behavior*, 278 <https://www.sciencedirect.com/science/article/pii/S0747563219301591> accessed 17 April 2022.
[235] ibid.
[236] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 5.

the context of AI several times [237] and the proposed Regulation states that one of the purposes with the proposal is to implement an ecosystem of trust by proposing a legal framework for trustworthy AI.[238] In combination with trustworthiness the proposal emphasises the need to have human-centric rules, so that individuals can trust that AI is used in a way that is secure and compliant with the law, including the respect of fundamental rights.[239]

When addressing the issue of trustworthy AI, the proposal refers to the Commission supporting the key requirements set out in the HLEG's Guidelines for trustworthy AI.[240] As trustworthiness and trustworthy AI are not concepts defined in the proposal, the Commission points to the Guidelines as one of the preparatory works on which the proposed minimum requirements, such as transparency and human oversight, have been based on.[241] As mentioned in chapter 3.3, although the Guidelines are not binding, the results of the framework are given attention and recognition in the proposal.[242] The aim of the Guidelines is to 'provide guidance for AI applications in general, building a horizontal foundation to achieve trustworthy AI.'[243] Moreover, the purpose is to foster sustainable and responsible AI innovations across Europe. It is only possible to fully enjoy the benefits of AI systems when trustworthiness is ensured, and safeguards are implemented to protect against any potential risks.[244] Nonetheless, different opportunities and challenges arise from AI systems. AI music recommendation systems will for instance not raise the same ethical issues as AI systems suggesting critical medical assessment and treatments. Moreover, AI systems used in a business-to-consumer context will not give rise to the

[237] European Council, 'Special meeting of the European Council (1 and 2 October 2020) – Conclusions' (n 199) 6; European Parliament, 'Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies 2020/2012(INL)' (2020) OJ C 404.

[238] COM(2021) 206 final, 1.

[239] ibid.

[240] ibid 9.

[241] ibid 14.

[242] ibid 9.

[243] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 5.

[244] ibid 5.

same challenges as systems used in a business-to-business context and public-to-citizen relationships.[245]

The framework of the Guidelines is divided into three chapters which all concern the different aspects of trustworthy AI. The first chapter manages the foundations of trustworthy AI and focuses on four ethical principles which are based on fundamental rights, namely respect for human autonomy, prevention of harm, fairness and explicability.[246] The second chapter presents the realisation of trustworthy AI and puts forward 7 key requirements which are to be evaluated continuously throughout the AI system's life cycle, one of which is transparency.[247] Lastly, the Guidelines present a trustworthy AI assessment list which can be adjusted to the specific AI application.[248] The realisation of trustworthy AI is considered to be a continuous process.[249] Furthermore, the White Paper on AI also emphasises the objective of trustworthy AI and presents policy options to enable a trustworthy development of AI in Europe whilst respecting the rights and values of EU citizens.[250]

The concept of trustworthy AI is presented as a set of three components which should all be met throughout the entire life cycle of an AI system:

1. The AI system should be lawful, respecting all applicable laws and regulations.

2. The system should be ethical, respecting ethical values and principles.

[245] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 5-6.
[246] ibid 11-12.
[247] ibid 8.
[248] ibid.
[249] ibid 20.
[250] COM(2020) 65 final, 2.

3. Lastly, the AI system should be robust, both from a technical and social perspective, seeing as AI systems can cause unintentional harm despite initial good intentions.[251]

Important to notice is that each of the requirements above is necessary but not sufficient on its own. Hence, it is not possible to achieve trustworthy AI if the three requirements do not work in harmony and overlap in their application. The components may at times conflict with one another. In those cases, there is an individual and joint responsibility to work towards securing that all three components help to protect the trustworthiness of AI systems.[252]

As mentioned above, the Commission underlines the importance of trust for AI systems by referring to a unique 'ecosystem of trust' being created through the future regulatory framework.[253] However, trust must be viewed from a wider perspective. Trust in the advancement, deployment and use of AI systems does not only concern the inherent properties of the technology, but also the qualities of the socio-technical systems involving AI applications. Trustworthy AI requires a holistic and systemic approach, not only a trustworthiness of the AI system itself. Trustworthy AI is therefore to be understood as trustworthiness regarding all processes and actors which are part of the AI system's socio-technical context.[254]

3.6.1 Fundamental rights as a basis for trustworthy AI

The proposal addresses that the use of AI with its specific characteristics, such as opacity and complexity, can unfavourably affect various fundamental rights which are enshrined in the CFR. The AI Act seeks to ensure a high level of protection for those fundamental rights and aims to, by using a clearly defined risk-based approach, address numerous sources of risks.[255] Furthermore, it is set forth in the proposal that 'with a set of requirements for

[251] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 35.

[252] ibid 5.

[253] COM(2020) 65 final, 3.

[254] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 5.

[255] COM(2021) 206 final, 11.

trustworthy AI and proportionate obligations on all value chain participants, the proposal will enhance and promote the protection of the rights protected by the Charter'.[256] Moreover, high-risk AI systems are regulated through a set of horizontal obligatory requirements for trustworthy AI and conformity assessment procedures.[257] The proposal aims to impose clear, predictable and proportionate obligations on providers and users to ensure that protection of fundamental rights is respected throughout the whole lifecycle of AI systems.[258] Thus, there is a clear connection between trustworthy AI and human rights in the AI Act, as it is expressly stated in the proposal that its purpose is to ensure human rights through trustworthy AI.[259]

As presented in chapter 2.4, human dignity is at the centre of the CFR and other important human rights instruments within the Union. Human dignity is also mentioned in the proposed Regulation and is one of the rights which the proposal especially aims to promote and enhance the protection of.[260] The proposal aims to protect the respect for private life and protection of personal data, which can be found in Article 7 and 8 of the CFR. Non-discrimination in Article 21 of the CFR and equality between women and men in Article 23 are also fundamental rights which are especially important to protect and promote in the context of AI. Human oversight is seen as a tool that helps to facilitate the respect of other fundamental rights by minimising the risk of biased or erroneous AI-assisted decisions in critical areas such as law enforcement, education and the judiciary.[261] Moreover, when designating an AI system as high-risk, the extent of the system's detrimental effects on the fundamental rights protected by the Charter is particularly important. The proposal includes numerous rights which fall under this importance, including the right to human dignity, respect for private and family life, protection of personal data and non-discrimination.[262] Additionally, the

[256] COM(2021) 206 final, 11.
[257] ibid 3.
[258] ibid.
[259] ibid 1.
[260] ibid 11.
[261] ibid 12.
[262] ibid 25.

proposal stresses the importance of highlighting the specific rights of children, as enshrined in Article 24 of the CFR.[263]

In its joint opinion to the AI Act the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) underline that the right to the protection of personal data as well as the right to private life form the basis of EU values, which is not only recognized in the Charter but also in Article 12 of the UDHR.[264] Other instruments also address the risk of AI affecting fundamental rights and how crucial these rights are in the context of AI. In its briefing the European Parliamentary Research Service (EPRS) brings forward the concern of AI jeopardising fundamental rights, such as freedom of expression, personal data protection, the right to non-discrimination and human dignity.[265] Moreover, the White Paper emphasises the importance of having European AI be based on EU values and fundamental rights.[266] The HLEG deems certain fundamental rights to be particularly apt to cover AI systems. The respect for human dignity, freedom of the individual, respect for democracy, justice and rule of law, equality, non-discrimination, and solidarity as well as citizens' rights are rights which in specified circumstances are legally enforceable within the EU.[267] However, as the use of AI systems may affect and implicate fundamental rights and their underlying values, ethical reflection can help identify what we *should* do with technology, rather than what we currently *can* do with it.[268]

Furthermore, apart from the comprehensive set of rights established in the CFR, ECHR and other instruments presented in chapter 2.4, documents drafted and documented by transnational governmental organs such as AI for Europe by the Commission as well as the Ethical Guidelines, include human

[263] COM(2021) 206 final, 25.

[264] EDPB and EDPS, 'Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' (n 132) 23.

[265] Madiega, 'Briefing EU Legislation Artificial Intelligence Act' (n 160) 2.

[266] COM(2020) 65 final, 3.

[267] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 10-11.

[268] ibid.

rights and apply a human rights framework.[269] The attention and necessity of human rights in the context of AI is also confirmed in a report published by Berkman Klein Center for Internet & Society which found that among 36 sets of AI principles, published by public and private agencies, human rights are particularly emphasised.[270] Moreover, three out of the five civil society-drafted AI principles documents explicitly adopt a human rights framework, and the vast majority of collected data and documents included references to human rights. [271] Not only is there an explicit need of a human rights perspective in the context of trustworthy AI, but scholars also believe that human rights can serve as a language that facilitates a deeper collaboration between civil society groups, AI researchers and the individuals affected and impacted by AI systems. [272]

If violations of fundamental rights still occur despite the efforts made in the proposal, the proposal sets forth that transparency and traceability of AI systems in combination with strong ex post controls will led to effective redress for affected individuals.[273] Although the Commission will be responsible for monitoring the effects of the proposal, AI providers are obliged to inform national competent authorities about malfunctioning or serious incidents that constitute a violation of fundamental rights obligations. This is to be done as soon as the providers become aware of these violations and will later lead to the necessary information being transmitted to the Commission.[274]

Furthermore, as the field of AI is continuously growing, social topics such as Fairness, Accountability, Transparency and Ethics in AI (FATE) is also becoming increasingly important.[275] Scholars and human rights advocates

[269] Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI' (n 12) 8-9.
[270] ibid 6.
[271] ibid 8-9.
[272] ibid.
[273] COM(2021) 206 final, 12.
[274] ibid 11-12.
[275] Shin and Park, 'Role of Fairness, Accountability, and Transparency in Algorithmic Affordance' (n 234) 277.

have continuously emphasised that rights cannot merely exist, individuals must be informed about their rights so that they can exercise them and in fact have significant bearing.[276] For instance, the ACM Conference on Fairness, Accountability, and Transparency (FAccT) holds annual conferences on the subject of FATE in order to discuss what role human rights play in the field of AI, ensuring these are not being overlooked.[277] Some scholars argue that the doctrine of human rights can help clarify what normative value systems shape a certain AI system as well as identify AI-related harms.[278] The role of human rights as a legal framework has implicitly formed numerous accountability initiatives within civil society bodies as well as within the AI industry.[279]

3.7 Conclusion

The Commission's extensive work behind the proposal can be traced back to 2015. The work has involved numerous actors and has resulted in the final draft being finished in December 2022. In March 2023 a final definition of AI was adopted by Parliament. The definition has been influenced by the definition used by the OECD.

The general objective of the proposed AI Act is 'to ensure the proper functioning of the single market by creating the conditions for the development and use of trustworthy artificial intelligence in the Union'.[280] Moreover, trustworthy AI and fundamental rights permeate the purpose and the objectives of the proposal. One of the specific objectives with the proposed Regulation is to ensure that AI systems respect existing Union law, such as Union values and fundamental rights.[281] Although the proposal lacks a definition of trustworthy AI, the Guidelines provide three components which should all be met throughout the life cycle of an AI system, one being that a system should be ethical. Trustworthy AI is a key element in the

[276] Prabhakaran and others, 'A Human Rights-Based Approach to Responsible AI' (n 13) 3.
[277] ibid.
[278] ibid 2.
[279] ibid.
[280] COM(2021) 206 final, 91.
[281] ibid 91-92.

Regulation, and it is seen as a set of requirements which ensure that fundamental rights are being respected. The AI Act explicitly states that its purpose is to achieve trustworthy AI by ensuring that the AI developed, deployed, and used respects fundamental rights. Trustworthy AI is essentially AI that respects fundamental rights. Thus, there is a clear commitment and connection between trustworthy AI and complying with fundamental rights. As described in chapter 2.4, EU measures protect EU fundamental rights, but they must also comply with the right to justification, seeing as this moral right is the normative foundation of national human rights. Thus, the right to justification is relevant in the context of trustworthy AI seeing as the rights in the proposal are rights which cannot be unjustifiably denied to the individual.

Furthermore, the proposal does address the use of AI with its specific characteristics, such as opacity as well as the risks AI imposes on the fundamental rights in the CFR. By targeting providers and users and imposing a risk-based approach the Commission addresses several sources of risks, ensuring a high level of protection for fundamental rights enshrined in the Charter.[282] Moreover, when an AI system is classified as high-risk, the extent of the system's detrimental effects on the fundamental rights protected by the CFR is of particular importance. [283]

Just as human dignity is the centerpiece of the CFR, human dignity is also promoted in the proposal as one of the rights that the Regulation aims to enhance the protection of.[284] Other fundamental rights that can be found emphasised in the AI Act is the right to private life and protection of personal data, rights which can be found in Articles 7 and 8 of the CFR. Furthermore, human oversight is seen as a tool to minimise the risk or bias of AI-assisted erroneous decisions in critical areas.[285] The point of protecting fundamental rights in EU law is, as mentioned in chapter 2.4, the need to maintain the unity, primacy and effectiveness of Union law.[286] Hence, the same type of

[282] COM(2021) 206 final, 11.
[283] ibid 25.
[284] ibid 11.
[285] ibid 12.
[286] Gill-Pedro, *EU law, fundamental Rights and National Democracy* (n 106) 115.

rights are reoccurring in the majority of the soft law instruments as well as in the proposed Regulation, ensuring that there is a uniform and consistent effectiveness of promoting and enforcing the rights enshrined in the Charter. The proposal imposes a standard of human rights protection, in this case trustworthy AI, which will ensure that AI systems are to be trusted in relation to fundamental rights. Trustworthy AI does not only protect the fundamental rights in the CFR, but also ensures that the right to justification is respected as this, again, is a moral right that underlies the human rights in the Member States. Moreover, according to the Guidelines trustworthy AI ought to be understood as trustworthiness regarding all processes and actors which are part of the AI system's socio-technical context, and not only the trustworthiness of the AI system itself.[287]

[287] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 5.

4 Transparency in the context of AI

4.1 About the chapter

As we have seen in the previous chapters fundamental rights and trustworthy AI connect with one another. By using a risk-based approach and emphasising the importance of trustworthy AI the Commission displays their commitment to human rights, a commitment that permeates the objectives and purpose of the proposal. In the following chapter transparency in the context of AI will be presented as well as the provisions in the proposal which regulate interpretability and explainability. These are neighbouring concepts of transparency[288] and central to the research question. Furthermore, it will be shown how these two fundamental concepts have been discussed in literature and in the field of AI. Lastly, the chapter is concluded with a conclusion that connects the measures in the proposal which aim to safeguard transparency through interpretability and explainability with the right to justification.

4.2 Transparency as a concept

Trust and its links to transparency has been studied in several social-scientific disciplines, including law, over a long period of time.[289] Moreover, transparency plays a crucial role in the general aim to develop more trustworthy AI[290] and is a multifaceted concept used by various disciplines.[291] The concept has recently gone through a resurgence due to the contemporary discourses concerning AI.[292]

Since the beginning of the policy process to regulate AI, all relevant documents, such as the Guidelines, the White Paper on AI and the

[288] Stefan Larsson, 'Transparency in artificial intelligence' (2020) 9(2) *Internet Policy Review*, 6 <https://policyreview.info/pdf/policyreview-2020-2-1469.pdf> accessed 27 March 2023.
[289] ibid 7.
[290] ibid 9.
[291] Helen Margetts, 'The Internet and Transparency' (2011) 82(4) *The Political Quarterly*, 1 < https://www.dhi.ac.uk/san/waysofbeing/data/citizenship-robson-margetts-2011b.pdf> accessed 29 March 2023.
[292] ibid.

Parliament's Report on AI Framework, have included transparency in their suggested legal or ethical framework. The Guidelines describe transparency as one of seven key requirements for the realisation of trustworthy AI[293] and is considered to be a key criteria in the proposal, as it focuses on the necessity of AI systems being transparent in the data, system, and business model.[294] However, while several requirements can already be found in existing regulatory or legal regimes, those concerning transparency, human oversight and traceability are not specifically covered under current legislation in many economic sectors.[295] According to the Commission transparency is viewed from three different aspects: traceability, explainability and communication.[296]

The proposed Regulation follows the direction towards transparency and explicitly devotes several provisions to the transparency of AI. Although the definition or degree of the concept is not explained in the proposal, the Commission explicitly lays down harmonised transparency rules for AI systems.[297] Transparency is also continuously mentioned in the preamble when referring to various AI systems and the need to impose transparency obligations for certain systems. [298] For high-risk systems, such as those used for law enforcement and migration, transparency is considered to be particularly important for the safeguarding of the fundamental rights of those affected, seeing as these forms of systems can cause adverse impacts.[299] Thus, the Commission sets forth that transparency is particularly essential for high-risk systems and states that a certain degree of transparency should be required for high-risk systems due to the opacity that makes certain AI systems incomprehensible for natural persons.[300] Moreover, transparency is seen as particularly important to certain AI systems. Title IV of the proposal concerns AI systems which pose specific risks of manipulation. These special

[293] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 14.
[294] COM(2019) 168 final, 5.
[295] COM(2020) 65 final, 9.
[296] COM(2019) 168 final, 5.
[297] COM(2021) 206 final, art 1(c).
[298] ibid recital 14.
[299] ibid recital 38 and 39.
[300] ibid recital 47.

transparency obligations apply for systems that generate or manipulate content (so called 'deep fakes'), interact with humans and systems that use biometric data to detect emotions or determine association with social categories. The reason why transparency is particularly important in these cases is due to character of the systems. By being informed of the circumstance, for instance interacting with an AI system, the individual is free to make informed choices or turn away from a given situation.[301]

As algorithms evolve and become more complex, the problem with opacity is of particular importance, especially when dealing with decision-making AI systems.[302] Transparency is a centrepiece of the Regulation and although the definition and degree of transparency required is yet to be defined in the proposal, the Union shows a clear commitment to the requirement of transparency. The difficulty to define and use transparency as a concept, relates to the fact that different fields denote the concept as the physical property of a material,[303] while others think of the concept as 'powerful means towards some desirable social end, for example, holding public officials accountable, reducing fraud and fighting corruption'.[304] Transparency serves as a tool to create trust and legitimacy and enables the realisation of other fundamental rights. Through this, individuals' capacity to make the right decision improves as it allows them to see what is taking place.[305] This type of clarity is a condition for accountability and ensures that those who make the decisions, in this case AI systems, do what they are supposed to do, namely act in the interest of people while simultaneously complying with the rule of law and the principles of democracy.[306]

---

[301] COM(2021) 206 final, 15.

[302] Fotios Fitsilis, *Imposing Regulation on Advanced Algorithms* (Springerbriefs in law 2019) 4.

[303] Larsson, 'Transparency in Artificial Intelligence' (n 288) 5.

[304] Hans Krause Hansen, Lars Thøger Christensen and Mikkel Flyverbom, 'Introduction: Logics of Transparency in Late Modernity: Paradoxes, mediation and governance' (2015) 18(2) *European Journal of Social Theory*, 118 <https://journals.sagepub.com/doi/10.1177/1368431014555254> accessed 27 March 2023.

[305] Anoeska Buijze, 'The Principle of Transparency in EU Law' (DPhil thesis, Utrecht University 2013) 62 <https://dspace.library.uu.nl/bitstream/handle/1874/269787/buijze%2Bapp.pdf?sequence=4> accessed 27 March 2023.

[306] ibid.

The theoretical backdrop of transparency is evidently vast and complex. When addressing transparency in the context of AI, literature often refers to explainability, interpretability as well as trust.[307] The report published by Berkman Klein Center for Internet & Society, mentioned in chapter 3.5.1, shows that transparency and explainability are closely linked principles and that these principles are some of the most frequently occurring individual principles, each mentioned in approximately three-quarters of the 36 documents analysed.[308] Transparency in the context of AI takes a system's perspective rather than only focusing on the components used or individual algorithms.[309] Due to this it is a less ambiguously broad term than algorithmic transparency.[310] Understanding transparency as an applied concept in the context of AI requires that it is understood in context, mitigated by literacies, explainability as well as a set of competing interests. Consequently, transparency in AI can best be viewed as 'a balancing of interests and a governance challenge demanding multidisciplinary development to be adequately addressed'.[311] Several scholars within the field of transparency support the notion of a wider transparency concept by holding systems accountable by looking across them, instead of privileging a form of accountability that needs to look *inside* the systems. By having this perspective of transparency, the AI systems are seen as systems that enact complexity by connecting to and intertwining with groups of humans and non-humans and not as systems that contain complexity. Transparency structures our thinking. Moreover, how the AI systems are understood has normative effects on the regulatory debates concerning how to regulate AI.[312]

[307] Larsson, 'Transparency in Artificial Intelligence' (n 288) 7.

[308] Fjeld and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI' (n 12) 41.

[309] Larsson, 'Transparency in Artificial Intelligence' (n 288) 10.

[310] ibid.

[311] ibid.

[312] ibid 6.

4.3 Transparency, interpretability and explainability

Interpretability and explainability are neighbouring concepts to transparency.[313] They promote trust and understanding in the same way as transparency does.[314] Although the proposed AI Act does not provide a definition of transparency or interpretability and explainability it does however articulate transparency requirements in multiple forms, notably under Articles 13, 14 and 52. Article 13 of the proposal sets out the criteria of interpretability for high-risk AI systems. The Article establishes:

> 1. High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured, with a view to achieving compliance with the relevant obligations of the user and of the provider set out in Chapter 3 of this Title.
>
> 2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users.

The Article enables AI providers to define the relevant degree and type of transparency with the purpose to achieve compliance with their and users' relevant obligations. The Article is applicable to a substantial part of AI systems, specifically those who present potential high risks to the fundamental rights of individuals.

Furthermore, Article 14 regulates human oversight:

> 1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that

[313] Larsson, 'Transparency in Artificial Intelligence' (n 288) 6.

[314] Kacper Sokol and Peter Flach, 'Explainability Is in the Mind of the Beholder: Establishing the Foundations of Explainable Artificial Intelligence' (2022) 2 <https://ui.adsabs.harvard.edu/abs/2021arXiv211214466S> accessed 1 April 2023.

they can be effectively overseen by natural persons during the period in which the AI system is in use.

2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.

The Article mandates that high-risk AI systems are designed with suitable human-machine tools so that humans, by overseeing the systems, can understand the full capacities and limitations of the system. Moreover, the Article highlights the necessity of a human overseeing the system so that safety and fundamental risks are not threatened.

Article 52 of the proposal specifies:

1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.

Articles 13 and 52 have similar names but imply different concepts. Article 13 focuses on high-risk AI systems and refers transparency to interpretability. The interactive AI systems regulated in Article 52 refers to a different aspect of transparency: communication regarding the presence of AI. Humans have the right to be informed that they are interacting with an AI system, which means that AI systems should not be represented as humans to those using the system. AI systems must be identifiable as AI systems and the limitations and capabilities of the systems should also be communicated to the users in a manner which is appropriate to the description of how users interact with a

system, a so-called use case.[315] Moreover, the option to decide against the interaction with an AI system in favour of human interaction should be provided as this would ensure that fundamental rights are being complied with.[316]

The provisions emphasise the necessity of different forms of transparency, something which is further highlighted in the recitals which also affirm the importance of interpretability and explainability. Recital 38 calls for explainable systems and underlines that AI systems which are not sufficiently transparent, documented, and explainable could impose on important procedural fundamental rights, such as the right to a fair trial.[317] Moreover, recital 47 of the proposal addresses the risk with opacity and underlines the necessity of interpretability by establishing that high-risk AI systems must be transparent to a certain degree, in order to ensure the opacity that might make certain systems too complex or incomprehensible for natural persons. A way of ensuring this interpretability is by having relevant documentation and instructions of use accompany high-risk AI systems. The instructions should be clear and concise as well as address possible risks to fundamental rights if needed.[318]

Similar to Article 14, recital 48 of the proposal emphasises the need for human oversight of high-risk AI systems. This type of AI 'should be designed and developed in such a way that natural persons can oversee their functioning'.[319] Furthermore, it is important to note that heavier regulatory obligations in the proposal apply in cases where the transparency and interpretability is at a higher risk.[320]

[315] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 18.
[316] ibid.
[317] COM(2021) 206 final, recital 38.
[318] ibid recital 47.
[319] ibid recital 48.
[320] Bordt and others, 'Post-Hoc Explanations Fail to Achieve their Purpose in Adversial Contexts' (n 17) 4.

4.4 What is interpretability?

When addressing transparency in the context of AI, literature often refers to explainability, interpretability as well as trust in the AI systems.[321] In the context of AI the term interpretability is reoccurring and defined as 'the ability to explain or to provide the meaning in understandable terms to a human'.[322]

The proposal does not, as presented in chapter 4.3, define the term despite the Act including several points which emphasise the importance of interpretability in the context of transparency. Although the preamble states that users should be able to interpret the output of a high-risk AI system and use it appropriately it does not explain what is meant with the term.[323] As the proposal does not include any definitions or explanations of the term, one must turn to soft law instruments. The HLEG does not either provide a clear definition of interpretability in the Guidelines, originally only classifying the term as a sub-component of explainability in its assessment list. The questions in the Guidelines regarding interpretability merely focus on if the design of the AI system was made with interpretability in mind, if the most interpretable model has been used, if there is a possibility to assess and analyse the training and testing data and if the interpretability can be assessed after the AI model has been developed.[324] However, in July 2020, the HLEG presented their final version of the Assessment list for Trustworthy AI (ALTAI). Following the first draft, over 350 stakeholders participated and contributed with feedback regarding the assessment list.[325] In comparison to the lack of a definition for interpretability in the first draft of the assessment list, the revised ALTAI introduces a definition of the term in its glossary.[326] The term according to

[321] Larsson, 'Transparency in Artificial Intelligence' (n 288) 7.

[322] Riccardo Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (2018) 51 *ACM Computing Surveys*, 5 <https://dl.acm.org/citation.cfm?doid=3271482.3236009> accessed 16 April 2023.

[323] COM(2021) 206 final, recital 47.

[324] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 29.

[325] AI HLEG, 'The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self Assessment' (2022) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342> accessed 7 March 2023.

[326] ibid 27.

the ALTAI is the concept of explainability, understandability or comprehensibility.[327] An AI system, or an element of such a system, is interpretable when 'it is possible at least for an external observer to understand it and find its meaning'.[328]

The concept of interpretability is evidently vast and complex, leading scholars to establish and define several dimensions of interpretability.[329] The first dimension is global and local interpretability. Global interpretability entails a model which is completely interpretable, meaning natural persons can understand the whole logic of the model and thus follow the entire reasoning behind the various possible results. Local interpretability includes a situation where the single decision is interpretable, meaning it is only possible to understand the reasons behind a specific decision or prediction.[330]Another important dimension of interpretability is the time limitation. This entails the time that the user is allowed to spend on understanding the explanation. The time availability for the user is strictly related to the context where the system must be used. For instance, in situations where the user needs to make a fast decision, due to an imminent disaster, it is preferable to have an explanation that is simple to understand. In other contexts, where the time of the decision is not a constraint, for example during a loan process, the user could prefer a more exhaustive and complex explanation.[331] Lastly, another essential dimension is the nature of the user expertise. Not every user has the same background knowledge and experience in the task. By acknowledging the user experience, the interpretability can also vary seeing as domain experts for instance may prefer a more sophisticated model over a smaller, and at times, opaquer one.[332]

[327] AI HLEG, 'The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self Assessment' (2022) 27 <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342> accessed 7 March 2023.

[328] ibid.

[329] Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (n 322) 6.

[330] ibid.

[331] ibid

[332] ibid.

An important aspect to note is that not all systems need to be interpretable. For instance, if we want to know if an image displays a dog or not and this information is not necessary for taking any sort of crucial decision or there are no consequences for unacceptable outcomes, then there is not a need for an interpretable model. Hence, in certain situations we can accept opaqueness, in the form of a black box. However, many times it is necessary to have systems that are interpretable. [333]

4.5 What is explainability?

The necessity of AI systems being explainable has and is being emphasised in AI research as well as in policy discussion.[334] In its White Paper the Commission underlines the importance of addressing current framework due to the specific characteristics of many AI systems, such as opacity ('the black box effect'), unpredictability and complexity. These characteristics increase the risks of non-compliance with existing EU law aimed to protect fundamental rights.[335] Moreover, the importance of research into the explainability of AI systems is also emphasised in the work of the Commission.[336] According to the HLEG, the principle of explicability is expressed through the demand for transparency.[337]

Explainability concerns the technical processes of AI systems as well as the human decisions related to the process, such as application areas of an AI system. Technical explainability requires that human beings can trace and understand the decisions made by an AI system.[338] Additionally, trade-offs may have to be made between enhancing an AI system's accuracy, which could reduce its explainability, or increasing its explainability which could have a negative impact on the accuracy.[339] By documenting the data sets and the processes that generate an AI system's decision, traceability facilitates

[333] Guidotti and others, 'A Survey of Methods for Explaining Black Box Models' (n 322) 6.

[334] Larsson, 'Transparency in Artificial Intelligence' (n 288) 28.

[335] COM(2020) 65 final, 12.

[336] COM(2018) 237 final.

[337] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 18.

[338] ibid.

[339] ibid.

explainability. Consequently, an increase in transparency is ensured. The traceability does not only concern the algorithms used but also the data gathering and data labelling. Decisions generated by the AI system must also be traceable and transparent seeing as this helps identifying why the decision was erroneous which could in turn help prevent future errors and mistakes.[340]

Being able to explain a decision that an AI system makes is crucial for building and upholding users' trust in AI systems. The processes and purpose of AI systems must therefore be transparent and openly communicated.[341] Demanding a suitable explanation of a specific AI system's decision-making process should be deemed as feasible. The explanation should be timely and adapted to the concerned stakeholder, for instance a researcher or regulator.[342] In various cases the whole decision-making process cannot be explained due to the black box issue. However every decision must be explainable to the extent possible to those who are directly and indirectly affected.[343] In situations where there is a lack of such essential information, decisions cannot be duly contested.[344] The Parliament has specified in Article 8(e) of the 'Framework of ethical aspects of artificial intelligence, robotics and related technologies' that AI systems are 'required to be developed, deployed and used in an easily explainable manner so as to ensure that there can be a review of the technical processes of the technologies'.[345] In the Guidelines explainability is found under the principle of ethical AI. AI systems should improve individuals' as well as society's wellbeing. In order to do so in a trustworthy manner, the HLEG has put forward four ethical imperatives which AI practitioners should aim to adhere to, one being explainability. Although many of these principles may fall within the scope of lawful AI and are already reflected in existing legal requirement, for which mandatory compliance is required, adherence to ethical principles goes beyond formal

[340] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 18.

[341] ibid 13.

[342] ibid 18.

[343] ibid 13.

[344] ibid.

[345] European Parliament, 'Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies 2020/2012(INL)' (2020) OJ C 404.

conformity with existing law.[346] The degree of explainability is highly dependent on the context and the severity of the potential consequences AI systems can cause.[347] Thus, AI systems must be approached in the context in which they are being used. For instance, inaccurate shopping recommendations generated by AI systems will not raise the same ethical concerns as AI systems evaluating whether a convicted person should be released on parole.[348] As mentioned in chapter 4.3, this is also reflected in the proposal by having heavier regulatory obligations apply in cases where the system is classified as high-risk.[349]

To implement the requirements of explainability non-technical and technical methods can be used. Non-technical methods serve to maintain trustworthy AI and should be assessed continuously. One example of such a method are codes of conduct, which ensures that organisations, when working with or on an AI system, document their intentions and secure them with standards of specific desirable values, such as fundamental rights.[350] Technical methods entail methods which ensure trustworthiness through the design, development and use phases of an AI system.[351] One way of ensuring explainability is by using technical explanation methods.[352] Explainable AI (XAI), which is a relatively new field within AI, focuses on the issue of explainability and tries to enable an easier understanding of the underlying mechanisms in AI systems.[353] XAI is described to handle the black box models. The notion of transparency is in this research filed narrower with a bigger focus on algorithmic models than for instance the notion of necessary transparency and explainability that is set forth by the HLEG.[354] Furthermore, research on XAI does not typically build on explanatory frameworks based on social science

[346] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 12.
[347] ibid 13.
[348] ibid.
[349] Bordt and others, 'Post-Hoc Explanations Fail to Achieve their Purpose in Adversial Contexts' (n 17) 4.
[350] European Commission, 'High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI' (n 18) 22.
[351] ibid 21.
[352] ibid.
[353] Larsson, 'Transparency in Artificial Intelligence' (n 288) 7.
[354] ibid.

and scholars have argued more could be done concerning this field of research.[355]

Furthermore, on the technical side explainability as a concept has progressed into its own field of research as mentioned above.[356] The current literature puts forward two different approaches towards explainability.[357] The first approach is to use a model, which could be a black box, and apply a different approach to explain the behaviour and decisions of the black box after the decisions have been made. This is referred to as 'post-hoc' explanations.[358] The second approach is to build models and systems which are obliged to be 'inherently interpretable'.[359] This approach questions XAI, namely 'post hoc' explanations, where a second model is used to explain the decision of the first black box model. By using the latter approach, where the system is inherently interpretable, systems instead provide their own explanations which is in line with what the system de facto calculates.[360]

4.6 The Black Box

As mentioned above, an explanation to why an AI model has generated a particular decision or output, is not always possible. This type of concern is attributable to the black box, which requires special attention, particularly since the concept of the black box is a neighbouring concept to transparency.[361] The term 'black box' is used to describe a system 'whose internal workings are opaque to the observer – its operation may only be traced by analysing its inputs and outputs'.[362] In the context of AI a black box is a data-driven algorithm which entails an automated process where humans only can observe the systems behaviour. Rudin points out two main sources

[355] Larsson, 'Transparency in Artificial Intelligence' (n 288) 7.
[356] Bordt and others, 'Post-Hoc Explanations Fail to Achieve their Purpose in Adversial Contexts' (n 17) 1.
[357] ibid.
[358] ibid.
[359] Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'. (2019) 1 <https://arxiv.org/abs/1811.10154> accessed 18 January 2023.
[360] ibid.
[361] Larsson, 'Transparency in Artificial Intelligence' (n 288) 6.
[362] Sokol and Flach, 'Explainability Is in the Mind of the Beholder: Establishing the Foundations of Explainable Artificial Intelligence' (n 314) 4.

of opaqueness for AI: the first being a proprietary system that operates as a black box but is transparent to its creators and the second a system that is too complex for any human to comprehend.[363] The latter establishes a continuous spectrum of understanding.[364]

4.7 Conclusion

Transparency has been studied over a long period of time and has in the context of AI been presented as crucial for the protection of fundamental rights. Moreover, the concept of transparency plays an essential role in the general aim to develop more trustworthy AI.[365] In the proposal Articles 13, 14 and 52 all focus on some aspect of transparency and the preamble consists of several points that underline the need for transparency. The Regulation also emphasises that the transparency requirement particularly applies to high-risk systems, as these forms of systems can cause serious impacts on the fundamental rights of those affected.[366] Some high-risk AI systems, for instance those who interact with humans, are considered to pose specific risks, leading to an increased need for transparency. By being informed of the circumstance, for instance interacting with an AI system, the individual can make informed choices or turn away from the situation.[367]

Furthermore, interpretability and explainability have shown to be neighbouring concepts to transparency.[368] Although the proposed AI Act does not provide a definition of transparency, interpretability or explainability, it does however articulate transparency requirements in multiple forms. Article 13 focuses on interpretability and implies that transparency is necessary for interpretability. The Article also implies that there are various forms and degrees of transparency depending on the AI system in question. Article 14 regulates human oversight, implying that there is a need for a human to

[363] Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (n 359) 2.

[364] Sokol and Flach, 'Explainability Is in the Mind of the Beholder: Establishing the Foundations of Explainable Artificial Intelligence' (n 314) 4.

[365] Larsson, 'Transparency in Artificial Intelligence' (n 288) 9.

[366] COM(2021) 206 final, recital 38 and 39.

[367] ibid 15.

[368] Larsson, 'Transparency in Artificial Intelligence' (n 288) 6.

oversee the system so that fundamental rights are not at risk. Lastly, Article 52 regulates the need for communication regarding the presence of AI. The legislative language jumps across several meanings of transparency, leading to a lack of clarity in relation to what the provisions de facto mean. Interpretability is merely mentioned in Article 13 and is also only limited to high-risk AI systems. The proposal does not explain how the two transparency rules in Articles 13 and 52 relate to one another.

Important to note is that depending on the system and its outcome opaque systems could be accepted. However, many times an interpretable AI system is fundamental for the aim to ensure trustworthy AI and thus also protect fundamental rights. There are several aspects of interpretability which have been discussed in literature but have not been presented in the proposal or in other relevant documents. What interpretability de facto entails can merely be found in recital 47 of the proposal which establishes that AI systems must be transparent to a certain degree. The recital specifies that one way of ensuring interpretability is by having relevant, clear and concise documentation and instructions of use accompany high-risk AI systems. The information should also address possible risks to fundamental rights if needed.

Transparency in AI can best be viewed as 'a balancing of interests and a governance challenge demanding multidisciplinary development to be adequately addressed'.[369] However, the main point of transparency in the proposal is not to define the degree of transparency. Through imposing measures on users and providers in the proposal the Union displays a commitment to make AI systems more transparent. As seen in previous chapters, the provisions of the proposed Regulation must comply with fundamental rights, something which is explicitly stated in the proposal. By imposing duties on users and providers to make systems explainable and interpretable, and thus also more transparent, the Union is showing an effort to ensure that fundamental rights are respected. The requirement of explainability could be seen in the same way as Forst's right to

[369] Larsson, 'Transparency in Artificial Intelligence' (n 288) 10.

justification, namely that human beings must be respected as agents who have the right to not be subjected to certain actions or norms that cannot be adequately justified to them.[370]

[370] Forst 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach' (n 32) 712.

5 Analysis and discussion

Human rights have moral, legal and political lives. They are rights that are meant to ensure that no individual is being treated in a way that could not be justified to them as a person equal to others. By grounding human rights on the right to justification, the political and social meaning of human rights is captured and in opposition to earlier and modern forms of social exclusion. The right to justification underlies all human rights and enables a power to demand justification and challenge false legitimations. It also allows individuals to be regarded as independent agents of justice and democracy, ensuring that their dignity and autonomy is not violated by merely viewing them as recipients of redistributive measures. Individuals who have the status of normative agency have a human right to particular forms of respect seeing as one cannot reasonably justify a denial of their basic claims. Respecting other individuals' human rights cannot be grounded on the view that the respect contributes to one own's good life or to the good life of the others. Essentially, structures and norms must be justified towards the individual as the right to justification is a fundamental moral and human right. If a restriction of a human right cannot be justifiable, then the limitation should not occur. Member States have obligations, under national human rights framework, to ensure that private organisations and businesses do not violate human rights. By harmonising these rights, the EU takes on the role of the state in ensuring respect for human rights, and thus also upholding the right to justification by imposing duties of transparency through the requirements of interpretability and explainability.

The question the thesis seeks to answer is whether the EU succeeds to uphold the right to justification by adequately safeguarding interpretability and explainability in the proposed AI Act. In order to answer this, one must initially highlight the connection between EU fundamental rights and national human rights. As established in the case of Siragusa, fundamental rights are part of Union law so that the unity, primacy and efficacy of Union law can be protected. As a result, the Union steps into the role of the guarantor of human

rights, not as national rights, but as EU fundamental rights. Article 6(3) of the TEU states that the fundamental rights as assured in the Convention and as they result from the constitutional traditions common to the Member States, shall constitute general principles of EU law. Through this provision, the EU displays that there is a clear commitment to respect fundamental rights in the way that it reflects national fundamental rights in the Member States. It can therefore be concluded that because national human rights are based on everyone's equal right to justification, the EU standard must also respect this. The moral standards of human rights that have been developed on a national level in Member States must also be reflected on an EU level. A just and democratic social order is an order where individuals have equal rights and where they can give their consent, not only their counterfactual consent but a consent which is based on institutionalized justification procedures. Thus, the proposed Regulation must ensure that the measures regarding interpretability and explainability respect the right to justification in an adequate manner.

Respect for fundamental rights is a central Union value, as shown in the Charter as well as in primary law. The EU continuously refers to this value in the proposed Regulation as well as its preparatory work. Moreover, fundamental rights in the proposed Regulation have a strong connection with the concept of trustworthy AI. The purpose and objectives of the proposed Regulation are permeated with trustworthiness. Trustworthy AI is seen as a set of requirements which ensure that fundamental rights are being respected. The AI Act explicitly states that its purpose is to achieve trustworthy AI by ensuring that the AI developed, deployed, and used respects fundamental rights. Trustworthy AI is essentially AI that respects these rights. Furthermore, trust and its connection to transparency has been studied over a long period of time in several social-scientific disciplines, including law. Interpretability and explainability are neighbouring concepts to transparency, thus, in order to establish how the prior concepts have been defined, one must analyse transparency and its meaning in the context of AI. The concept of transparency has played a fundamental role in developing more trustworthy AI and has since the beginning of the policy process been included in several

soft law instruments and their suggested ethical or legal framework. In the proposal transparency is particularly important for high-risk systems, due to the opacity that makes certain AI systems incomprehensible for natural persons. Additionally, certain systems require more transparency than others due to its character. In those cases, it is important to ensure that the individual can make an informed choice and decide freely if they want to engage with the AI system in question.

The AI Act devotes three provisions to the transparency of AI, namely Articles 13, 14 and 52. The requirement of transparency in the proposal ought to be viewed as a property that is promoted by the Union and built into the proposed Regulation, by imposing the requirements of interpretability and explainability. Although Article 13 enables AI providers to define the relevant degree and type of transparency with the purpose to achieve compliance with their and users' relevant obligations, the Article leaves open what interpretability de facto requires from a technical perspective. Moreover, the Article implies that transparency is necessary for interpretability and that there are different degrees and types of transparency, something that is not further explained or developed in the proposal, except for recital 47. Recital 47 does shed light on the issue of interpretability by emphasising having high-risk AI systems be accompanied with clear and concise instructions and documentation. However, the provision does not explicitly provide examples of types of transparency besides interpretability and does not make it clear for providers of AI systems whether ensuring interpretability is sufficient to adhere with the requirement of transparency. The revised ALTAI does provide somewhat of a definition of interpretability; however, it is not explained what is exactly meant with 'an external observer'. Observers and users do not all have the same level of knowledge or user expertise, something which must be acknowledged and not overlooked.

Another aspect of ambiguity with the interpretability transparency obligation enshrined in Article 13 is its correlation with explainability. Explainability is also expressed through the demand for transparency. While Article 13 focuses on the need for interpretability, recital 38 of the proposal

calls for explainable systems. This leads to a lack of clarity about what distinguishes the different concepts and what the EU aims; interpretability, explainability or both? If the Union does want AI systems to fulfil both, which the preparatory work implies, the degree and technical meaning of the concepts must be further explained. An unexplainable AI system does not only violate the requirement of transparency and trustworthy AI but also violates the moral right to justification.

As presented in chapter 2, Member States in the Union have obligations, under national human rights framework, to ensure that human rights are not being violated. Chapter 4 displays that the Union is imposing duties on users and providers to make AI systems explainable and interpretable as an effort to ensure that human rights are respected in the light of the right to justification. Human rights require us to have our right to justification respected. Thus, when opaque decisions are made about us, failing to fulfil the criteria of interpretability and explainability, the right to justification has not been respected either. By imposing duties of interpretability and explainability the right to justification is to be respected. However, to be able to fully safeguard interpretability and explainability and thus the right to justification, one must fully understand the meaning and extent of the concepts. EU law and non-binding acts have discussed the two concepts and have, to a certain extent, defined them. If the concepts themselves cannot be adequately explained in the proposal nor in soft law instruments, the legal certainty of human rights and the application of the proposed Regulation could be severely affected.

By not respecting the right to justification through sufficient measures in Articles 13,14 and 52 individuals are no longer regarded as independent agents of justice and democracy. It is only through ensuring the right to justification that individual's dignity and autonomy is not violated. Human dignity is also promoted in the proposal as one of the rights that the Regulation aims to enhance the protection of. According to Forst dignity means that 'a person is to be respected as someone who is worthy of being given adequate reasons for actions or norms that affect him or her in a relevant

way'. Thus, one could establish that human dignity and the right to justification have something in common: one must be given justifiable reasons for norms that has some form of effect on the individual. Consequently, if an AI system generates a decision or an outcome which cannot be adequately explained or interpreted, human dignity and the right to justification are violated. Human rights require us to have our right to justification respected and so when opaque and unexplainable decisions are made about us, our right to justification has not been respected either. The principle of justification enables a power to demand justification and challenge false legitimations. If interpretability and explainability are not ensured, the power and the right to demand justification and challenge decisions that an AI system makes is violated. Humans who are exposed to decisions made by AI which cannot be explained or interpreted are not either given the status of independent agents of justice, resulting in their dignity and autonomy being negatively affected. Seeing as dignity permeates several fundamental sources of human rights within the Union, this matter demands meticulous consideration and should not be underestimated. It is therefore of my opinion that the measures in the AI Act, which aim to safeguard transparency through interpretability and explainability, do not sufficiently uphold the right to justification. The Union is showing a clear commitment to fundamental rights by incorporating these measures in the proposal, however that does not mean that the measures are adequately defined or clear. Considering the AI Act could have a Brussels Effect, like the GDPR, the proposed Regulation must secure legal certainty. Providing a predictable and safe legal environment will ensure effective safeguards for the protection of fundamental rights and freedoms, something that the Union is clearly aiming to do but have yet to succeed with.

Bibliography

Literature

Books

Dauses M.A, *The Protection of Fundamental Rights in the Legal Order of the European Union* (1st edn, Frankfurt am Main: Lang 2014)

di Federico G, *The EU Charter of Fundamental Rights: from declaration to binding instrument* (1st edn, Springer 2011)

Fitsilis F, *Imposing Regulation on Advanced Algorithms* (Springerbriefs in law 2019)

Forst R, *Justice, democracy and the right to justification Rainer Forst in dialogue* (1st edn, Bloomsbury Academic 2014) <https://library.oapen.org/handle/20.500.12657/58767> accessed 20 January 2023

Gill-Pedro E, *EU law, fundamental Rights and National Democracy* (1st edn, Routledge 2018)

Hettne J and Otken Eriksson I, *EU-rättslig metod: teori och genomslag i svensk rättstillämpning* (2nd edn, Norstedts Juridik 2011)

Kleineman J, 'Rättsdogmatisk metod' in Maria Nääv and Mauro Zamboni (eds.), *Juridisk metodlära* (2nd edn, Studentlitteratur 2018)

Russell S.J, *Human compatible: Artificial Intelligence and the problem of Control* (1st edn, Allen Lane 2019)

Journal Articles

Besson S, 'Human Rights and Democracy in a Global Context: Decoupling and Recoupling' (2011) 4(1) *Ethics & Global Politics* pp. 19-50 <https://www.researchgate.net/publication/50434479_Human_rights_and_democracy_in_a_global_context_Decoupling_and_recoupling> accessed 21 January 2023

Bordt S and others, 'Post-Hoc Explanations Fail to Achieve Their Purpose in Adversarial Contexts' (2022) 2022 ACM Conference on Fairness, Accountability, and Transparency <https://arxiv.org/pdf/2201.10295.pdf> accessed 1 March 2023

Carriço G, 'The EU and Artificial Intelligence: A Human-Centred Perspective' (2018) 17(1) *European View* pp. 29-36 <https://journals.sagepub.com/doi/full/10.1177/1781685818764821> accessed 8 May 2023

Dietterich T.G and Horvitz E.J, 'Rise of concerns about AI: reflections and directions' (2015) 58(10) *Communications of the ACM*

<https://mags.acm.org/communications/october_2015/?folio=38&&pg=40#pg40> accessed 14 April 2023

Douglas-Scott S, 'The European Union and Human Rights after the Treaty of Lisbon' (2011) 11(4) *Human Rights Law Review*, 647 <https://doi.org/10.1093/hrlr/ngr038> accessed 17 April 2023

Fjeld J and others, 'Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI' (2020) Berkman Klein Center for Internet & Society Research Publication No. 2020-1 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482> accessed 15 March 2023

Forst R, 'The Justification of Human Rights and the Basic Right to Justification: A reflexive approach', (2010) 120(4) *Ethics* pp. 711–740. <https://www.jstor.org/stable/10.1086/653434> accessed 1 February 2023

Gill-Pedro E, 'The Most Important Legislation Facing Humanity? The Proposed EU Regulation on Artificial Intelligence' (2021) 4(1) *Nordic Journal of European Law* pp. IV-X

Guidotti R and others, 'A Survey of Methods for Explaining Black Box Models' (2018) 51 *ACM Computing Surveys* <https://dl.acm.org/citation.cfm?doid=3271482.3236009> accessed 16 April 2023

Günther, M and Kasirzadeh, A, 'Algorithmic and human decision making: for a double standard of transparency' (2021) 37 *AI & Soc* pp. 375–381 <https://link.springer.com/article/10.1007/s00146-021-01200-5> accessed 29 March 2023

Hansen HK, Christensen LT and Flyverbom M, 'Introduction: Logics of Transparency in Late Modernity: Paradoxes, mediation and governance' (2015) 18(2) *European Journal of Social Theory* pp. 117-131 <https://journals.sagepub.com/doi/10.1177/1368431014555254> accessed 27 March 2023

Larsson S, 'Transparency in artificial intelligence' (2020) 9(2) *Internet Policy Review* <https://policyreview.info/pdf/policyreview-2020-2-1469.pdf> accessed 27 March 2023

Margetts H, 'The Internet and Transparency' (2011) 82(4) *The Political Quarterly* pp. 518-521 <https://www.dhi.ac.uk/san/waysofbeing/data/citizenship-robson-margetts-2011b.pdf> accessed 29 March 2023

Molinari F and others, 'AI Watch. Beyond pilots: sustainable implementation of AI in public services' (Publications Office of the European Union 2021) JRC 126665, EUR 30868 EN, <https://www.standict.eu/node/5035> accessed 14 april 2023

Prabhakaran V and others, 'A Human Rights-Based Approach to Responsible AI' (2022) <https://arxiv.org/abs/2210.02667> accessed 15 March 2023

Rudin C, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) <https://arxiv.org/abs/1811.10154> accessed 18 January 2023

Shin D and Park YJ, 'Role of Fairness, Accountability, and Transparency in Algorithmic Affordance' (2019) 98 *Computers in Human Behavior* pp. 277-284<https://www.sciencedirect.com/science/article/pii/S0747563219301591> accessed 17 April 2022

Siegmann C and Anderljung M, 'The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market' (2022) <https://arxiv.org/pdf/2208.12645.pdf> accessed 9 May 2023

Sokol K and Flach P.A, 'Explainability Is in the Mind of the Beholder: Establishing the Foundations of Explainable Artificial Intelligence' (2022) <https://ui.adsabs.harvard.edu/abs/2021arXiv211214466S>
accessed 1 April 2023

Table of legislation

European Union Law

Charter of Fundamental Rights [2012] OJ C 326/391

Consolidated Version of the Treaty on European Union [2012] OJ C 326/47

Consolidated version of the Treaty on the Functioning of the European Union [2012] OJ C326/1

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) [2022] OJ L152/1

International Law

Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14, 4 November 1950

Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR)

Official sources and documents

European Union

European Commission

Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions – A Digital Single Market Strategy for Europe, COM(2015) 192 final

Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions on the Mid-Term Review on the implementation of the Digital Single Market Strategy – A Connected Digital Single Market for All, COM(2017) 228 final

Communication from the Commission - Artificial Intelligence for Europe, COM(2018) 237 final

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, COM(2018) 795 final

Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions Empty – Building Trust in Human-Centric Artificial Intelligence, COM(2019) 168 final

White Paper on Artificial Intelligence – A European Approach to excellence and trust, COM(2020) 65 final

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, COM(2020) 66 final

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Shaping Europe's digital future, COM(2020) 67 final

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final

Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM(2022) 496 final

European Council

Special meeting of the European Council (1 and 2 October 2020) – Conclusions (2020) EUCO 13/20 <https://www.consilium.europa.eu/media/45910/021020-euco-final-conclusions.pdf>accessed 10 May 2023

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - General approach (2022) 14954/22

European Court of Human rights

Guide on Article 8 of the European Convention on Human Rights' (2022) <https://www.echr.coe.int/documents/guide_art_8_eng.pdf>
accessed 23 March 2023

Opinions

Opinion 2/94 of the Court on Accession by the Community to the European Convention for the Protection of Human Rights and Fundamental Freedoms, [1996] ECR 1996 I-01759, EU:C:1996:140

Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' (2021) <https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf> accessed 27 March

The Organization for Economic Cooperation and Development

OECD, 'Artificial Intelligence & Responsible Business Conduct' (2019) <https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf> accessed 2 March 2022

OECD, 'Recommendation of the Council on Artificial Intelligence' (2022) OECD/ LEGAL/0449, < https://legalinstruments.oecd.org/api/print?ids=648&lang=en> accessed 4 May 2023

Expert reports and papers

European Union

AI HLEG, 'The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self Assessment' (2022) <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342> accessed 7 March 2023

Eager J, 'Opportunities of Artificial Intelligence, Study for the committee on Industry, Research and Energy, Policy Department for Economic, Scientific and Quality of Life Policies' (European Parliament 2020) PE 652 713 <http://www.europarl.europa.eu/RegData/etudes/STUD/2020/652713/IPOL_STU(2020)652713_EN.pdf> accessed 27 February 2023

EUR-Lex,'EU hierarchy of norms'<https://eur-lex.europa.eu/EN/legal-content/glossary/european-union-eu-hierarchy-of-norms.html> accessed 27 January 2023

European Commission, 'Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence' (2021) <https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682> accessed 8 May 2023

European Commission, 'Types of EU law' <https://commission.europa.eu/law/law-making-process/types-eu-law_en> accessed 27 January 2023

European Digital Rights and others, ‘An EU Artificial Intelligence Act for Fundamental Rights – A Civil Society Statement’ (2021) <https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf> accessed 20 February 2023

High-Level Expert Group on Artificial Intelligence – Ethics Guidelines for Trustworthy AI (2019) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 1 February 2020

Kritikos M, ’Artificial Intelligence ante portas: Legal & ethical reflections’ (European Parliamentary Research Service 2019) PE 634.427 <https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/634427/EPRS_BRI(2019)634427_EN.pdf> accessed 27 March 2023

Madiega T, ‘Briefing EU Legislation Artificial Intelligence Act’ (European Parliamentary Research Service 2022) <https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf> accessed 13 February 2023

Zamfir I, ‘At a Glance - the Universal Declaration of Human Rights and Its Relevance for the European Union’ (European Parliamentary Research Service 2018) <https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/628295/EPRS_ATA(2018)628295_EN.pdf> accessed 15 April 2023

Doctoral thesis

Buijze A, ’The Principle of Transparency in EU Law’ (DPhil thesis, Utrecht University 2013) <https://dspace.library.uu.nl/bitstream/handle/1874/269787/buijze%2Bapp.pdf?sequence=4> accessed 27 March 2023

Websites

European Union

EDPS.europa.eu ‘Artificial Intelligence Act: A Welcomed Initiative, but Ban on Remote Biometric Identification in Public Space Is Necessary’ (European

Data Protection Supervisor 23 May 2021) <https://edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative_en> accessed 3 May 2023

European Commission, 'A European approach to Artificial Intelligence' (2023), <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> accessed 14 February 2023

European Commission, 'Data Act' (2022) <https://digital-strategy.ec.europa.eu/en/policies/data-act> accessed 8 May 2023

European Council, 'Artificial Intelligence Act: Council Calls for Promoting Safe AI That Respects Fundamental Rights' (2022) <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/> accessed 10 May 2023

European Parliament, 'AI Act: A Step Closer to the First Rules on Artificial Intelligence' (2023) <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> accessed 3 May 2023

Other websites

Bertuzzi L, 'EU Lawmakers Set to Settle on OECD Definition for Artificial Intelligence' (9 March 2023) <https://www.euractiv.com/section/artificial-intelligence/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/?utm_source=substack&utm_medium=email> accessed 10 March 2023

——, 'AI Act: European Parliament Headed for Key Committee Vote at End of April' (31 March 2023) <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-european-parliament-headed-for-key-committee-vote-at-end-of-april/> accessed 2 April 2023

——, 'EU Launches AI Blueprint in Bid to Become World Leader' (21 April 2021) <https://www.euractiv.com/section/digital-single-market/news/commission-launches-ai-package-proposal/> accessed 28 April 2023

——, ‘AI Act: MEPs Close in on Rules for General Purpose AI, Foundation Models’ (24 April 2023) <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-meps-close-in-on-rules-for-general-purpose-ai-foundation-models/?utm_source=substack&utm_medium=email> accessed 2 May 2023

——, ‘AI Regulation Filled with Thousands of Amendments in the European Parliament’ (7 June 2022) <https://www.euractiv.com/section/digital/news/ai-regulation-filled-with-thousands-of-amendments-in-the-european-parliament/> accessed 3 May 2023

GDPR.eu, ‘What Is GDPR, the EU’s New Data Protection Law? - GDPR.eu’ (7 November 2018) <https://gdpr.eu/what-is-gdpr/> accessed 10 May 2023

OECD.ai, ‘Putting the OECD AI Principles into Practice: Progress and Future Perspectives - OECD.AI’ (2021) <https://oecd.ai/en/mcm> accessed 3 May 2023

OECD.org, ‘Forty-Two Countries Adopt New OECD Principles on Artificial Intelligence - OECD’ (22 May 2019) <https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm> accessed 4 May 2023

Table of Cases

CJEU

Cruciano Siragusa v Regione Sicilia Case (C-206/13) [2014] EU:C:2014:126

Netherlands v European Parliament and Council (C-377/98) [2001] ECR I-7079