TU WIEN Informatics

European Union's Artificial Intelligence Act (AI Act)

lessons learned from the history of artificial intelligence and their consideration towards a future-proof legal regulation

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science

in

Media Informatics and Visual Computing

by

Livia Rose Ecker
Registration Number 01526839

to the Faculty of Informatics

at the TU Wien

Advisor: ao.Univ.-Prof. Mag. Dr.iur. Markus Haslinger

Vienna, 29th July, 2024

Livia Rose Ecker

Markus Haslinger

Erklärung zur Verfassung der Arbeit

Livia Rose Ecker

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 29. Juli 2024

Livia Rose Ecker

Acknowledgements

I would like to express my deepest appreciation to my supervisor. His guidance, his patience and his profound belief in my work and my abilities formed the basis of my success in completing this thesis. I would also like to express my deepest gratitude to myself. Thanks for persevering and completing this work.

Abstract

Artificial Intelligence (AI) had, still has, and is going to have an enormous influence on our daily lives. With its unique characteristics this technology provides crucial competitive advantages and benefits. However, it also bears the risk of potential harm to society and thereby its trustworthiness is negatively impacted. Nevertheless, trust in AI systems is mandatory in order to take advantage of its benefits. That is why the European Union had already been following the path of achieving trustworthy AI for many years. Finally, they introduced the European Artificial Intelligence Act (AI Act) on the 21st April 2021 which has already been agreed on on the 9th December 2023. Its aim is to support the innovation of AI while preventing its potential risks.

Within this thesis the future-proof approach of the AI Act was examined. First, it was analyzed to which extent challenges and their solutions of the development history of AI have been considered in the establishment of the new regulation. Second, current loopholes in the legal system of the EU and how the AI Act has covered them as well as how the EU ensured that they implemented a future-proof approach was discussed.

It was reflected that future AI governance must take greater account of the development history of AI, as this could speed up its processes and avoid repeated errors. Further, the unique characteristics of AI were not only a complicated factor throughout its development, it also challenged the enforcement of a regulatory framework. First, it is challenging to deliberately formulate a legal framework that covers every aspect of the technology. Second, any implementation phase of a new regulation is time-consuming and costly. That is why the enforcement of the AI Act bears the risk of AI providers from other countries overtaking or undercutting EU partners in the development of AI solutions as they do not have to comply with the regulation.

Besides that, the results have shown that the AI Act serves as a crucial foundation in the field of AI governance worldwide. Due to its emphasis on innovation as well as ethical considerations it reflects the need of benefiting from AI while mitigating its potential harm. However, moving forward it is essential to closely monitor the enforcement of the AI Act to ensure that the right balance of innovation and risk prevention is given.

Finally, it is important to note that the AI Act is not a standalone piece of legislation. It must be seen in the wider context of the EU law. Of particular note are the GDPR, the accompanied proposal of the AILD and the revised PLD.

Contents

Acronyms

AGI Artificial General Intelligence. 15, 87, 91

AI Artificial Intelligence. vii, ix, x, 1–4, 7–12, 14–31, 33–38, 41, 43–53, 55–85, 87–95, 97, 98

AI Act European Artificial Intelligence Act. vii, 3–5, 59, 67–69, 71–74, 76–79, 81–84, 87–95, 97, 98

AI HLEG High-Level Expert Group on Artificial Intelligence. 33, 34, 37, 41, 60, 65, 89, 91

AILD European Artificial Intelligence Liability Directive. vii, 5, 67, 83–85, 93, 98

ANI Artificial Narrow Intelligence. 14, 91

ANN Artificial Neural Network. 10, 20, 21, 27

ASI Artificial Super Intelligence. 15, 87, 91

BD Big Data. 9–11, 28, 29, 33, 88

CEPEJ European Commission for Efficiency of Justice. 62

CFR European Charter of Fundamental Rights. 34, 35, 43–45, 47, 56, 70, 91, 92

DL Deep Learning. 9, 10, 21, 26–28, 33, 88, 89

DNN Deep Neural Networks. 10

EC European Commission. 51, 65

EESC European Economic and Social Committee. 56, 59

EGE European Group on Ethics in Science and New Technologies. 58

EU European Union. 4, 43, 47, 59, 70, 71, 97

FRIA Fundamental Rights Impact Assessment. 81, 91, 92

GDPR European General Data Protection Regulation. vii, ix, 43, 47–51, 60, 89, 91–93, 95, 98

GPAI models General purpose AI models. 68, 72, 73, 79–81

GPAI systems General purpose AI systems. 79

GPS General Problem Solver. 21, 22

ML Machine Learning. 9, 10, 26, 33, 88, 89

NN Neural Network. 9, 10, 26, 89

NTF New Technologies formation. 63, 66, 88

OECD Economic Co-operation and Development. 71, 90

PLD European Product Liability Directive. vii, 5, 43, 51, 52, 65, 67, 82–84, 90, 91, 93, 98

PLF Product Liability Directive. 63, 65

SNN Simulated Neural Network. 10

CHAPTER 1

Introduction

"We may eventually have to worry about all-powerful machine intelligence. But first we need to worry about putting machines in charge of decisions that they don't have the intelligence to make."

— Jon Kleinberg, Sendhil Mullainathan

1.1 Problem

Artificial Intelligence (AI) had, still has and is going to have an enormous influence on our daily life's.[1,2] Nowadays it even affects our way of behaving and thinking.[3] Yet, the current hype about AI makes it seem like a new technology was discovered whether in fact it already exists for decades.[4,5] It was due to exponentially cheaper computing and the broad availability of data in recent years[6] that its potential increased enormously and therefore it gained in value, importance and popularity.[7,8]

AI has the potential to replace existing technologies, products or services in a fundamental way and therefore can be classified as disruptive innovation.[9,10] Compared to other

[1] Ruschemeier (2023), "AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal", p. 362.
[2] Couch (2023), "Artificial Intelligence: Past, Present and Future", pp. 1, 2.
[3] Ruschemeier (2023), "AI as a challenge for legal regulation", p. 362.
[4] Chen et al. (2023), "Systematic analysis of artificial intelligence in the era of industry 4.0", p. 89.
[5] Panesar (2020), "What is Artificial Intelligence?"
[6] Panesar (2020), "What is AI?".
[7] Li et al. (2022), "Liability Rules for AI-Related Harm: Law and Economics Lessons for a European Approach", p. 1.
[8] Akinrinola et al. (2024), "Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability", p. 50.
[9] Ruschemeier (2023), "AI as a challenge for legal regulation", p. 362.
[10] Secinaro et al. (2021), "The role of artificial intelligence in healthcare: a structured literature review", p. 16.

technologies, it is outstanding because of its complexity, opacity and autonomy.[11] It provides the ability of evaluating complex scenarios with the outcome of a decision or a technical trigger without any additional human intervention.[12] Through that, decisions or operations can be improved or optimized which gives the opportunity to benefit the environment and society.[13] Therefore, crucial competitive advantages can be the result of the application of AI that might take companies and countries to the next level.[14] Especially in high-impact sectors the need is given to take advantage of the power of AI, including climate climate change and health.[15] Nonetheless, AI also bears the potential to harm society or an individual and thereby its trustworthiness is negatively impacted.[16]

Due to the fact that AI systems are trained on real life data, already existing ethically controversial topics in our society are a part of these systems as well.[17] Therefore, grievances might be perpetuated, reinforced or even worse, new ones could arise because of its autonomously learning process.[18,19] Additionally, its complexity impedes a proper traceability. AI utilizes neural networks that are imitating neurons of the human brain which are further known for their abstruseness.[20] As a consequence, developers are sometimes not able to comprehend the whole process themselves. If developers are not capable of that, neither will be end users. Considering the lack of human intervention, it allows for a state where nobody is in the position to take responsibility if a risk actually comes into operation.[21] This accountability gap as well hinders the trust of society.[22] As can be seen, many outstanding challenges need to be solved in order to achieve trustworthiness in the field of AI.

Still, for AI bringing advantages to society, trust in decisions made by these systems is required.[23,24] Otherwise society might avoid or even refuse the use of that technology.[25] Therefore, achieving trustworthiness of AI is the main goal to follow in order to make use of the advantages that the technology is carrying.[26] For years the European Union

[11] Li et al. (2022), "Liability Rules for AI-Related Harm", p. 1.
[12] Meyer (2021), "Rechtliche Herausforderungen der Künstlichen Intelligenz und ihre Bewältigung", p. 25.
[13] 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*, p. 1.
[14] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[15] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[16] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[17] Natasa et al. (2023), "Artificial Intelligence: Friend or Foe? Experts' Concerns on European AI Act", p. 5.
[18] Akinrinola et al. (2024), "Ethical dilemmas in AI development", p. 51.
[19] Raji et al. (2020), "Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing", p. 1.
[20] Chen et al. (2023), "Systematic analysis", p. 96.
[21] Raji et al. (2020), "Closing the AI Accountability Gap", p. 2.
[22] Ruschemeier (2023), "AI as a challenge for legal regulation", p. 363.
[23] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 5.
[24] Natasa et al. (2023), "Friend or Foe?", p. 6.
[25] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 5.
[26] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 4.

started to discuss possible solutions of ensuring the trustworthiness of AI, however, it was not until 2017 that it had been seen as an independent field of governance[27]. Later on in 2018, the European Union and its member states agreed on cooperation on AI governance and started to develop strategies and plans to realize the cooperation on AI.[28,29,30]

Their first step towards a possible regulation was the introduction of an ethical framework to ensure trustworthiness of AI that was based on fundamental rights of the European Union.[31] However, that was based on the assumption that all legal rights and obligations that apply throughout the life cycle of AI remain binding and must continue to be complied with.[32] Nevertheless, many analysis of the current applicable laws have pointed out possible concerns and loopholes in the current legal framework of the European Union.[33,34] As a result the need for a regulatory framework addressing AI became clear.[35] The published 'White Paper' in 2020 then was the final trigger of establishing a legal framework related to AI.[36] As a result the Commission finally proposed the AI Act in 2021, a regulatory framework addressing, mitigating and preventing risks and challenges posed by AI.[37] Thus, the establishment of any legal system is a rather slow process as it is a time consuming task due to involved parties, their negations and the cycle of its formal adaption. Thereof, the final agreement on the AI Act only took place in the end of 2023.[38]

However, the field of AI continues to develop exponentially.[39] Therefore, in addition to ethical and legal challenges arising from AI, the AI Act must also implement a future-proof approach in order to keep pace with its constant evolution and increasing capabilities.[40] That is why the European Union is committed to take a balanced approach to ensure both, that it maintains its technological leadership and that newly developed technologies and their functioning are in line with the values, fundamental rights and principles of the

[27] EESC, *Opinion of the European Economic and Social Committee on 'Artificial intelligence — The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society'*, p. 1.
[28] Stix (2022), "The Ghost of AI Governance Past, Present, and Future: AI Governance in the European Union", p. 4.
[29] Commission, *Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Comittee and the Commitee of the Regions Artificial Intelligence for Europe*, pp. 5-16.
[30] Commission (2023), *Commission welcomes political agreement on Artificial Intelligence Act*, p. 2.
[31] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 2.
[32] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 6.
[33] Sartor (2020), "Artificial intelligence and human rights: Between law and ethics", p. 712.
[34] (2023), "Reconciling Artificial Intelligence (AI) With Product Safety Laws", p. 1.
[35] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[36] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[37] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[38] Legislative Train Schedule, *Artificial intelligence act In "A Europe Fit for the Digital Age"*, accessed on 7.5.2024.
[39] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[40] Commission, *Proposal Artificial Intelligence Act*, p. 1.

Union, so that citizens can benefit from their use.[41]

1.2 Objective and Motivation

The aim of this thesis is to examine whether the technological history of AI was taken into account during the establishment of the EU AI Act and how it was ensured that it is following a future-proof approach. Thereof, the necessity of considering the technical progress as well as arising concerns and challenges of AI in the approach of legally regulating it must be set out. Further, current loopholes of the legal system and how the AI Act tries to overcome them must be taken into account. Finally, elements that are ensuring that the AI Act follows a future-proof approach must be investigated. Resulting research questions are defined as followed:

- *What conclusions can be drawn from comparing the history of AI with the journey of the European Union towards achieving trustworthy AI?*
- *How does the AI Act ensures that it follows a future-proof approach and what existing legal concerns were taken into account during its development?*

1.3 Approach

Artificial Intelligence as a Research Field will first discuss the lack of a globally unique accepted definition of the term AI and its resulting subdivision into smaller subareas. Second, categorizations that had been established overtime due to the intangible wide-ranging characteristics of AI will be discussed. Finally, the ups and downs in the history of AI will be laid out as well as key finding on how to prevent another down in the future of this technology.

The future goal of achieving trustworthy AI layed out the importance of achieving trustworthy AI to enable an ongoing development and application of AI. It will further discuss the approach of the European Union of an ethical framework to ensure trustworthiness and why that approach is insufficient as an appropriate legal regulation of the technology is of importance due to its opacity, complexity and autonomy.

Applicable legal acts and concerns posed by AI is complementary to the previous chapter, as it will discuss current applicable legal sources and their concerns posed by AI. Only the most relevant primary and secondary laws for the future implementation of the AI Act will be taken into account as the scope of this thesis does not allow otherwise.

European Union's journey towards trustworthy AI will lay out the start of AI centered governance and their path towards the establishment of the AI Act. It will highlight the main concerns and challenges underlying the proposal for the AI Act.

[41] Commission, *Proposal Artificial Intelligence Act*, p. 1.

Finally, *European Artificial Intelligence Act - a future proof solution?* will discuss most important regulations, provisions and obligations of the AI Act as well as its interplay with the revised PLD and the AILD. Further, important changes during the development of the regulation from its proposal of the Commission to its legislative resolution of the Parliament will be pointed out.

CHAPTER 2

Artificial Intelligence as a Research Field

The fear of the unknown is a natural human instinct that serves as a warning system against potential dangers.[1] Hence it is not surprising that humanity tend to fear and mistrust new technologies including AI. However, it is a misconception that AI is a newly invented technology since it has been researched on for seven decades[2,3] and its roots can be traced back even further[4]. In Greek mythology there was already the idea of an artificial replica of a human being.[5] Back then, Pandora was the idea of a woman created by gods, endowed with human-like qualities but bearing the potential of unpredictable consequences.[6]

AI being seen as a new invention is not a phenomenon unique of today. Its history had not experienced a linear progression of success, rather it was marked by periods of great enthusiasm as well as periods of decreased interest and funding. Many subareas got established over time, following different approaches to converge towards the aim of AI. Each period of enthusiasm was triggered by advancements in one of these subareas while periods of decreased interest and funding happened due to disappointment. Those fluctuating phases and the rise and fall of these subareas gave the impression of newly invented technologies whether in fact it always were approaches towards AI.

[1] Leong (2023), "Rethinking Human Motivation Psychology: The Hierarchy of Human Fear Model", p. 2.
[2] Toosi et al. (2021), "A Brief History of AI: How to Prevent Another Winter (A Critical Review)", p. 4.
[3] Biore (2017), "Understanding AI in a world of big data", p. 29.
[4] Heanlein et al. (2019), "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence", p. 6.
[5] Haesik (2022), "Historical Sketch of Artificial Intelligence", p. 9.
[6] Mayor (2018), "What Pandora's Box tells us about AI".

Understanding the history of AI provides the context for a better understanding of its current state and its possible future development. Many breakthroughs were due to earlier concepts and experiments and therefore it is important to observe past events. Further, since AI has already faced several challenges in the past understanding its history is the foundation of avoiding repeated mistakes. These overcome challenges should also be taken into account in the future studies and regulations on AI.

2.1 Definition of Artificial Intelligence

Since the term AI was coined, many attempts were made to find a suitable definition.[7,8,9] However, that is almost impossible as its development is always changing and evolving and the term intelligence itself has no unique definition itself.[10] As various disciplines followed different approaches, each definition attempt focused on another subarea of AI which made the resulting definition not comprehensive enough to be accepted as a global one. Although a broad spectrum of definitions already exists, to this day no generally accepted definition of what AI actually involves could be agreed on.[11,12,13] People from various fields of expertise are still trying to create a suitable definition that covers the whole extensive scope of AI.[14] However, even converging towards a comprehensive description is almost impossible due to its multidisciplinary nature.[15,16]

Artificial Intelligence is a research field that deals with the idea to simulate human intelligence within a technical environment.[17] Therefore, the aim of AI systems is to include the same characteristics as of human intelligence including thinking, reasoning and interacting.[18,19] Due to its complexity and interdisciplinary nature, a variety of subdomains have been developed were each of them focuses on the implementation of different characteristics of AI.[20] These subdomains are including reasoning, planning, learning, communicating and perception[21] and can be used to approach and better understand the term AI.[22,23]

[7] Haesik (2022), "Historical Sketch of AI", p. 4.
[8] Schuett (2019), "A Legal Definition of AI", p. 1.
[9] Samoili et al. (2020), "AI Watch Defining Artificial Intelligence", p. 7.
[10] Mitchell (2021), "Why AI is Harder Than We Think", p. 8.
[11] Samoili et al. (2020), p. 7.
[12] Haesik (2022), "Historical Sketch of AI", p. 4.
[13] Schuett (2019), "A Lgeal Definition of AI", p. 3.
[14] Samoili et al. (2020),
[15] Haesik (2022), "Historical Sketch of AI", p. 4.
[16] Schuett (2019), "A Lgeal Definition of AI", p. 1.
[17] (2022), "A Machine Learning-Based Model for Epidemic Forecasting and Faster Drug Discovery", p. 5.
[18] Haesik (2022), "Historical Sketch of AI", p. 4.
[19] Shinde et al. (2018), "A Review of Machine Learning and Deep Learning Applications", p. 1.
[20] Deng (2018), "Artificial Intelligence in the Rising Wave of Deep Learning: The Historical Path and Future Outlook", p. 173.
[21] Samoili et al. (2020), p. 11.
[22] Samoili et al. (2020), p. 11.
[23] Deng (2018), "AI in the Rising Wave of DL", p. 173.

Within the scope of this paper we will especially consider the learning subdomain as it is one of the most important ones.[24] The subdomain learning encompasses subareas as Machine Learning, Neural Network and Deep Learning (DL).[25] The development of AI as well as its subareas was and still is heavily depended on data. Thus, the invention of Big Data was a massive milestone in the history of AI and still is of great importance. The relationships of these terms are visually represented in figure 2.2.

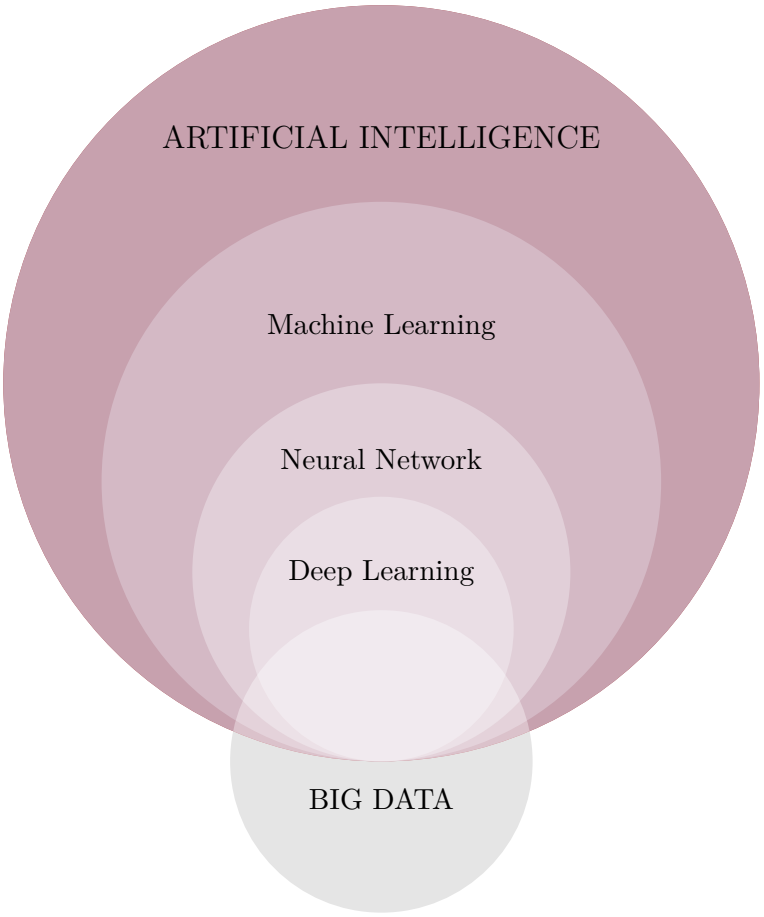

Figure 2.1: Artificial Intelligence, Machine Learning, Deep Learning and Big Data

2.1.1 Machine Learning (ML)

Machine Learning is a subset of AI[26,27] that enables systems the ability to automatically learn based on experiences and their improvement over time without any further human

[24] Deng (2018), "AI in the Rising Wave of DL", p. 173.
[25] Deng (2018), "AI in the Rising Wave of DL", p. 173.
[26] Santosh and Wall (2022), *AI, Ethical Issues and Explainability—Applied Biometrics*, p. 3.
[27] (2022), p. 5.

assistance.[28] Due to its self-improving behaviour such systems do not need to be explicitly programmed.[29] Learning types of ML can be classified into three categories: (1) supervised learning where labelled input datasets and known responses are applied to develop a regression or classification model that further can be applied to new datasets in order to generate predictions, (2) unsupervised learning where unlabelled data gets applied and the system itself is able to classify and process that data and further learn from its inherent structure and (3) reinforcement learning where algorithms learn through their interaction between the AI system and its environment.[30]

2.1.2 Neural Network (NN)

Neural Network is a subset of ML which is based on the knowledge of the way that neurons of the human brain work.[31] More precisely, it is a specific architecture of ML consisting of artificial neurons that are able to navigate signals between each other.[32] Neural Network is also known as Artificial Neural Network (ANN)[33] or Simulated Neural Network (SNN) and it is the backbone of Deep Learning.

2.1.3 Deep Learning (DL)

Deep Learning is an advancement of the concept of NN.[34,35] If a NN consists of three or more layers it is considered DL as each additional layer of depth optimizes the accuracy of predictions.[36] It provides the ability for systems to learn from structured as well as unstructured data.[37] Thus, DL algorithms are indeed ML algorithms, its NN architecture requires less human intervention and it has the ability to handle unstructured data.[38] DL is also known under the term Deep Neural Networks (DNN).[39]

2.1.4 Big Data (BD)

Still, no uniform definition for Big Data exists.[40] Yet, it can be seen as massive complex datasets which are impossible to be analyze by traditional statistical modeling tools.[41]

[28] Hassanien et al. (2021), *Enabling AI Applications in Data Science*, p. 433.
[29] Hassanien et al. (2021), *Enabling AI Applications*, p. 433.
[30] BioStrand, *AI, ML, DL, and NLP: An Overview*, accessed on 5.5.2024.
[31] Biore (2017), "Understanding AI", p. 30.
[32] Biore (2017), "Understanding AI", p. 30.
[33] (2022), p. 7.
[34] Taye (2023), "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions", p. 2.
[35] Hassanien et al. (2021), *Enabling AI Applications*, p. 186.
[36] Shinde et al. (2018), "A Review of ML and DL Applications", p. 3.
[37] Hassanien et al. (2021), *Enabling AI Applications*, p. 186.
[38] Shinde et al. (2018), "A Review of ML and DL Applications", p. 3.
[39] Fradkov (2020), "Early History of Machine Learning", p. 1387.
[40] Reusch (2023), "Handlungsfähigkeit durch, trotz und gegenüber (Big) Data und KI: Eine Bestandsaufnahme mit Hilfe des Frankfurt-Dreiecks", p. 1.
[41] Demigha (2020), *The impact of Big Data on AI*, p. 1395.

Through established methods of AI these sets can be explored an analyzed to gain information and insights.[42] Five characteristics are defined to point out the unique technical and analytical requirements of Big Data which are called '5Vs'.[43]

Volume stands for the quantity of data. These immense amount of data are challenging traditional storage and processing capabilities.[44]

Velocity stands for the pace at which data is generated. Distributed processing techniques are required because of that speed and quantity.[45]

Variety stands for the variety of data types. It takes raw, structured, semi-structured and unstructured data and therefore challenge traditional analytic systems to analyze and process them. Distinct processing capabilities and specialist algorithms are needed.[46]

Veracity stands for the quality of data. It can be differentiated into high and low veracity. High veracity data is mainly valuable to analyze with the outcome of a meaningful result. Contrary, low veracity data contains mostly meaningless data which is also called noise.[47]

Value stands for generating value from the available data. It is the most important quality of BD. Data itself is useless unless it gets converted or analyzed to further gain valuable insights.[48]

2.2 Categorizations in virtue of a missing definition

In virtue of the absence of a globally accepted definition, people attempted to establish categorizations of AI in order to define its scope and possibilities more precisely.[49] As different disciplines were requiring different categorization approaches it resulted in the establishment of several categorizations.[50] That might give the impression of a huge overload, but each of them is of importance. Even the introduction of new categorizations in the future cannot be ruled out since it is important to get a better understanding of the extent of AI as long as no definition can be agreed on.

[42] Biore (2017), "Understanding AI", p. 31.
[43] Biore (2017), "Understanding AI", p. 23.
[44] Kusak (2022), "Quality of data sets that feed AI and big data applications for law enforcement", p. 211.
[45] Kusak (2022), "Quality of data sets that feed AI", p. 211.
[46] Kusak (2022), "Quality of data sets that feed AI", p. 211.
[47] Kusak (2022), "Quality of data sets that feed AI", p. 212.
[48] Kusak (2022), "Quality of data sets that feed AI", p. 212.
[49] Haesik (2022), "Historical Sketch of AI", p. 4.
[50] Haesik (2022), "Historical Sketch of AI", p. 4.

2.2.1 Research

After the term AI was coined people attempted to reproduce human intelligence within a technical environment.[51] However, age-old questions that were already a hurdle for many other research areas impeded this process.[52] Stuart Russel and Peter Norvig were investigating into these unresolved issues and introduced a breakdown of four main AI research categories.[53,54] The base concept is the distinction between the thinking and acting process, whereby the perspective from which these processes are viewed can be either human-based or rational: think humanly, act humanly, think rationally, act rationally.[55,56]

Think humanly - the cognitive modeling approach

As the name implies, human thought processes are the focus of this research category.[57,58] Important to note is that human thought processes are studied regardless of their legitimacy or accuracy.[59,60] In order to create systems with the ability to think like a human being, the actual way of working of the human mind needs to be understood.[61] Three ways have been established to learn about human thoughts: catching our own thoughts as they go by (introspection), observing a person in action (psychological experiments) and observing the brain in action (brain imaging).[62] Therefore, a sufficiently accurate theory of mind must be established first to enable the progress of AI.[63] Cognitive science is concerned with solving this problem.[64,65] It unites AI systems and experimental techniques from psychology to construct precise and testable theories of the human mind.[66,67]

Think rationally - the laws of thought approach

As the name implies, rational thinking processes are the focus of this research category.[68] Those processes heavily depend on two different areas: logic and probability.[69] Aristotle

[51] Haesik (2022), "Historical Sketch of AI", p. 4.
[52] Haesik (2022), "Historical Sketch of AI", p. 4.
[53] Haesik (2022), "Historical Sketch of AI", p. 4.
[54] (2021), "Autonomy in AI Systems: Rationalizing the Fears", p. 39.
[55] Russel et al. (2010), *Artificial Intelligence: A Modern Approach*, pp. 1-2.
[56] Haesik (2022), "Historical Sketch of AI", p. 4.
[57] Haesik (2022), "Historical Sketch of AI", p. 4.
[58] Russel et al. (2010), *A Modern Approach*, p. 3.
[59] Haesik (2022), "Historical Sketch of AI", p. 4.
[60] Russel et al. (2010), *A Modern Approach*, p. 3.
[61] Russel et al. (2010), *A Modern Approach*, p. 3.
[62] Russel et al. (2010), *Artificial Intelligence: A Modern Approach*, p. 20.
[63] Russel et al. (2010), *A Modern Approach*, p. 3.
[64] Haesik (2022), "Historical Sketch of AI", p. 4.
[65] Russel et al. (2010), *A Modern Approach*, p. 3.
[66] Haesik (2022), "Historical Sketch of AI", p. 4.
[67] Russel et al. (2010), *A Modern Approach*, p. 3.
[68] Haesik (2022), "Historical Sketch of AI", p. 4.
[69] Russel et al. (2010), *A Modern Approach*, p. 21.

was one of the first to attempt to describe the process of "right thinking" and established syllogisms, a catalog of certain types of logical conclusions.[70] The study of these laws of thought initiated the field called logic.[71] As logic requires the world to be certain, which in fact is a seldom achieved condition, probability needs to be considered as well.[72] In principle, it enables the modeling of rational thought, which leads from raw perceptual information to an understanding of how the world works and to predictions about the future.[73] However, rational thinking does not imply intelligent behaviour.[74] To achieve that, a theory of rational acting is needed.[75]

Act humanly - the turing test approach

As the name implies, the behaviour of human beings is the focus of this research category.[76] The turing test was introduced as a tool to interrogate into the age old questions whether machines are being able to think.[77] It is a question-answer game where the outcome depends on a human being being able to distinguish between another human being and a computer.[78] Therefore it laid the foundation of testing if machines are able to act like a human being instead of testing if machines are acting intelligent.[79] To pass this test the computer would need four capabilities: natural language processing to enable a successful communication in human language, knowledge representation to store retrieved information, automated reasoning to draw conclusions and answer questions and machine learning to detect and extract patterns and further adapt to new circumstances.[80] Other researchers have further introduced the total turing test to involve interactions with objects and persons in the real world. To achieve that two more capabilities are required: computer vision and speech recognition to perceive the world and robotics to move around and manipulate objects.[81] However, no big effort was shown in trying to pass the turing test.[82] Researchers believe that it is more important to study the underlying principles of intelligence rather than to duplicate it.[83]

[70] Russel et al. (2010), *A Modern Approach*, p. 21.
[71] Russel et al. (2010), *A Modern Approach*, p. 21.
[72] Russel et al. (2010), *A Modern Approach*, p. 21.
[73] Russel et al. (2010), *A Modern Approach*, p. 21.
[74] Russel et al. (2010), *A Modern Approach*, p. 21.
[75] Russel et al. (2010), *A Modern Approach*, p. 21.
[76] Haesik (2022), "Historical Sketch of AI", p. 4.
[77] Russel et al. (2010), *A Modern Approach*, p. 20.
[78] Russel et al. (2010), *A Modern Approach*, p. 20.
[79] Russel et al. (2010), *A Modern Approach*, p. 20.
[80] Russel et al. (2010), *A Modern Approach*, p. 20.
[81] Russel et al. (2010), *A Modern Approach*, p. 20.
[82] Russel et al. (2010), *A Modern Approach*, p. 20.
[83] Russel et al. (2010), *A Modern Approach*, p. 20.

Act rationally - the rational agent approach

As the name implies, rational behaviour is the focus of this research category.[84] Within this category AI is viewed as the construction of rational agents[85], one that acts to achieve the best expected outcome with uncertainty taking into account.[86] Skills provided by the turing test are important to allow an agent to act rational. Further, rational thinking is a prerequisite of rational acting.[87] However, rational acting sometimes requires to go beyond drawing conclusions based on logic and probability.[88] Compared to other approaches it is more general and better suited for scientific development.[89]

2.2.2 Emulation of Human Capability

This categorization subdivides AI systems based on their capability to emulate human beings.[90] It is focused on the process of how a system is learning and how far it can apply its knowledge. In 1980 the philosopher John Searle introduced the categories weak AI and strong AI which later on were replaced by Artificial Narrow Intelligence and Artificial General Intelligence.[91] Further it was complemented by the category of Artificial Super Intelligence.

Artificial Narrow Intelligence (ANI)

Artificial Narrow Intelligence is the only type of this categorization that has already been realized.[92] Systems belonging to this category have limited capabilities and cannot grow in their abilities.[93] Therefore they can only perform tasks which they are designed for and lack the capability of general problem solving.[94] However these machines appear to be intelligent as they even can outperform human beings if trained properly.[95]

This category is also known as **weak AI**[96] and the hypothesis about weak AI says that it is possible for machines to act as if they were intelligent.[97] Since machines are able to outperform human beings in tasks that they are designed for and they are only acting intelligent instead of actually being intelligent this hypothesis is proven.

[84] Haesik (2022), "Historical Sketch of AI", p. 4.
[85] Haesik (2022), "Historical Sketch of AI", p. 5.
[86] Russel et al. (2010), *A Modern Approach*, p- 22.
[87] Haesik (2022), "Historical Sketch of AI", p. 5.
[88] Russel et al. (2010), *A Modern Approach*, p- 22.
[89] Russel et al. (2010), *A Modern Approach*, p- 22.
[90] Haesik (2022), "Historical Sketch of AI", p. 6.
[91] Russel et al. (2010), *A Modern Approach*, p. 1032.
[92] Kalota (2024), "A Primer on Generative Artificial Intelligence", p. 2.
[93] Dorr (2022), "Types of Artificial Intelligence, Explained", p. 2.
[94] Kalota (2024), "A Primer on Generative AI", p. 2.
[95] Kalota (2024), "A Primer on Generative AI", p. 2.
[96] Kalota (2024), "A Primer on Generative AI", p. 2.
[97] Russel et al. (2010), *A Modern Approach*, p. 1020.

Artificial General Intelligence (AGI)

Many science fiction books are already making use of the idea of Artificial General Intelligence but in real life it is not yet implemented.[98,99] Systems belonging to this category will have the same capacity as human beings and therefore can execute almost any given task if trained properly.[100] To make them indistinguishable to a human being its aim is to create machines that are constantly improving in their capabilities by being able to think, act and learn from experiences.[101] Based on that they can construct an own mind, make decisions and act independently in different environments without any human intervention.[102] They will have their own intelligence and the capability of general problem solving.[103]

This category is also known as **strong AI**.[104] and the hypothesis about strong AI says that machines that act as if they were intelligent are not only simulating thinking, they are actually thinking.[105] As soon as AGI is realized this hypothesis will be proven.

Artificial Super Intelligence (ASI)

Artificial Super Intelligence, describes machines with capabilities that are going far beyond human capabilities regardless of the area.[106] They are significantly more intelligent than human beings and have their own needs, beliefs and desires.[107] Due to today no machines exist that nearly reaches this level.[108] It is only a hypothetical concept.[109]

2.2.3 Functionality

This categorization subdivides AI systems into four distinctive groups based on their functionality.[110] Its functionality is characterized by the way a system applies its learning capabilities to process data, respond to stimuli and interact with its environment.[111]

[98] Haesik (2022), "Historical Sketch of AI", p. 5.
[99] Panesar (2020), "What is Artificial Intelligence?"
[100] Kalota (2024), "A Primer on Generative AI", p. 2.
[101] Kalota (2024), "A Primer on Generative AI", p. 2.
[102] Dorr (2022), "Types of AI, Explained", p. 2.
[103] Dorr (2022), "Types of AI, Explained", p. 2.
[104] Kalota (2024), "A Primer on Generative AI", p. 2.
[105] Russel et al. (2010), *A Modern Approach*, p. 1020.
[106] Kalota (2024), "A Primer on Generative AI", p. 2.
[107] Kalota (2024), "A Primer on Generative AI", p. 2.
[108] Dorr (2022), "Types of AI, Explained", p. 39.
[109] Kalota (2024), "A Primer on Generative AI", p. 2.
[110] Haesik (2022), "Historical Sketch of AI", p. 6.
[111] Dorr (2022), "Types of AI, Explained", p. 38.

Reactive machines

Reactive machines are the most basic type of AI.[112] They are designed for one specific task and cannot grow in their abilities.[113,114] Based on statistical models and algorithms they draw conclusions from patterns found in data.[115] Huge amounts of data is analyzed to produce an accurate output since reactive machines do not have any memory.[116] This lack of storing experiences makes it impossible to respond with a decision based on past experiences.[117] Its outcome is only based on taught or recalled data.[118] Reactive machines are reliable in completing specific tasks which they are trained for. However they lack interaction, emotion and consciousness and further can easily be tricked.

Limited memory

Limited memory machines enable the ability of a first-stage learning process.[119] Such machines have the same capability of decision making as reactive machines do but additionally they are able to learn from past input.[120] Large volumes of data and experimental knowledge is stored for a short period of time.[121,122] Based on that data these machines are able to learn, make decisions and update experimental knowledge.[123,124] The resulting decisions are more accurate than from reactive machines since stored data gets filtered through and inferences about what might happen are made.[125] However its short time storage makes it transient.

Theory of mind

The term theory of mind co-opted from psychology which describes the process of accessing mental states of others and understanding them.[126] It is an important aspect of acting socially.[127] Theory of mind AI therefore has the aim to understand intents (i.e.: emotions, beliefs, thoughts, needs and goals) of individuals it is interacting with whether it is a person or another AI machine.[128,129] Until today this level of AI is under development

[112] Haesik (2022), "Historical Sketch of AI", p. 6.
[113] Panesar (2020), "What is AI?".
[114] Haesik (2022), "Historical Sketch of AI", p. 6.
[115] Dorr (2022), "Types of AI, Explained", p. 38.
[116] Dorr (2022), "Types of AI, Explained", p. 38.
[117] Haesik (2022), "Historical Sketch of AI", p. 6.
[118] Panesar (2020), "What is AI?".
[119] Dorr (2022), "Types of AI, Explained", p. 38.
[120] Dorr (2022), "Types of AI, Explained", p. 38.
[121] Dorr (2022), "Types of AI, Explained", p. 38.
[122] Haesik (2022), "Historical Sketch of AI", p. 6.
[123] Dorr (2022), "Types of AI, Explained", p. 38.
[124] Haesik (2022), "Historical Sketch of AI", p. 6.
[125] Dorr (2022), "Types of AI, Explained", p. 38.
[126] Dorr (2022), "Types of AI, Explained", p. 39.
[127] Dorr (2022), "Types of AI, Explained", p. 36.
[128] Dorr (2022), "Types of AI, Explained", p. 39.
[129] Haesik (2022), "Historical Sketch of AI", p. 6.

and only fully exists as the concept of what it should be.[130,131] The roadblock of its development is the understanding part. Machines are already able to identify mental states but they do not understand what they have identified and why it is important.[132]

Self-awareness

Self-awareness is the ultimate goal of AI but it is far down the road since it only exists as a hypothetical concept.[133,134] The idea behind it is an AI that is not only capable of being conscious of others but also of itself. This means that such systems will be aware of themselves and their internal states.[135,136] The ability to understand, evoke and even feel emotions could put it at odds with human intentions.[137]

2.3 Ups and downs in the history of AI

In general, if a new technology gets discovered its life cycle follows the so-called Gartner Hype Curve, consisting of five stages.[138] First, the technology trigger happens, where researchers and developers start to pay attention to the new discovered technology.[139] Second, a peak of inflated expectations occurs due to implementation progress and its therefore obtained publicity.[140,141] Third, a trough of disillusionment follows due to exaggerated expectations.[142,143] During this stage people that are investigating into the technology are stumbling across disadvantages or problems which further results into frustration and disappointment.[144] Fourth, if solutions could be found for those challenges the technology experiences a slope of enlightenment since its potential can be seen again.[145] Otherwise it disappears.[146] Fifth, if the technology survived it is reaching its plateau of productivity where it gets established into society.[147,148]

Initially, the life cycle of AI seemed to be similar.[149] Since the term was coined (=

[130] Dorr (2022), “Types of AI, Explained”, p. 39.
[131] Haesik (2022), “Historical Sketch of AI”, p. 6.
[132] Dorr (2022), “Types of AI, Explained”, p. 39.
[133] Haesik (2022), “Historical Sketch of AI”, p. 6.
[134] Dorr (2022), “Types of AI, Explained”, p. 39.
[135] Dorr (2022), “Types of AI, Explained”, p. 39.
[136] Haesik (2022), “Historical Sketch of AI”, p. 6.
[137] Dorr (2022), “Types of AI, Explained”, p. 39.
[138] Haesik (2022), “Historical Sketch of AI”, p. 6.
[139] Haesik (2022), “Historical Sketch of AI”, p. 6.
[140] Haesik (2022), “Historical Sketch of AI”, p. 6.
[141] Schuchmann (2019), “Analyzing the Prospect of an Approaching AI Winter”, p. 19.
[142] Haesik (2022), “Historical Sketch of AI”, p. 6.
[143] Schuchmann (2019), “Analyzing the Prospect”, p. 19.
[144] Haesik (2022), “Historical Sketch of AI”, p. 6.
[145] Haesik (2022), “Historical Sketch of AI”, p. 6.
[146] Haesik (2022), “Historical Sketch of AI”, p. 6.
[147] Haesik (2022), “Historical Sketch of AI”, p. 6.
[148] Schuchmann (2019), “Analyzing the Prospect”, p. 19.
[149] Haesik (2022), “Historical Sketch of AI”, p. 7.

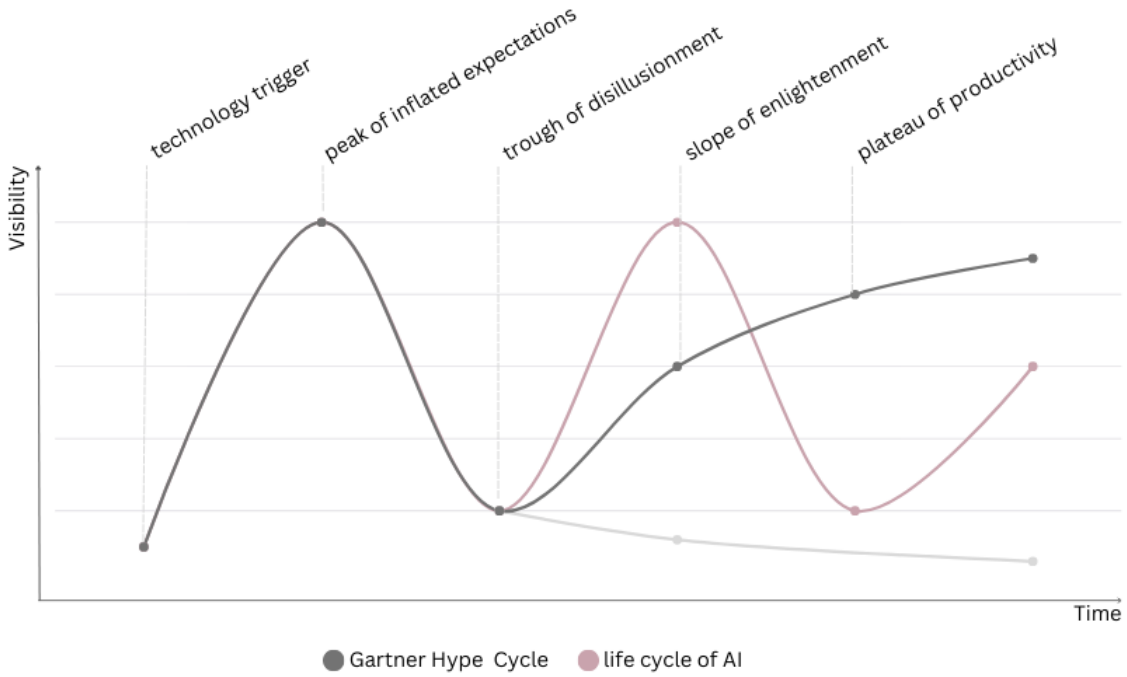

Figure 2.2: Gartner Hype Cycle compared to the life cycle of AI

technology trigger) a period full of optimistic predictions and massive investments started (= peak of inflated expectations).[150] Those exaggerated predictions and other technological limitations led to the first period of disappointment and reduced funding, also referred to as the first AI winter (= trough of disillusionment).[151] After some years the potential of the technology could be seen again and another period full of optimistic outlooks and increased funding started (= slope of enlightenment).[152] Normally, if a technology did not disappeared the next stage would be the establishment into society.[153] However, the life cycle of AI experienced a second trough of disillusionment, also known as the second AI winter.[154] Thus, the technology managed to face another slope of enlightenment and got established into society.[155] Due to today, the challenge of preventing another trough of disillusionment and keep the technology alive to benefit from its advantages still remains.[156,157]

[150] Mitchell (2021), "Why AI is Harder Than We Think", p. 1.
[151] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[152] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[153] Haesik (2022), "Historical Sketch of AI", p. 6.
[154] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[155] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[156] Harguess et al. (2022), "Is the Next Winter Coming for AI? Elements of Making Secure and Robust AI", p. 3.
[157] Haesik (2022), "Historical Sketch of AI", p. 9.

However, the first AI winter is a rather controversial topic as some are of the opinion that it never took place.[158,159] They argue that the first down in the history of AI corresponds to the standards set out by the Gartner Hype Curve[160] and further its time span was very short.[161] Still, it is possible to point out specific characteristics that are reflected in both down phases of AI.[162,163] During both periods a massive decrease in public interest and funding took place[164,165] which are still lingering in society as it negatively influenced the trustworthiness of AI. Therefore within this thesis, both down phases in the history of AI will be referred to as AI winters. Despite that, only the most important events in the history of AI will be discussed as within the scope of this thesis it is not possible to cover them completely.

2.3.1 Before the term was coined: 1955 and earlier

The first milestone in the history of AI was the invention of computers and robots as their accompanied opportunities arose the idea of creating machines that are able to decode language, make decisions and carry out targeted actions.[166] It was the trigger of investigating on the feasibility of AI even though the term itself did not yet existed.[167] At the same time, many science fiction authors have incorporated the potential power of AI into their stories.[168] Due to the autonomy and intelligence that could be achieved by this technology, these narratives often were exaggerated.[169] Nevertheless, in 1942 a science fiction author named Isaac Asimov published a short story named "Runaround" in which he introduced the three laws of robotics.[170,171] Up until now they are still relevant for ethical considerations about AI.[172]

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.[173,174]
2. A robot must obey the orders given it by human beings except where such orders

[158] Haigh (2024), “How the AI Boom Went Bust”, p. 22.
[159] Schuchmann (2019), “Analyzing the Prospect”, p. 6.
[160] Schuchmann (2019), “Analyzing the Prospect”, p. 6.
[161] Haigh (2024), “How the AI Boom Went Bust”, p. 22.
[162] Schuchmann (2019), “Analyzing the Prospect”, p. 6.
[163] Schuchmann (2019), “Analyzing the Prospect”, p. 16.
[164] Haesik (2022), “Historical Sketch of AI”, p. 3.
[165] Schuchmann (2019), “Analyzing the Prospect”, p. 16.
[166] Couch (2023), “Artificial Intelligence: Past, Present and Future”, pp. 1-2.
[167] Panesar (2020), “What is AI?”, p. 2.
[168] Haesik (2022), “Historical Sketch of AI”, p. 10.
[169] Russel et al. (2010), *A Modern Approach*, p. 1052.
[170] Couch (2023), “Past,Present and Furutre”, p. 2.
[171] Toosi et al. (2021), “How to Prevent Another Winter”, p. 4.
[172] Toosi et al. (2021), “How to Prevent Another Winter”, p. 4.
[173] Couch (2023), “Past,Present and Furutre”, p. 2.
[174] Haesik (2022), “Historical Sketch of AI”, p. 10.

would conflict with the First Law.[175,176]

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[177,178]

In 1943 the first step towards the implementation of AI was taken by inventing the McCulloch-Pitts Neuron, a mathematical model of an artificial neuron based on physiology and function of neurons in the brain.[179,180] Their inventors also stated that any computable function could be modeled as a network of such neurons.[181,182] However, this concept did not received much attention.[183] That changed after other people made use of it as it became an important milestone in the history of AI[184], especially for the invention of Artificial Neural Network[185]. Still the question remained whether machines were capable of thinking independently.

The next peak was reached in 1950[186,187,188,189] when Alan Turing dealt with the question of whether machines are able to think.[190] He circumvented it with the introduction of the imitation game[191], better known as the Turing Test[192,193,194], a behavioral intelligence test of machines.[195] The outcome depends on whether a human being is able to distinguish between a computer and a human being based on a question-answer game.[196,197] If the distinction was not possible the intelligence of machines could be considered.[198,199] This was followed by a period of many inventions that later on were recognized as AI as the term was officially coined.[200,201]

[175] Couch (2023), "Past,Present and Furutre", p. 2.
[176] Haesik (2022), "Historical Sketch of AI", p. 10.
[177] Couch (2023), "Past,Present and Furutre", p. 2.
[178] Haesik (2022), "Historical Sketch of AI", p. 10.
[179] Haesik (2022), "Historical Sketch of AI", p. 10.
[180] Hassanien et al. (2021), *Enabling AI Applications*, p. 151.
[181] Russel et al. (2010), *A Modern Approach*, p. 16.
[182] Toosi et al. (2021), "How to Prevent Another Winter", p. 5.
[183] Toosi et al. (2021), "How to Prevent Another Winter", p. 5.
[184] Toosi et al. (2021), "How to Prevent Another Winter", p. 5.
[185] Hassanien et al. (2021), *Enabling AI Applications*, p. 151.
[186] Turing (1950), "I.—COMPUTING MACHINERY AND INTELLIGENCE", pp. 433-434.
[187] Chen et al. (2023), "Systematic analysis of artificial intelligence in the era of industry 4.0", p. 94.
[188] Russel et al. (2010), *A Modern Approach*, p. 17.
[189] Panesar (2020), "What is AI?", p. 2.
[190] Panesar (2020), "What is AI?", p. 2.
[191] Turing (1950), "Computing Machinery and Intelligence", pp. 1021.
[192] Chen et al. (2023), "Systematic analysis", p. 94.
[193] Russel et al. (2010), *A Modern Approach*, p. 17.
[194] Panesar (2020), "What is AI?", p. 2.
[195] Turing (1950), "Computing Machinery and Intelligence", pp. 1021.
[196] Turing (1950), "Computing Machinery and Intelligence", pp. 433-434.
[197] Panesar (2020), "What is AI?", p. 2.
[198] Turing (1950), "Computing Machinery and Intelligence", pp. 433-434.
[199] Panesar (2020), "What is AI?", p. 2.
[200] Toosi et al. (2021), "How to Prevent Another Winter", p. 6.
[201] Russel et al. (2010), *A Modern Approach*, p. 35.

2.3.2 Golden years of AI: 1956 - 1973

John McCarthy and Marvin Minsky organised the Dartmouth conference in 1956, with the aim to establish a new field of research for machines that are able to simulate human intelligence.[202] To accomplish that, eleven scientists from various fields of research were invited.[203] They stated that "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it"[204,205,206] which led to the definition of AI as the science and engineering of making intelligent machines.[207] The term AI was finally coined[208,209,210] and the participants of the workshop are considered the founders of AI.[211,212] Although that conference created a lot of excitement about AI, still it largely was a theoretical concept.

However, after the term AI was coined almost two decades of significant achievements followed[213] and due to that, substantial funding was provided for its research.[214] In 1958 psychologist Frank Rosenblatt introduced the perceptron[215], an ANN architecture based on the McCulloch-Pitts Neuron.[216,217,218] More precisely, an algorithm that was able to classify images into two possible categories like dog or cat and women or man. It followed the connectionist approach of replicating human intelligence which is based on the brain that consists of billions of interconnected neurons that can jointly generate intelligence.[219] Therefore it was the the first attempt of replicating human-like activities and laid the foundation for ANN which later on enabled DL.[220]

Soon after that, in 1959 Newell and Simon which also attended at the conference, presented the General Problem Solver (GPS).[221] An algorithm that mimics the thinking activity of a human being in the process of solving a problem.[222] It was the first program that embodied the thinking humanly approach.[223] Afterwards, other programs that were also based on

[202] Heanlein et al. (2019), p. 7.
[203] Haesik (2022), "Historical Sketch of AI", p. 4.
[204] Hassanien et al. (2021), *Enabling AI Applications*, p. 104.
[205] Panesar (2020), "What is AI?", p. 2.
[206] Haesik (2022), "Historical Sketch of AI", p. 4.
[207] Haesik (2022), "Historical Sketch of AI", p. 4.
[208] Hassanien et al. (2021), *Enabling AI Applications*, p. 104.
[209] Panesar (2020), "What is AI?", p. 2.
[210] Haesik (2022), "Historical Sketch of AI", p. 4.
[211] Heanlein et al. (2019), p. 7.
[212] Toosi et al. (2021), "How to Prevent Another Winter", p. 6.
[213] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[214] Heanlein et al. (2019), p. 7.
[215] Kanal (2003), "Perceptron", p. 1383.
[216] Zamora-Cárdenas et al. (2020), "McCulloch-Pitts Artificial Neuron and Rosenblatt " s Perceptron : An abstract specification in Z", p. 16.
[217] Schuchmann (2019), "Analyzing the Prospect", p. 9.
[218] Russel et al. (2010), *A Modern Approach*, p. 836.
[219] Richbourg (2018), *Deep Learning: Measure Twice, Cut Once*, p. 3.
[220] Richbourg (2018), *DL: Measure Twice, Cut Once*, p. 3.
[221] Haocheng (2017), "A brief history and technical review of the expert system research", p. 1.
[222] Haocheng (2017), "history of expert system reseach", p. 1.
[223] Russel et al. (2010), *A Modern Approach*, p. 37.

the model of cognition were experiencing success as the GPS did which further led to the formulation of the famous physical symbol system hypothesis: "a physical symbol system has the necessary and sufficient means for general intelligent actions".[224,225] In 1963 Minsky proposed a simplified approach for AI research, called microworlds.[226,227] Its proposal was to concentrate on designing programs being capable of intelligent behaviour in smaller artificial environments[228] rather than on actual representation and reasoning in formal logic[229]. Therefore, microworlds only simulated a subset of the real-world.

As many breakthroughs were achieved during this period, many scientists from various disciplines gave enthusiastic outlooks on AI[230], including Marvin Minsky[231]. In 1970 he gave an interview to Life Magazine in which he stated that a machine with the general intelligence of an average human being could be developed within the next three to eight years.[232]

2.3.3 First AI Winter: 1973 - 1980

The optimistic outlooks given during the golden years of AI triggered a hype around AI, in which the media and the public had high expectations on. Those enthusiastic beliefs of experts stemmed from promising performances of early AI systems on simple tasks.[233] However, when those systems were applied to more difficult problems, they all failed.[234] Reason for that were three key factors:[235,236]

1. Many early AI systems were based on the approach of introspection of thinking humanly as the GPS was.[237,238] Therefore they were merely relying on the replication of the way human beings are performing a task.[239,240] That approach was missing a careful analysis of sequences of a task, defining an acceptable solution and the implementation of an algorithm that reliably produces such solutions.[241,242]

[224] Richbourg (2018), *DL: Measure Twice, Cut Once*, p. 3.
[225] Russel et al. (2010), *A Modern Approach*, p. 37.
[226] Toosi et al. (2021), “How to Prevent Another Winter”, p. 7.
[227] Russel et al. (2010), *A Modern Approach*, p. 38.
[228] Toosi et al. (2021), “How to Prevent Another Winter”, p. 7.
[229] Russel et al. (2010), *A Modern Approach*, p. 37.
[230] Russel et al. (2010), *A Modern Approach*, p. 20.
[231] Heanlein et al. (2019), p. 7.
[232] Heanlein et al. (2019), p. 7.
[233] Russel et al. (2010), *A Modern Approach*, p. 39.
[234] Russel et al. (2010), *A Modern Approach*, p. 39.
[235] Toosi et al. (2021), “How to Prevent Another Winter”, p. 8.
[236] Russel et al. (2010), *A Modern Approach*, pp. 39-40.
[237] Russel et al. (2010), *A Modern Approach*, p. 39.
[238] Toosi et al. (2021), “How to Prevent Another Winter”, p. 8.
[239] Russel et al. (2010), *A Modern Approach*, p. 39.
[240] Toosi et al. (2021), “How to Prevent Another Winter”, p. 8.
[241] Russel et al. (2010), *A Modern Approach*, p. 39.
[242] Toosi et al. (2021), “How to Prevent Another Winter”, p. 8.

2. The complexity of many problems that were attempted to be solved with AI was not recognized.[243] Reason for that was the oversimplification of microworlds which was introduced by Minsky.[244] To find a solution most early problem-solving systems tried out different combinations of steps until a suitable outcome was found.[245] That approach was perfectly suited for microworlds since they contain very view objects and therefore only a few possible actions.[246] It was assumed that scaling-up to more complex problems will be possible if faster hardware and larger memory capacity is provided.[247] However, developments in complexity theory have disproved this.[248]

3. Some fundamental limitations on basic structures that are being used to generate intelligent behaviour were found.[249,250,251] In 1969 Marvin Minsky pointed out the inability of a single-layer perceptron to implement the logical XOR function.[252,253,254] Additionally, he pointed out the lack of an effective learning algorithm for multi-layer networks.[255] An issue already known by Rosenblatt.[256] Although that finding did not applied to multilayered networks, funding for neural network research drastically decreased to almost nothing.[257,258]

As a result, those enthusiastic promises could not be kept which further lead to a major gap between the reality and the expected outcome.[259] In 1973 the Lighthill report was published[260,261,262], an evaluation of the current state of AI which publicly pointed out its inadequate performance.[263,264] It was the trigger for a significant decline in public interest and major cuts in funding of AI.[265,266] Those events drastically slowed down

[243] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.
[244] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.
[245] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.
[246] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.
[247] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.
[248] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.
[249] Russel et al. (2010), *A Modern Approach*, p. 40.
[250] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.
[251] Kanal (2003), "Perceptron", p. 1384.
[252] Russel et al. (2010), *A Modern Approach*, p. 40.
[253] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[254] Kanal (2003), "Perceptron", p. 1384.
[255] Russel et al. (2010), *A Modern Approach*, p. 836.
[256] Russel et al. (2010), *A Modern Approach*, p. 836.
[257] Russel et al. (2010), *A Modern Approach*, p. 40.
[258] Kanal (2003), "Perceptron", p. 1384.
[259] Heanlein et al. (2019), p. 7.
[260] Heanlein et al. (2019), p. 7.
[261] Schuchmann (2019), "Analyzing the Prospect", p. 11.
[262] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[263] Heanlein et al. (2019), p. 7.
[264] Schuchmann (2019), "Analyzing the Prospect", p. 11.
[265] Heanlein et al. (2019), p. 7.
[266] Toosi et al. (2021), "How to Prevent Another Winter", p. 8.

the advancement of AI[267] and therefore initiated the first AI winter which lasted until 1980[268,269].

2.3.4 The era of expert systems: 1980 - 1986

Soon before the first AI winter occurred it started to show that early attempts of AI were suiting perfectly fine for simple problems.[270] Although they were intended as general-purpose search mechanisms they lacked the ability to scale up to more complex problems and are therefore referred to as weak methods.[271] Alternatively to that, expert systems were introduced. They are utilized with domain-specific information for stronger reasoning but in narrower areas of expertise, represented mainly as if–then rules.[272] Researchers believed that expert systems are a more robust approach.[273]

In 1965 Ed Feigenbaum introduced the first effective knowledge-intensive system named DENDRAL.[274,275] It was a program to help chemists identify unknown organic molecules.[276] However, it was not until 1971 that Feigenbaum and other researchers started to investigate into the extent in which expert systems could be applied to other areas.[277] Another milestone was reached in 1972 with the invention of MYCIN, a system to diagnose blood infections.[278] It was considered at least as correct as some of the experts and significantly better than junior doctors.[279] Compared to DENDRAL it had to incorporate a factor of uncertainty which initially seemed to fit well with the way doctors assess the impact of evidence on diagnosis.[280]

These progresses achieved by expert systems brought out the commercial value of AI and marked the end of the first AI winter.[281,282] A new hype around AI was created and the industry boomed from a few million dollars in 1980 to billions of dollars in 1988.[283] In 1981, Japan started the Fifth Generation project, a ten year investigation plan into intelligent systems to keep up with the new boom of AI.[284,285] One year later, the US

[267] Heanlein et al. (2019), p. 7.
[268] (2022), "KI-Grundlagen", p. 12.
[269] Schuchmann (2019), "Analyzing the Prospect", p. 12.
[270] Russel et al. (2010), *A Modern Approach*, p. 40.
[271] Russel et al. (2010), *A Modern Approach*, p. 40.
[272] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[273] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[274] Haocheng (2017), "history of expert system reseach", p. 1.
[275] Russel et al. (2010), *A Modern Approach*, p. 40.
[276] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[277] Russel et al. (2010), *A Modern Approach*, p. 41.
[278] Russel et al. (2010), *A Modern Approach*, p. 41.
[279] Russel et al. (2010), *A Modern Approach*, p. 41.
[280] Russel et al. (2010), *A Modern Approach*, p. 41.
[281] Khan et al. (2021), "Advancements in Microprocessor Architecture for Ubiquitous AI - An Overview on History, Evolution, and Upcoming Challenges in AI Implementation", p. 5.
[282] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[283] Russel et al. (2010), *A Modern Approach*, p. 42.
[284] Khan et al. (2021), "History, Evolution, and Upcoming Challenges in AI", p. 5.
[285] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.

formed a consortium to ensure national competitiveness, named Microelectronics and Computer Technology Cooperation (MCC).[286,287] Finally in 1982 the first successful commercial expert system was developed by McDermott, named R1.[288,289] It was used in the digital equipment industry for configuring orders for new computer systems.[290,291] A turnover of $40 million dollars could be achieved by the company within almost four years.[292,293]

2.3.5 Second AI Winter: 1987 - 1993

Despite all funding and efforts made during the early 1980s[294], a second AI winter occurred in 1987[295]. Neither the Fifth Generation project nor the Strategic Computing Initiative met their ambitious goals.[296,297] Again, optimistic promises that had been made could not been kept.[298] Expert systems turned out to be difficult to maintain and build and therefore the industry declined sharply and finally collapsed.[299] Two key factor led to that collapse. First, expert systems lacked the ability to learn from their past experiences.[300] New or updated knowledge had to be implemented by actual experts which made it very difficult to keep them up to date.[301] Second, reasoning methods used by these systems broke down due to uncertainty.[302] Already in 1984 John McCarthy stated his considerations about the inability to incorporate common-sense knowledge into expert systems.[303] He highlighted it with an example were the application of MYCIN, the first expert system incorporating the certainty factor, would have led to the death on a patient.[304]

The AAAI conference that originally started in 1980 had attracted over 6000 visitors by 1986.[305] That drastically decreased by 1991 to 2000 visitors.[306] Similar to that, the lost of interest is also reflected in the amount of published AI-related articles which can be

[286] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[287] Russel et al. (2010), *A Modern Approach*, p. 41.
[288] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[289] Russel et al. (2010), *A Modern Approach*, p. 41.
[290] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[291] Russel et al. (2010), *A Modern Approach*, p. 41.
[292] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[293] Russel et al. (2010), *A Modern Approach*, p. 41.
[294] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[295] Haesik (2022), "Historical Sketch of AI", p. 8.
[296] Russel et al. (2010), *A Modern Approach*, p. 41.
[297] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[298] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[299] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[300] Russel et al. (2010), *A Modern Approach*, p. 42.
[301] Haigh (2024), "How the AI Boom Went Bust", p. 26.
[302] Russel et al. (2010), *A Modern Approach*, p. 42.
[303] Schuchmann (2019), "Analyzing the Prospect", p. 13.
[304] Schuchmann (2019), "Analyzing the Prospect", p. 13.
[305] Schuchmann (2019), "Analyzing the Prospect", pp. 12-14.
[306] Schuchmann (2019), "Analyzing the Prospect", pp. 12-14.

observed from the The New York Times.[307] A decrease started to show in 1987 and had its lowest point in 1995.[308] The reputation of the research field AI really suffered during the second AI winter and people were advised to not use this term anymore.[309,310] While many researchers hesitated to put effort into this field again some of them continued their work.[311] However, they used different banners, as NN and ML to follow the advise of not using the term itself.[312] Compared to the end of the first AI winter, the end of the second AI winter started to show in 1993 but was rather gradual. It was marked by progress made in the established subareas of AI including NN, ML and DL.

2.3.6 The return of neural networks: 1986 - present

The idea of back-propagation can be traced back to the 1960s when Rosenblatt attempted to create a multi-layer perceptron as well as a therefore suitable training algorithm.[313] However, he was not able to find a suitable solution and therefore failed at solving the problem that Minsky layed out.[314] Thus, the issue being well known[315], unfortunately the scientific field was no longer believed in and funding for neural networks was cut in the first AI winter. Despite the remarkable reduce in funding and interest, the work still continued at a lower pace.[316] As the limitations of expert systems have become apparent the attention shifted back to neural networks due to two important discoveries: Parallel Distributed Processing (PDP) and back-propagation.[317,318]

The research on back-propagation never fully stopped and at least four groups have reinvented the learning algorithm in the mid 1980s.[319] Back-propagation is a method for storing information to train the network as it is running.[320,321] That allows information to be used during the training phase and therefore enables the network to use errors in training deep learning models.[322] It was the first successful approach of enabling machines with the ability to learn. While it was applied to many learning problems in computer science and psychology the widespread dissemination of the results in the collection 'Parallel Distributed Processing', published in 1986 by Rumelhart and McClelland, caused great excitement.[323] The creation of PDP was also based on earlier research, however

[307] Schuchmann (2019), "Analyzing the Prospect", p. 14.
[308] Schuchmann (2019), "Analyzing the Prospect", p. 14.
[309] Mitchell (2021), "Why AI is Harder Than We Think", p. 2.
[310] Toosi et al. (2021), "How to Prevent Another Winter", p. 9.
[311] Khan et al. (2021), "History, Evolution, and Upcoming Challenges in AI", p. 6.
[312] Khan et al. (2021), "History, Evolution, and Upcoming Challenges in AI", p. 6.
[313] Kanal (2003), "Perceptron", p. 1383.
[314] Pires et al. (2023), "Artificial Neural Networks: History and State of the Art", p. 4.
[315] Russel et al. (2010), *A Modern Approach*, p. 836.
[316] Priyam, *The Evolution of Parallel Distributed Processing*, accessed on 6.05.2024.
[317] Russel et al. (2010), *A Modern Approach*, p. 42.
[318] Priyam, *The Evolution of PDP*, accessed on 6.05.2024.
[319] Pires et al. (2023), "ANN: History and State of the Art", p. 5.
[320] Kanal (2003), "Perceptron", p. 1384.
[321] Priyam, *The Evolution of PDP*, accessed on 6.05.2024.
[322] Russel et al. (2010), *A Modern Approach*, p. 42.
[323] Russel et al. (2010), *A Modern Approach*, p. 42.

only gained attention because of findings stated in the collection. It is an approach to ANN that emphasizes the parallel nature of the processing that occurs in the human brain.[324]

Both are so called connectionist models and can be seen as a competitor to symbolic models invented by Newell and Simon and to the logistic approach of McCarthy and others.[325] Philosophically, their discovery gave a new objective within cognitive psychology of whether human understanding relies on symbolic logic or distributed representations.[326] It might be that connectionist models are better suited for the messiness of the real world as they have the capability to learn.[327] The resurgence of interest into the field of ANN and connectionism laid the foundation for the development of DL and therefore transformed the field of AI.[328] These found developments are still relevant for many of the AI systems that are in use today.

2.3.7 Machine Learning as the rebirth of AI: 1987 - present

The invention of AI initially was a competitor against existing limitations of other fields like control theory and statistics.[329] However, as the brittleness of expert systems started to show, researchers were becoming more conservative and shifted their focus towards a new more scientific approach.[330] Already established theories like statistic-based methods were gaining on popularity.[331] Boolean logic was replaced by probability, hand-coding by machine learning and philosophical claims by experimental results.[332] Additionally, shared benchmark problem sets became the norm for demonstrating progress, since they allow to measure the performance of AI models and systems accurately.[333] One of the most important ones for further advancements in image object recognition was ImageNet.[334]

This pattern is well represented in the history of speech recognition.[335] In 1970s a variety of different approaches was realized which all were rather ad-hoc and fragile and only worked on a small number of selected examples.[336] In 1980s an approach based on mathematical theory which was trained on large corpus of real speech data dominated the field, named hidden Markov models (HMMs).[337] Therefore it was based on a collection of mathematical results from several fields over the last decades and due to the training

[324] Priyam, *The Evolution of PDP*, accessed on 6.05.2024.
[325] Russel et al. (2010), *A Modern Approach*, p. 42.
[326] Russel et al. (2010), *A Modern Approach*, p. 42.
[327] Russel et al. (2010), *A Modern Approach*, p. 42.
[328] Priyam, *The Evolution of PDP*, accessed on 6.05.2024.
[329] Russel et al. (2010), *A Modern Approach*, p. 43.
[330] Russel et al. (2010), *A Modern Approach*, p. 42.
[331] Russel et al. (2010), *A Modern Approach*, p. 42.
[332] Russel et al. (2010), *A Modern Approach*, p. 42.
[333] Russel et al. (2010), *A Modern Approach*, p. 42.
[334] Russel et al. (2010), *A Modern Approach*, p. 44.
[335] Russel et al. (2010), *A Modern Approach*, p. 43.
[336] Russel et al. (2010), *A Modern Approach*, p. 43.
[337] Russel et al. (2010), *A Modern Approach*, p. 43.

procedure HMMS improved their accuracy steadily.[338] However, advantages in DL later on proved to be better suited.[339]

1988 marked an important milestone for the reinforcement learning and the reunion of AI and other fields that initially were seen independent.[340] Judea Pearl published his book called 'Probabilistic Reasoning in Intelligent Systems' which further led the foundation for the acceptance of probability and decision theory in AI.[341] Within this book he presented Bayesian networks, an efficient formalism for representing uncertain knowledge as well as practical algorithms for probabilistic reasoning.[342] In the same year Rich Suttons connected reinforcement learning to the theory of Markov decision process which first was developed and applied in the field of operations research.[343] That led to the application of reinforcement learning in robotics and process control and further deep theoretical foundations were acquired.[344]

Consequentially to this increased interest in statistics, machine learning, data and optimization, sub-fields that were seen completely separate to AI started to reunite.[345] That led to remarkable advantages for the application as well as for the theoretical understanding of core problems of AI.[346]

2.3.8 The revolution of Big Data: 2001 - present

The world was experiencing a digital transformation as people and technology were no longer separable.[347] Computing power has made remarkable progress and the World Wide Web was invented which further enabled the creation of very large data sets, known as BD.[348] That initiated the development of learning algorithms especially making use of the advantages of these data sets since the majority of data in these sets are unlabeled.[349]

In the work of 1995 by Yarovsky on word-sense disambiguation, issues with the occurrence of many words existed.[350] In example, the word "plant" was difficult to classify since it was not labeled and therefore not knowable if it refers to flora or factory.[351] However, with the availability of big data, suitable learning algorithms could achieve an accuracy of 96% on the identification of the intended sense in a sentence.[352] A similar phenomenon

[338] Russel et al. (2010), *A Modern Approach*, p. 43.
[339] Russel et al. (2010), *A Modern Approach*, p. 43.
[340] Russel et al. (2010), *A Modern Approach*, p. 43.
[341] Russel et al. (2010), *A Modern Approach*, p. 43.
[342] Russel et al. (2010), *A Modern Approach*, p. 43.
[343] Russel et al. (2010), *A Modern Approach*, p. 43.
[344] Russel et al. (2010), *A Modern Approach*, p. 43.
[345] Russel et al. (2010), *A Modern Approach*, p. 43.
[346] Russel et al. (2010), *A Modern Approach*, p. 43.
[347] Reusch (2023), "Handlungsfähigkeit", p. 1.
[348] Russel et al. (2010), *A Modern Approach*, p. 44.
[349] Russel et al. (2010), *A Modern Approach*, p. 44.
[350] Russel et al. (2010), *A Modern Approach*, p. 44.
[351] Russel et al. (2010), *A Modern Approach*, p. 44.
[352] Russel et al. (2010), *A Modern Approach*, p. 44.

could be found in the field of computer vision by filling in holes of photographs.[353] A method was introduced by Hays and Efros that blended in pixels from similar images of a data set.[354] However that method worked poorly for data sets only consisting of thousands of images.[355] With the invention of big data this method finally crossed a threshold of quality.[356] Especially the availability of millions of images in the ImageNet database initiated a new revolution for the field of computer vision.[357]

Big Data and the shift towards machine learning were the trigger of regaining commercial attractiveness for AI. Up until today, BD is ubiquity[358] for AI, since data is the oil to fuel it.[359] Up until today a synergistic relationship between these terms exists.[360] AI requires massive amounts of data and BD is in the need of analytic and processing advantages enabled by AI.[361] If AI is taken into account, so is BD.[362]

2.3.9 Deep Learning as the return of neural networks: 2011 - present

The concept of deep neural networks had been around since the 1970s.[363] Back then they they could not be scaled up to a larger network, which therefore resulted in challenges that could not be solved with the current development state of AI.[364] In the 1990s they experienced success in handwritten digit recognition with the invention of convolutional neural networks.[365] However, it was not until 2011 that deep learning methods really took off.[366] Due to new achievements like fast parallel computing chips, big data and innovations in training methods they could finally be scaled up.[367] These newly gained capability of parallel processing enabled the training of large neural networks which are now known as deep neural networks.[368] Therefore, to fully gain advantages from deep learning powerful hardware, large amounts of data as well as a few algorithmic tricks are a necessity.[369]

The rise of deep learning first occurred in speech recognition and then in visual object recognition, a breakthrough that revolutionized the landscape of AI research and application.[370] Its success has gained back the interest in AI among students, companies,

[353] Russel et al. (2010), *A Modern Approach*, p. 44.
[354] Russel et al. (2010), *A Modern Approach*, p. 44.
[355] Russel et al. (2010), *A Modern Approach*, p. 44.
[356] Russel et al. (2010), *A Modern Approach*, p. 44.
[357] Russel et al. (2010), *A Modern Approach*, p. 44.
[358] Reusch (2023), "Handlungsfähigkeit", p. 1.
[359] Biore (2017), "Understanding AI", p. 41.
[360] Demigha (2020), *The impact of Big Data on AI*, p. 1399.
[361] Demigha (2020), *The impact of Big Data on AI*, p. 1399.
[362] Demigha (2020), *The impact of Big Data on AI*, p. 1399.
[363] Mitchell (2021), "Why AI is Harder Than We Think", pp. 2-3.
[364] Mitchell (2021), "Why AI is Harder Than We Think", pp. 2-3.
[365] Russel et al. (2010), *A Modern Approach*, p. 44.
[366] Russel et al. (2010), *A Modern Approach*, p. 44.
[367] Mitchell (2021), "Why AI is Harder Than We Think", pp. 2-3.
[368] Mitchell (2021), "Why AI is Harder Than We Think", pp. 2-3.
[369] Russel et al. (2010), *A Modern Approach*, p. 45.
[370] Hassanien et al. (2021), *Enabling AI Applications*, p. 154.

investors, governments, media and society.[371] Up to today deep neural network are the magic behind all the major advantages in AI that we have seen in the past decades. That includes speech recognition, machine translation, chat bots, image recognition, game playing and many other things.[372]

2.4 Prevention of another AI winter

The recurrence of another AI winter is not unpredictable. AI winters are characterized by a loss of confidence in the technology which leads to a loss of interest from society and government and further to reduced funding.[373,374] Preceding to that is a period of big hype and enthusiasm about AI.[375] Throughout its history important factors have been crystallized that are contributing to the prevention of another AI winter as well as its ongoing development.

The importance of realistic outlooks. Optimistic forecasts given by experts are raising high expectations in society. These forecasts are often based on the excitement about a breakthrough that has been reached. Since initially it was thought that the solved problem is the missing piece of success an oversimplification of the remaining challenges might happen and enthusiastic promises are being made. For this reason, these forecasts are often not fulfilled in the promised time which leads to unmet expectations by society and government. Further, the resulting disappointment leads to society avoiding the application of this technology and a decrease in funding. This disappointments are often lingering even after a hype was reached and therefore are creating concerns about the technology. With realistic forecasts enough time is given to complete the research and development of AI and show its true potential without causing disappointment.

The importance of funding. Funding is the foundation of AI since it is fueling its research and development. Without the provided funding for its innovation, research projects, computing resources, infrastructure, commercialization and considerations of ethical and legal aspects, the field of AI would not be where it is now. Even if a technology is faced with a loss of interest it is of great importance to still provide funding. First, the reason of this setback needs to be analyzed in order to give an estimation about its future potential. Only if no recognizable potential can be found a reduction in funding is justified. This goes hand in hand with the importance of giving realistic outlooks on the future development of a technology. A realistic outlook does not lead to unfulfilled expectations and funding might not be decreased.

[371] Russel et al. (2010), *A Modern Approach*, p. 45.
[372] Mitchell (2021), "Why AI is Harder Than We Think", p. 3.
[373] Haesik (2022), "Historical Sketch of AI", p. 3, 7.
[374] Schuchmann (2019), "Analyzing the Prospect", p. 16.
[375] Schuchmann (2019), "Analyzing the Prospect", p. 16.

The importance of interrelated disciplines. A variety of disciplines like philosophy, mathematics, economics, neuroscience, psychology, computer science, cybernetics and linguistics are the origin of the research field AI. Concepts found by these disciplines are the base of the development of AI and therefore progress made in these fields enable its further progress. Thus, maintaining research and development in other subareas is of the same importance as fostering AI. Big Data is one of the most important discipline linked to AI since it unleashed its potential and strongly contributed to its current hype.

The importance of subareas. No globally accepted definition of AI exists up to today which makes it hard to grasp the full scope of AI. Thus, it is well known that AI is characterized by its complexity, opacity and autonomy. These factors create the impression of an uncontrollable extent of power that might be achieved by this technology and therefore it is not gaining the trust of society. The establishment of subareas made it possible to define their scope more comprehensively and therefore technologies of these subareas were more tangible and welcomed. Additionally, as AI faced a period of aversion after the second AI winter those subareas enabled the reestablishment of this technology.

The importance of categorizations. Complex topics are often difficult to define in a way that all important aspects are covered. This makes research more difficult as perspectives are broad but very vaguely defined. AI has been faced with that issue and to overcome it different categorizations have been introduced over time. Each category of such a categorization has a precisely defined perspective and therefore it can be invested on in more detail. These established categorizations also allow to classify the progress of AI, which enables a better understanding of the current state of the technology and its potential future development. Different approaches were made to categorize AI and each of them is of importance to reach a comprehensive definition of AI due to their findings.

CHAPTER 3

The future goal of achieving trustworthy AI

Ever since the term AI was coined its research and development faced several waves of rapid progress, as discussed in *Ups and downs in the history of AI*. Especially the establishment of the two subareas ML and DL as well as the progress in BD fueled the idea of a reality where advantages of AI might push the wellbeing and prosperity of individuals, organizations and societies to a new level.[1] However, it started to show that it not only brings up great new possibilities, it is also accompanied by a variety of novel ethical, legal and social challenges.[2] To overcome them multiple calls were made to establish requirements, guidelines and regulations towards a safer development and application of AI.[3,4,5] Those approaches are referred to as beneficial AI, responsible AI or ethical AI.[6,7] Irrespective of the terminology they all aimed for the same goal: shaping the progress of AI to maximize its benefits while its risks and dangers are mitigated or prevented.[8,9] For that reason, in early 2019 the High-Level Expert Group on Artificial Intelligence published their 'Ethics guidelines for trustworthy AI'.[10,11,12]

[1] Thiebes et al. (2021), "Trustworthy artificial intelligence", p. 448.
[2] Kaur et al. (2022), "Trustworthy Artificial Intelligence: A Review", pp. 4-5.
[3] Thiebes et al. (2021), "Trustworthy AI", p. 448.
[4] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", pp. 6-7.
[5] Kaur et al. (2022), "Trustworthy AI: A Review", p. 2.
[6] Thiebes et al. (2021), "Trustworthy AI", p. 448.
[7] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", pp. 6-7.
[8] Thiebes et al. (2021), "Trustworthy AI", p. 448.
[9] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", pp. 6-7.
[10] Thiebes et al. (2021), "Trustworthy AI", p. 448.
[11] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 2.
[12] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 4.

3.1 The importance of trustworthiness

Trustworthy has its origins in the word trust, which is defined as the "firm belief in the reliability, truth, or ability of someone or something" according to the Oxford English Dictionary or the "belief that you can depend on someone or something" according to the Dictionary of Cambridge.[13] Trust is essential for a sustainable development of society as it sets the base for good relationships.[14] Nobody can control their external environment and therefore potential dangers will always exist.[15] Allowing ourselves to trust our environment enables the continuous interaction with it, although those risks remain.[16]

Not only relationships between human beings are heavily depending on trust, it is also a key factor for the relationship of human beings and technology.[17] Without trust in a technology, people would try to avoid its application.[18] That further impedes to make use of the advancements a technology entails.[19] Therefore, in order to gain advantages of a technology it is indispensable to ensure its trustworthiness.[20] Further, ensuring trustworthiness is a prerequisite to enable the development, deployment and use of AI.[21] Because of that the European Union strives to turn into a hub for trustworthy AI to position itself as a global leader.[22]

3.2 Ethics guidelines for trustworthy AI

With the aim of promoting trustworthy AI 'Ethics Guidelines for Trustworthy AI' were published by the High-Level Expert Group on Artificial Intelligence.[23] According to the group trustworthy AI can be narrowed down to three components: being lawful, ethical and robust through its whole life cycle.[24] In more detail it must comply with all applicable laws and regulations as well as respecting given ethical principles and values while its robustness from a technical perspective taking into account its social environment must be given at anytime in order to prevent unintentional harm.[25]

Based on the European Charter of Fundamental Rights (CFR) a framework for achieving the goal of trustworthy AI was introduced within these guidelines.[26] It has the focus on

[13] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 4.
[14] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 4.
[15] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 4.
[16] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", pp. 4-5.
[17] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 5.
[18] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 5.
[19] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 5.
[20] Liu et al. (2022), "Trustworthy AI: A Computational Perspective", p. 5.
[21] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 4.
[22] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 5.
[23] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 2.
[24] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 5.
[25] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 5.
[26] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 2.

trustworthy AI		
(1) lawful	**(2) ethical**	**(3) robust**
respecting all applicable laws and regulations	respecting ethical principles and values	both from a technical perspective while taking into account its social environment to prevent unintended harm

Table 3.1: Three components of trustworthy AI

the latter two components: fostering and securing ethical and robust AI.[27] Therefore it does not explicitly deal with the component of being lawful[28], as the guidelines hold onto the assumption that all legal rights and obligations that apply throughout the life cycle of AI remain binding and must continue to be complied with.[29] In order to ensure ethical behaviour and robustness the framework layed out the foundation of trustworthy AI by identifying four ethical principles based on fundamental rights that must be adhered to:[30]

1. Principle of respect for human autonomy: AI should never be used to manipulate or otherwise inappropriately guide human beings.[31] The CFR specifies the importance of respecting the freedom and autonomy of human beings.[32] Therefore users must always be able to self-determine over themselves, and be able to partake in any democratic process.[33]

2. Principle of prevention of harm: AI should never be used to cause or exacerbate any harm nor any other adverse effect on human beings.[34] AI systems as well as their application environment must be safe and secure in a way that they are technically robust and malicious use can be ruled out.[35]

3. Principle of fairness: Individuals and groups must be free from unfair bias, discrimination and stigmatisation and further it must be possible to challenge and effectively attack decisions made by AI systems and humans operating them.[36]

4. Principle of explicability: Capabilities and the purpose of AI systems need to

[27] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 2.
[28] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 2.
[29] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 6.
[30] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 7.
[31] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 12.
[32] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 12.
[33] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 12.
[34] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 12.
[35] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 12.
[36] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, pp. 12-13.

be openly communicated while processes must be transparent in order to enable the explainability of decisions made.[37] The black-box problem is a special case for this principle which might require other explicability measures.[38] Explicability is essential to enable users trust in AI.[39]

These principles were translated into key requirements that AI systems should implement through their entire life cycle.[40] If these key requirements are met an AI system deems to be trustworthy in the aspect of ethical and robust. These key requirement are:[41]

1. Human agency and oversight: As the principle of respect for human autonomy states, the application of any AI system should support human autonomy and decision-making.[42] To act as an enabler for a fair and democratic society AI is required to support users in their ability to act while at the same time promoting their fundamental rights as well as allowing for human oversight.[43]

2. Technical robustness and safety: Technical robustness is closely linked to the principle of prevention of harm and crucial for achieving trustworthy AI.[44] Any application of AI needs to be resilient, reliable and secure.[45] To ensure their intended behaviour, the development of AI must incorporate a preventative approach to risks with the aim of minimizing unintentional and unexpected harm while preventing from unacceptable harm.[46]

3. Privacy and data governance: Privacy, a fundamental right especially challenged by AI, is closely linked to the principle of prevention of harm.[47] It is necessary for AI systems to fully respect privacy and data protection through their whole life cycle.[48] Further, to prevent any harm related to privacy, adequate governance in terms of privacy and data protection, quality, integrity and access to data must be provided.[49]

4. Transparency: Transparency is closely linked to the principle of explicability and also refers to elements relevant for the development and application of any AI system like data, the system itself and involved business models.[50] Characteristics like the opacity, complexity and autonomy of AI often makes it difficult to retrace all steps

[37] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 13.
[38] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 13.
[39] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 13.
[40] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 14.
[41] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 14.
[42] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 15.
[43] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 15.
[44] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 16.
[45] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 16.
[46] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 16.
[47] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 17.
[48] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 17.
[49] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 17.
[50] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.

of the decision making process and are therefore not transparently communicable.[51] Therefore suitable traceability mechanisms are needed to overcome the issue of AI systems and their decision not being explainable.[52] Further, users must be aware of their use of AI systems.[53]

5. Diversity, non-discrimination and fairness: Closely linked with the principle of fairness is the necessity of avoiding unfair bias.[54] AI systems must act fair in consideration of all different groups and cultures of society.[55] Vulnerable groups should not be marginalized as well as the exacerbation of discrimination and prejudice must be prevented.[56] To foster diversity, AI systems must be accessible to anybody regardless of any disability.[57]

6. Societal and environmental well-being: Closely linked to the principles of fairness and prevention of harm is the necessity of AI systems being sustainable, environmentally friendly and benefiting all human beings, including future generations.[58] Therefore, human beings, the environment as well as other sentient beings should be taken into account as stakeholders.[59]

7. Accountability: This requirement is closely linked to the principle of fairness.[60] The autonomy of AI often makes it difficult to define anybody as responsible or accountable for their decisions.[61] Therefore it is important to set up mechanisms that ensure responsibility and accountability through the whole life cycle of AI regardless of decisions made are being correct or incorrect.[62]

Additionally, the group proposed technical and non-technical methods for ensuring the compliance of these key requirements.[63] They even introduced a non-technical approach to operationalize these requirements, called 'Assessment List of Trustworthy AI'.[64] Finally, the AI HLEG discussed some critical concerns raised by AI that arise if any component of trustworthy AI gets violated.[65] Partially they are already covered by European law.[66] However, the compliance with legal requirements does not rule out that the existing legal framework might not address the full range of ethical concerns that might arise.[67]

[51] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.
[52] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.
[53] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.
[54] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.
[55] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.
[56] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.
[57] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 18.
[58] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 19.
[59] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 19.
[60] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 19.
[61] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 19.
[62] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 19.
[63] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 20.
[64] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, pp. 24-31.
[65] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 33.
[66] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 33.
[67] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 33.

The discussed concerns includes the identification and tracking of individuals with AI, human beings using AI without their awareness, the violation of fundamental rights by AI enabled citizen scoring, lethal autonomous weapon systems and currently unknown concerns that arise due to the factor of uncertainty.[68]

3.3 The crucial interrelation of ethic and law

As the 'Ethics Guidelines for Trustworthy AI' exposed, ethic and law are two essential players in achieving the goal of trustworthy AI.[69] These two terms themselves have a crucial relationship between each other as their interplay enables the ongoing development and refinement of both systems.[70] Thus, their interrelation is also the main reason why these terms are often being confused or even viewed as the same.[71] Since both fields of knowledge are important for shaping human behaviour, societal norms and the functioning of the legal system, merging these fields might not seem critical. However to establish a balance of individual freedoms, cultural diversity and the need for societal order and justice their distinction is of great importance.

Both fields are regulating relationships between citizens themselves as well as citizens and the state with the aim of a peaceful coexistence between all human beings.[72,73,74] Also both are serving as a guideline of behavior for society by upholding a set of moral values and benefiting people from adhering to them in order to prevent violations.[75] If both concepts have the same goal and both pursue it with a set of moral rules by which society should behave, then how does ethics differ from law? Ethics is more of an internal system, while law is an external control mechanism.[76] More precisely, ethics are internal guidelines about how people should act, while law sets out external rules which people must be followed.[77,78]

> *Ethics* is the study of moral principles and values that are steering the behaviour of human beings.[79] These rules of conduct support human beings in their decision on what is good and bad, right or wrong and morally justifiable or not.[80] Also it serves as a guideline on how people should live and interact with each other.[81] These principles and values are agreed on and adopted by each individual itself and therefore can

[68] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, pp. 33-35.
[69] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 5.
[70] Pollmaecher (2022), "The DGPPN congress 2022: Ethics, law and mental health", p. 1091.
[71] Key Differences, *Difference between Law and Ethics*, accessed on 28.04.2024.
[72] Pollmaecher (2022), "DGPPN congress 2022", p. 1091.
[73] Tzafestas (2018), "Ethics and Law in the Internet of Things World", p. 105.
[74] Gundugurti et al. (2022), "Ethics and Law", p. 7.
[75] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[76] Gundugurti et al. (2022), "Ethics and Law", p. 7.
[77] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[78] Key Differences, *Difference: Law and Ethics*, accessed on 16.01.2023.
[79] Tzafestas (2018), "Ethics and Law in the IoT World", p. 102.
[80] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[81] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.

derive due to cultural, religious, philosophical, or individual perspectives.[82] They might evolve over time due to changing societal norms and values and are therefore dynamic. By a certain degree they are restricted by the official law.

	Ethic	Law
Definition	Ethics are a collection of fundamental concepts and principles of an individual human character. It branches from moral philosophy and helps to classify actions or thoughts into good and bad.	Law refers to a systematic body of rules and regulations. It controls and react to the actions and interactions of and between individual members of its society.
Objective	Ethics are a guideline to help people decide what is right or wrong with the goal to act in the most possible appropriate way.	Law has the intention to maintain a peaceful social order and provide the same amount of protection to all citizens.
Who is it governed by?	Ethics are governed by each individual because of their individual values and beliefs but also by legal and professional norms.	Law is governed by the government and differs for each country/state.
What is it?	Ethics is a set of guidelines.	Law is a set of rules and regulations.
How is it expressed?	Ethics are only abstract and differ because of individual values of different people.	Expressed and published in writing by the government.
What happens when violated?	There is no punishment for violations of ethics if they do not intersect with a violation of law.	Violation of law is prohibited and may lead to punishment like imprisonment or fine or both.
Is it binded?	Ethics do not have a binding nature.	Law has legal binding.

Table 3.2: Comparison of the terms ethic and law

Law establishes normative rules and regulations with the aim to maintain order,

[82] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.

resolve conflicts and protect rights and interests of individuals.[83] It clearly specifies how people must or must not behave and is enforced by imposing penalties, punishments or both.[84] By that it prevents people from behaving in a way that negatively affects another persons life. These rules and regulations are set up by the government and are obligatory to each person in this country.[85] Therefore each person is accountable to the same laws which further should protect fundamental human rights and undermine the foundation of the rule of law.[86] Although laws can be amended as part of legislative procedures, they are quite stable and ensure predictability and consistency in the legal system. By a certain degree they are shaped by ethical principles.

Since ethics is a guideline for people to behave appropriately, and law is the system to control their behavior, ethics should be the foundation on which law arises.[87] However, ethics is based on each individuals awareness of what is right or wrong and may vary widely.[88] For that reason law is crucial to protect human fundamental rights by establishing a minimum standard of ethical codes of conduct for society.[89] Nevertheless, ethics often exceeds the legal minimum and therefore goes way beyond law.[90] Values of society are continuously changing which implies a steady movement in the ethical point of view. That results in the need of constant adaptions to the legal framework.[91] However, amending or establishing laws is a time consuming task which makes it almost impossible to harmonise current ethical and legal positions.[92] As a consequence of these occurrences, not every legal decision is ethically correct and vice versa.[93,94] Neither law and justice nor lawful and correct actions entail each other which creates a constant tension between ethics and law.[95]

In conclusion, ethics is the base on which law arises while law is the regulation to enable ethically justifiable coexistence. While ethics often exceeds the legal minimum and therefore goes far beyond the law, law at least governs situations arising from different established ethical values.[96] To achieve the fairest and most equitable behaviour ethics and law are essential and therefore both should be considered and respected.[97,98]

[83] Tzafestas (2018), "Ethics and Law in the IoT World", p. 98.
[84] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[85] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[86] Tzafestas (2018), "Ethics and Law in the IoT World", p. 106.
[87] Pollmaecher (2022), "DGPPN congress 2022", p. 1091.
[88] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[89] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[90] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[91] Pollmaecher (2022), "DGPPN congress 2022", p. 1091.
[92] Pizzi (2020), "AI for humanitarian action: Human rights and ethics", p. 167.
[93] Pollmaecher (2022), "DGPPN congress 2022", p. 1091.
[94] Tzafestas (2018), "Ethics and Law in the IoT World", p. 105.
[95] Pollmaecher (2022), "DGPPN congress 2022", p. 1091.
[96] Tzafestas (2018), "Ethics and Law in the IoT World", p. 98.
[97] Tzafestas (2018), "Ethics and Law in the IoT World", p. 98.
[98] Gundugurti et al. (2022), "Ethics and Law", p. 7.

3.4 The need for a legal framework addressing AI

The importance of trustworthiness declared the necessity of ensuring trustworthiness in the context of AI. To achieve that goal the AI HLEG published '*Ethics guidelines for trustworthy AI*'. Within this publication they stated the importance of being lawful, ethical and robust in achieving trustworthiness, thus, they did not incorporate the part of being lawful as they hold onto the assumption that all legal rights and obligations remain binding and must continue to be complied with. *The crucial interrelation of Ethic and Law* as well stated the importance of considering both, ethic and law to achieve the fairest and most equitable behaviour.

To a certain extent the current legal framework of the European Union regulates the use of AI, however it is insufficient when it comes to specific challenges brought up by the characteristics of AI.[99] Its opacity, autonomy and complexity poses new, previously unregulated challenges for the existing framework resulting into loopholes in the current in force law.[100] Notwithstanding the fact that most AI systems pose limited or no risks, certain AI systems have a higher potential for risks which might lead to harmful, undesirable outcomes. The mitigation and prevention of such risks must be legally covered to enable trustworthiness for the users of such systems to further enable the ongoing application of AI.

[99] Sartor (2020), "Artificial intelligence and human rights: Between law and ethics", p. 712.
[100] (2023), "Reconciling Artificial Intelligence (AI) With Product Safety Laws", p. 1.

CHAPTER 4

Applicable legal acts and concerns posed by AI

The need for a legal framework addressing AI stated the necessity of establishing a legal framework regulating the application of AI as the current in force law is only partially capable of doing so. Reason for that is the lack of covering new arising challenges brought up by the unique characteristics of AI. Within the scope of this thesis it is not possible to discuss all relevant legal sources regarding the application of AI. Therefore, only the most relevant primary and secondary laws will be explained as well as their potential concerns arising through the uniqueness of AI.

The most important primary law is the European Charter of Fundamental Rights (CFR), as it already was the foundation of the established '*Ethics guidelines for trustworthy AI*. Further, two most important secondary laws were selected. Since data is the oil to fuel AI it is of great importance to inspect the existing most relevant data protection law, the European General Data Protection Regulation (GDPR). Finally, AI is raising concerns regarding its potential of harm and therefore persons affected by that must be appropriately compensated which is regulated by the European Product Liability Directive (PLD). *European Union's journey towards trustworthy AI* is going to underpin the importance of the CFR, the GDPR and the PLD in AI governance even more.

4.1 European Charter of Fundamental Rights (CFR)

Due to progress in society, scientific and technological developments and the expansion of policies in the EU, a need for a legislation act on fundamental rights legally binding for every member state was given. The European Charter of Fundamental Rights was

declared in 2000[1] and came into force in December 2009.[2] Its aim is to protect and promote individuals rights and freedoms as well as to protect individuals against the power of organizations.[3] That is achieved by providing rights for individuals that they can assert against the state or an organ of the state.[4] 54 articles build the framework of protecting human rights by addressing and regulating the topics dignity, freedom, equality, solidarity, citizens' rights and justice.[5] Due to the complexity, limited interpretability, accompanied bias and degree of autonomy, AI has the potential to interfere with these fundamental rights regardless of the field of application.[6,7] It is not possible to cover all interference of AI and the CFR but some critical ones will be discussed below.

4.1.1 Dignity

The core value of the CFR is human dignity since it constitutes the foundation of all fundamental rights besides being a fundamental right itself.[8] Potential concerns of AI to dignity can be clearly seen through three dimensions of possible violations.[9] First, a violation by humiliating any individual as putting them in a state of losing autonomy over their own representation.[10] Second, a violation through instrumentalization as treating individuals as interchangeable and merely as a means to an end.[11] Finally, a violation by rejecting an individual's gift as treating individuals as superfluous without recognizing their contributions, aspirations and potentials.[12]

AI systems must process personal data in respect of human dignity since art. 1 insists that human dignity is inviolable and must be respected and protected at any time.[13,14] While society is structured and governed with the focus on human beings, AI systems might not be designed to enhance human dignity unless it is somehow ensured.[15] Additionally, the right to life[16] and the right to the integrity of the person must be respected.[17] As discussed in section 4.1.2, AI bears the risk to perpetuate existing discrimination. This can lead to a violation of human dignity, the right to life or the right of integrity

[1] European Union, *Charter of Fundamental Rights of the European Union*, last pahge.
[2] ENNHRI, *Implementation of the EU Charter of Fundamental Rights*, p. 2.
[3] Janssen et al. (2022), "Practical fundamental rights impact assessments", p. 201.
[4] Janssen et al. (2022), "Practical FRIA", p. 208.
[5] European Union, *Charter of Fundamental Rights*.
[6] Janssen et al. (2022), "Practical FRIA", p. 201.
[7] Gerards et al. (2020), *Getting the future right – Artificial intelligence and fundamental rights – Report*, p. 7.
[8] Aizenberg et al. (2020), "Designing for human rights in AI", p. 5.
[9] Aizenberg et al. (2020), "Designing for human rights in AI", p. 5.
[10] Aizenberg et al. (2020), "Designing for human rights in AI", p. 5.
[11] Aizenberg et al. (2020), "Designing for human rights in AI", p. 5.
[12] Aizenberg et al. (2020), "Designing for human rights in AI", p. 5.
[13] European Union, *Charter of Fundamental Rights*, art. 1.
[14] Gerards et al. (2020), *Getting the future right*, p. 60.
[15] Donahoe et al. (2019), "Artificial Intelligence and Human Rights", p. 217.
[16] European Union, *Charter of Fundamental Rights*, art. 2.
[17] European Union, *Charter of Fundamental Rights*, art. 3.

throughout the decision processes of an AI system.[18]

4.1.2 Equality

Title III is probably the most critical chapter of the CFR when it comes to the application of AI.[19] It enshrines that everyone is equal before the law[20] and at its heart aims to eliminate any nature of discrimination[21,22]. AI systems challenge these laws because of their potential to encode discriminatory biases.[23] Such systems are trained on data collected from the real world which might be biased.[24] Therefore, underrepresented groups, existing inequalities and prejudice, racism and many things more might be maintained by AI systems.[25] Especially in domains that already have a history of discrimination this is crucial.[26] AI can exacerbate any underlying societal problems and inequalities.[27]

4.1.3 Freedoms

The title of freedom is closely linked to human dignity.[28] Art. 6 lays out the right for liberty and security of any person.[29] The use of predictive policy systems and recidivism risk assessments might falsely label persons as high risk because of their demographics correlating with any data of previously arrested persons.[30] Art. 7 enshrines the right for respect for private life and family[31] while art. 8 enshrines the protection of personal data[32]. They are closely related to each other and set the foundation for other fundamental rights[33] like the freedom of thought, conscience and religion[34], the freedom of expression and information[35] and freedom of assembly and of association.[36] The application of AI often implies automated processing of large amount of data which interferes with art. 8 as well as art. 7.[37] Finally, art. 11 enshrines the right of sharing and obtaining information without any impairment.[38] In example, personalization algorithms are applied in news

[18] Gerards et al. (2020), *Getting the future right*, p. 60.
[19] Gerards et al. (2020), *Getting the future right*, p. 68.
[20] European Union, *Charter of Fundamental Rights*, art. 20.
[21] European Union, *Charter of Fundamental Rights*, art. 21-26.
[22] Aizenberg et al. (2020), "Designing for human rights in AI", p. 8.
[23] Janssen et al. (2022), "Practical FRIA", p. 205.
[24] Janssen et al. (2022), "Practical FRIA", p. 205.
[25] Janssen et al. (2022), "Practical FRIA", p. 205.
[26] Janssen et al. (2022), "Practical FRIA", p. 205.
[27] Janssen et al. (2022), "Practical FRIA", p. 205.
[28] Aizenberg et al. (2020), "Designing for human rights in AI", p. 7.
[29] Aizenberg et al. (2020), "Designing for human rights in AI", p. 7.
[30] Aizenberg et al. (2020), "Designing for human rights in AI", p. 7.
[31] European Union, *Charter of Fundamental Rights*, art. 7.
[32] European Union, *Charter of Fundamental Rights*, art. 8.
[33] Gerards et al. (2020), *Getting the future right*, p. 61.
[34] European Union, *Charter of Fundamental Rights*, art. 10.
[35] European Union, *Charter of Fundamental Rights*, art. 11.
[36] European Union, *Charter of Fundamental Rights*, art. 12.
[37] Gerards et al. (2020), *Getting the future right*, p. 62.
[38] Aizenberg et al. (2020), "Designing for human rights in AI", p. 8.

recommender systems.[39] Taking into account the ongoing debate in balancing freedom of speech against any kind of hate speech or disinformation the application of these AI based systems might cause a potential violation of this right.[40]

4.1.4 Solidarity

The impact of AI technologies on social protection systems and the lives of individuals relying on them can potentially be very problematic which is becoming more and more apparent.[41] Social security and social assistance is enshrined by art. 34 as respecting their ability to exercise any individuals rights and therefore uphold their dignity by providing protection in example in the case of maternity, illness as well as industrial accidents.[42] The application of AI based on statistical correlations might judge based on data collected from population compared to the data of this individual.[43] These special circumstances of an individual might not be considered within this comparison and therefore might lead to an unfair decision.[44] In example if an individual is newly moving to the European Union and applies for a new job an unfair decision might be taken based on the lack of data about their earlier job history.[45] Further, targeted advertising supported by AI must implicate the consumers awareness of their option to op-out. If that is not ensured consumers might be faced with unwanted advertising eventually leading to manipulation of the consumers preferences which further conflicts with art.38 of consumer protection.

4.1.5 Citizen's Rights

Under art. 41 the right to good administration is enshrined, including art. 41(2)(b) the right of an individual to inspect his or her file as well as art. 41(2)(c) the obligation of the authority to provide adequate reasons for its decisions.[46,47] Applying AI systems in the context of administrative procedures not only enables improvements in analytic abilities and decision making processes.[48,49] Also questions arise on how to ensure that individuals have access to their files as the number is potentially high and how to ensure that the obligation of authorities to give sufficient reasons is fulfilled despite the lack of transparency of AI systems.[50]

[39] Aizenberg et al. (2020), "Designing for human rights in AI", p. 8.
[40] Aizenberg et al. (2020), "Designing for human rights in AI", p. 8.
[41] Gerards et al. (2020), *Getting the future right*, p. 79.
[42] Aizenberg et al. (2020), "Designing for human rights in AI", p. 8.
[43] Aizenberg et al. (2020), "Designing for human rights in AI", p. 8.
[44] Aizenberg et al. (2020), "Designing for human rights in AI", p. 8.
[45] Gerards et al. (2020), *Getting the future right*, p. 79.
[46] European Union, *Charter of Fundamental Rights*, art. 41.
[47] Wróbel (2022), "Artificial intelligence systems and the right to good administration", pp 213-214.
[48] Wróbel (2022), "AI systems and the Art. 41 CFR", p. 217.
[49] Gerards et al. (2020), *Getting the future right*, p. 81.
[50] Gerards et al. (2020), *Getting the future right*, p. 81.

4.1.6 Justice

One of the most used CFR right in legal proceedings is art. 47 the right to an effective remedy before a tribunal and to a fair trial.[51,52] New challenges arise since decisions taken by AI are not excluded.[53] Due to the lack of transparency resulting from the complexity of AI information that is important for individuals to defend themselves can be withhold which might further even prevent a fair trial.[54,55,56] Additional, AI bears the risk of discrimination in its decision making process as already stated in section 4.1.2. That challenges art. 48, the right of presumption of innocence and right to defence.[57] If the decision is biased due to incomplete data or incorporated grievances, an individual might be falsely suspected.[58,59]

4.2 European General Data Protection Regulation (GDPR)

The digital age was setting the need for a new legal framework safeguarding EU members by ensuring the protection of their data.[60,61] The majority was requiring a standardized data protection right across the EU regardless of the data processing location.[62] Therefore the GDPR was passed in May 2016, came into force on 25th May 2018[63,64] and replaced the Data Protection Directive 95/46/EC.

Compared to the Data Protection Directive the GDPR contains internet related terms.[65] Reason for that was that challenges arising from the application of internet were not present at the time that the Data Protection Directive was introduced.[66] However, these challenges were well present when the GDPR was drafted and therefore were specifically focused.[67] That brings up the issue of not addressing challenges related to AI

[51] Gerards et al. (2020), *Getting the future right*, p. 75.
[52] European Union, *Charter of Fundamental Rights*, art. 47.
[53] Gerards et al. (2020), *Getting the future right*, p. 75.
[54] Gerards et al. (2020), *Getting the future right*, p. 75.
[55] Završnik (2020), "Criminal justice, artificial intelligence systems, and human rights", p. 578.
[56] Leslie et al. (2021), "Artificial intelligence, human rights, democracy, and the rule of law: a primer", p. 15.
[57] European Union, *Charter of Fundamental Rights*, art. 48.
[58] Gerards et al. (2020), *Getting the future right*, p. 75.
[59] Završnik (2020), "Criminal justice, artificial intelligence systems, and human rights", p. 578.
[60] Commission, *Datenschutz in der EU*.
[61] Kunkel et al. (2021), "Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall einschließlich Profiling im Sinne des Art. 22 DSGVO – Praxisrelevanz und Wirksamkeit der Norm in Zeiten von Big Data und KI", p. 9.
[62] Commission, *Datenschutz in der EU*.
[63] Commission, *Datenschutz in der EU*.
[64] Kunkel et al. (2021), "Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall", p. 9.
[65] Parliament et al. (2020), *The impact of the General Data Protection Regulation (GDPR) on artificial*, p. 35.
[66] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 35.
[67] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 35.

as this technology was not relevant enough to be incorporated regardless of its boom in 2012. Therefore, while the GDPR contains terms referring to the internet, it does not contain the term AI itself nor any term of a related concept.[68] Not even the term big data is mentioned within this regulation regardless of its hype in 2001.[69] Nevertheless, the regulation is relevant to AI and some of the provisions are challenged by the its uniqueness.[70] It is not possible to cover all interference of AI and the GDPR but the most critical ones will be discussed below.

4.2.1 Training of AI systems

Training processes of AI systems heavily rely on data as data is the oil to fuel AI. If personal data contributes to applied training data sets, challenges arise regarding the GDPR. In example if an AI system is tasked to simply distinguish between an e-mail address and a telephone number it first needs to be fed and trained with large amounts of data.[71] These data sets would need to include e-mail addresses and telephone numbers in order to extract patterns and how to distinguish between them both.[72] Both attributes are considered as personal data under art. 4(1) of the GDPR and therefore must comply with requirements of the GDPR.[73]

Training Data

Art. 4(2) enshrines that the collection and use of personal data is referred to as processing.[74] Therefore it needs to be ensured that the processing of these data sets is in line with requirements layed out by the GDPR. Art. 6 states the requirement for the identification of a correct legal basis for the processing of personal data.[75] If an AI system would be trained to extract eye colors of individuals, training data sets would include pictures of individuals which falls under the processing of special categories of personal data.[76] Art. 9(1) prohibits the processing of special categorised personal data[77,78] with the exception enshrined in art. 9(2)[79]. However, the legal basis for applied training data not only depends on the classification of personal data it also depends on its origin.[80]

[68] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 35.
[69] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 35.
[70] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 35.
[71] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[72] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[73] European Union, *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, art. 4(1).
[74] European Union, *GDPR*, art. 4(2).
[75] European Union, *GDPR*, art. 6.
[76] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[77] European Union, *GDPR*, art. 9(1).
[78] Gerards et al. (2020), *Getting the future right*, p. 51.
[79] European Union, *GDPR*, art. 9(2).
[80] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.

The simplest legal way of gathering training data is if it initially gets collected as training data with participants participating willingly, as it would fall under art. 6(1)(f).[81,82] Another simple way to cover a legal gathering of data is if it can be based on consent according to art 6(1)(a).[83] Thus, if existing data initially collected for another purpose rather than training AI gets processed new challenges arise.[84] That is referred to as change of purpose and must be measured against the requirements enshrined in art. 6(4) which states that the new purpose must be compatible with its initial purpose.[85] For the use of personal data to train AI, such an assumption would be almost impossible, to justify.[86]

Another problematic case is the processing of data collected from other sources, as it is almost impossible to identify if it initially was collected lawfully as well as identifying its origin purpose.[87] Additional, the large volume of data makes it almost impossible to identify data subjects and therefore they are not known to the new data processor.[88] That rules out the option to obtain consent as art. 6(1)(a) enshrines, and the only considerable legal basis left is the legitimate interest pursuant under art. 6(1)(f).[89] To enable that a balancing of interests must be carried out, which in particular must include the purpose of the planned processing of the data.[90] The question of the feasibility of identifying the person related to the used data must also be taken into account.[91]

Rights of the data subject

If personal data is processed, the data subject has data subject rights that must be guaranteed by the data controller.[92] The implementation of these rights is already challenging, yet, it even gets more complicated with the application of AI.[93] Among others, main obligations of the controller are enshrined in art. 15-18.[94] Art. 15 lays out the obligation to provide information to the data subject about their processed data as well as their stored personal data.[95] Additionally, according to art. 16-18 controllers have the obligation to to rectify, delete and restrict the processing of personal data if inquired by the data subject.[96] Data processing falls in general also under profiling according to

[81] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[82] European Union, *GDPR*, art. 6(1).
[83] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[84] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[85] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[86] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[87] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[88] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[89] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[90] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[91] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[92] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[93] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[94] European Union, *GDPR*, art. 15-18.
[95] European Union, *GDPR*, art. 15.
[96] European Union, *GDPR*, art. 16-18.

art. 4(4) which particularly effects the obligation of information about personal data to a data subject according to art. 15(1)(h).[97]

As already discussed in the section on training data, big data also makes it more difficult to enforce and fulfill the rights of data subjects.[98] Large volumes of processed data makes it almost impossible to trace and identify personal data that is being processed.[99] Therefore, data subject might not be identified and beyond that their data subject rights might not be respected.[100] Further, the black box problem is one of the greatest issue in the field of AI as well as its governance.[101] Due to complexity, autonomy and opacity it is almost impossible to fully understand processes of AI systems and therefore providing a detailed explanation is almost impossible.[102] That might result into an unfulfilled claim for information against a data subject, as the necessary insights into these processes are not accessible.[103] Similar to that, it is impossible to follow the request of any data subject of deleting their personal data due to possible unintentionally stored data.[104]

4.2.2 Automated decision-making and profiling

AI enables the ability of automated decision-processes. In example, it might be used to determine whether or not an individual is entitled for receiving a loan solely on the base of a persons income, expenses and other personal data.[105] Data collecting during these processes can result into the creation of one or more profiles of an individual.[106] Based on these profiles decision-making processes might further act with a result ranging from less harmful to critical.[107] Thus, an individual should not be acted on based on an exclusively data-driven decision from a machine responsible evaluation[108] and therefore art. 22(1) of the GDPR banned automated case-by-case decisions in general, including profiling.[109,110,111] However, that only applies if the individual experiences any legal or significant effect.[112,113] Also important to note it, that not automated data processing itself was banned, only the decision making based on that process.[114] Further, automated

[97] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[98] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[99] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[100] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[101] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[102] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[103] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[104] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[105] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[106] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[107] Hilchenbach et al. (2023), "AI and the GDPR", accessed on 6.5.2024.
[108] Kunkel et al. (2021), "Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall", p. 10.
[109] European Union, *GDPR*, art. 22.
[110] Kunkel et al. (2021), "Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall", p. 9.
[111] Gerards et al. (2020), *Getting the future right*, p. 63.
[112] European Union, *GDPR*, art. 22(1).
[113] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 75.
[114] Kunkel et al. (2021), "Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall", p. 12.

decision making is not prohibited if the outcome is inspected and revised by a human being.[115]

4.2.3 Personal data in re-identification

As previous sections stated, the main objective of the GDPR is the prevention of any kind of fully or partly automated processing of personal data as well as non-automated processing of personal data if data gets stored.[116] Art. 4(1) defines personal data as any information that can be linked to an identified or identifiable natural person.[117] Recital (26) states that the regulation does not apply to anonymous data, regardless if data is already anonymous or if it has been anonymized.[118] However, AI enables the possibility to re-identify data due to their ability of connecting non-identified data to the related individual. The re-identification is equivalent to processing personal data and might be assimilated as collecting new personal data.[119] This is particularly critical when considering critical personal data.[120]

4.3 European Product Liability Directive (PLD)

In 1985 the Product Liability Directive (85/374/EEC) was adopted to ensure a safe environment for consumers across Europe.[121] This introduced strict liability for defective products and the resulting consumer claims for damages.[122] According to this directive, a producer of a product is liable for any damage caused a defect of the product if the injured party is able to prove the damage, the defect as well as their connection.[123] Compensation for personal injury and property damage can be claimed by injured parties up to ten years after a product has been placed on the market.[124] In 19999 the directive was amended by Directive (1999/34/EC) which solely redefined the term product.[125] Since then the EC's liability regime relied on the PLD, however, in 2018 several shortcomings of the directive regarding AI were identified.[126]

4.3.1 Product definition

According to art. 2 a product is defined as all movables even if incorporated into another movable or into an immovable, including electricity.[127]

[115] European Union, *GDPR*, art. 22(1).
[116] European Union, *GDPR*, Art. 2(1).
[117] European Union, *GDPR*, art. 4(1).
[118] European Union, *GDPR*, recital (26).
[119] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 74.
[120] Parliament et al. (2020), *The impact of the GDPR on AI*, p. 53.
[121] European Union, *Product Liability Directive*, pp. 1-2.
[122] Tambiama (2023), *Artificial intelligence liability directive*, p. 2.
[123] Tambiama (2023), *AILD*, p. 2.
[124] Tambiama (2023), *AILD*, p. 2.
[125] European Union, *Amendment of the Product Liability Directive*, art. 1.
[126] Ziosi et al. (2023), "The EU AI Liability Directive (AILD): Bridging Information Gaps", p. 2.
[127] European Union, *Amendment PLD*, art. 2.

Many aspects of AI are challenging the definition of the term product layed out in this directive. First, as AI systems, products and services are closely interacting, it is almost impossible to make a clear distinction between these components.[128,129] Further, due to the ambiguity of whether a software is a product or is rather classified as only a part of a product it is questionable if a software is legally covered by the concept of product.[130] Similar to that, it is also unclear whether involved data and updates of an AI system are included in the concept of product.[131]

4.3.2 Producer definition

> According to art. 3(1) a producer means the manufacturer of a finished product, the producer of any raw material or the manufacturer of a component part and any person who, by putting his name, trade mark or other distinguishing feature on the product presents himself as its producer.[132]

Origins of the PLD were risk occurring due to mass-production.[133] Within the area of mass-production it was feasible to carry out precise and predictive risk analysis and therefore it was quite reasonable to hold the manufacturer accountable for any damage of its product.[134] Considering the current rise of technology strongly relying on AI and its included unpredictability, the impossibility of a precise and predictive risk analysis must be taken into account. Further, the life cycle of AI might involve many parties and not only a single manufacturer that could be held accountable.[135] Developers, operators and other involved parties may be the cause of the emerged harm.[136] Therefore it is questionable if the concept of produced as defined in the PLD may be outdated due to a needed responsibility shift for components as software AI systems and data systems.[137,138]

4.3.3 Defect definition

> According to art. 6(1) a product is deemed to be defective if it does not provide for the safety that a person can reasonably expect at the time it was put into circulation.[139]

[128] (2020), "Producer Liability for AI-Based Technologies in the European Union", p. 78.
[129] (2023), "The revision of the product liability directive: a key piece in the artificial intelligence liability puzzle", p. 253.
[130] Ziosi et al. (2023), "The EU AILD: Bridging Information Gaps", p. 2.
[131] (2020), "Liability for AI-Based Technologies", p. 78.
[132] European Union, *PLD*, art. 3(1).
[133] Li et al. (2022), "Liability Rules for AI-Related Harm: Law and Economics Lessons for a European Approach", p. 624.
[134] Li et al. (2022), "Liability Rules for AI-Related Harm", p. 618.
[135] (2020), "Liability for AI-Based Technologies", p. 81.
[136] Li et al. (2022), "Liability Rules for AI-Related Harm", p. 2.
[137] (2018), *Evaluation of Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products*, p. 54.
[138] Li et al. (2022), "Liability Rules for AI-Related Harm", p. 625.
[139] European Union, *PLD*, art. 6.

In order for a consumer to hold the producer liable the defect of a product must be proven.[140] Taking into account that any AI system is based on software as well as the fact that error-free software rarely exists it is difficult to apply the concept of defect to software.[141] Further, if an AI system is already in operation it has the ability to learn and therefore might result into unpredictable outcomes.[142] It is unclear whether unpredictable outcomes causing any damage could be seen in the scope of defect.[143] Finally, if the product was already put into circulation and an update is being performed, it is questionable who should be held liable.[144,145]

4.3.4 Burden of proof

> According to art. 4 the injured person shall be required to prove the damage, the defect and the causal relationship between defect and damage.[146]

Characteristics of AI as its opacity, complexity and autonomy create the potential of impeding the process of injured parties identifying the responsible producer as well as identifying and proving any fault.[147] Injured parties might not have the needed technological knowledge to interpret the decision-process of an AI system and therefore might not be able to identify and prove the defect nor the causal link between that defect and the damage suffered.[148] That missing capability of identifying the defect makes it almost impossible to single out a liable person.[149] Identifying any potentially liable person is further difficult due to the involvement of many parties throughout the life cycle of AI.

[140] European Union, *PLD*, art. 4.
[141] Ziosi et al. (2023), "The EU AILD: Bridging Information Gaps", p.
[142] (2023), "The revision of PLD", p. 257.
[143] (2020), "Liability for AI-Based Technologies", p. 78.
[144] (2020), "Liability for AI-Based Technologies", p. 79.
[145] (2023), "The revision of PLD", p. 257.
[146] European Union, *PLD*, art. 4.
[147] (2023), "The revision of PLD", p. 256.
[148] (2023), "The revision of PLD", p. 256.
[149] (2023), "The revision of PLD", p. 256.

CHAPTER 5

European Union's journey towards trustworthy AI

Artificial Intelligence (AI) is already part of our daily lives and its usage and impact will continue to grow.[1,2] However, as discussed in *The future goal of achieving Trustworthy AI* and *Applicable legal acts and concerns posed by AI* concerns about its ethical and legal aspects are increasing as well[3] and are negatively affecting the trust of society. This resulting lack of trust goes hand in hand with a lack of acceptance of AI systems.[4] Therefore, in order to benefit from the advancing development of AI and enable its further application, trust of society is required.[5] For years, the European Union is already tackling the mission of enhancing the application of AI across the European Union.[6]

Their journey towards a regulatory framework regarding AI started around 2017. At the beginning the topic AI was a bit neglected and therefore not handled in an appropriate extent. Despite many recommendations that were given to establish a legal framework regarding AI the focus was primarily on combating and dealing with its ethical implications. Finally, on the 21st April 2021 the European Commission proposed a legal regulatory framework to ensure trustworthy AI with a special interest in respecting Union rights and values while benefiting of the advantages that AI has to offer. That might put the European Union in the position of a global hub for trustworthy AI. However, establishing or amending the legal system is a time consuming process, since it requires thorough analysis, consultation and legal drafting and the complexity of AI further decelerates this process.

[1] FLI, *The AI Act*, accessed on 28.04.2023.
[2] Janssen et al. (2022), "Practical fundamental rights impact assessments", p. 201.
[3] Robles (2020), "Artificial intelligence: From ethics to law", p. 1.
[4] Wartner, *Vertrauen in die Künstliche Intelligenz*.
[5] Wartner, *Vertrauen in die KI*.
[6] Commission (2023), *Commission welcomes political agreement on Artificial Intelligence Act*, p. 2.

5.1 The start of AI centered governance

The report of 'Recommendations to the Commission on Civil Law Rules on Robotics', published on 27th January 2017, was one of the earliest milestones from the European Parliament on AI governance.[7] Although the main topic was robotics, it also addressed AI, as both were one of the most important technological trends of the century.[8,9] One of the origins of these recommendations was the discussion about Asimov's laws, listed in *Before the term was coined*, only being directed at the people creating and interacting with robots including robots assisted by AI practices since those laws could not be converted into machine code.[10] It laid out the importance of complementing and adapting the European legal framework by establishing ethical principles respected in the development, programming and use of AI.[11,12] These principles should not replace the need of a legal regulatory regarding AI but merely complement it.[13] Additionally it stated the possible necessity of creating a generally accepted flexible definition of AI to not hinder its innovation[14,15] as well as the importance about liability issues arising from AI.[16] However the main focus still relied on robotics and AI was not seen as an independent field in governance.[17]

Shortly afterwards, on the 31st August 2017 the European Economic and Social Committee (EESC) presented their 'Opinion on AI'.[18] Within this opinion they recommended the EU to take the lead in establishing clear global policy frameworks regarding AI that are based on EU values, especially on the European Charter of Fundamental Rights (CFR).[19] In line with the report of the parliament, it pointed out the need of establishing codes of ethics for the development, application and use of AI and additionally advocated an human-in-command approach to ensure human control at all time.[20] They stated that AI systems need to be verified, validated and monitored not only from a technical but also from an ethical, safety and societal perspective.[21] For this purpose a standardisation systems for AI should be developed that is based on values from important areas including

[7] Stix (2022), "The Ghost of AI Governance Past, Present, and Future: AI Governance in the European Union", p. 2.
[8] Parliament, *European Parliament resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*, p. 2.
[9] Stix (2022), "The Ghost of AI Governance", p. 2.
[10] Parliament, *Recommendations to the Commission on Civil Law Rules on Robotics*, p. 3.
[11] Parliament, *Recommendations to the Commission on Civil Law Rules on Robotics*, p. 6.
[12] Stix (2022), "The Ghost of AI Governance", p. 2.
[13] Parliament, *Recommendations to the Commission on Civil Law Rules on Robotics*, p. 15.
[14] Parliament, *Recommendations to the Commission on Civil Law Rules on Robotics*, p. 1.
[15] Stix (2022), "The Ghost of AI Governance", p. 2.
[16] Parliament, *Recommendations to the Commission on Civil Law Rules on Robotics*, pp. 4-5.
[17] Parliament, *Recommendations to the Commission on Civil Law Rules on Robotics*.
[18] EESC, *Opinion of the European Economic and Social Committee on 'Artificial intelligence — The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society'*, p. 1.
[19] EESC, *Opinion on AI*, p. 2.
[20] EESC, *Opinion on AI*, p. 2.
[21] EESC, *Opinion on AI*, p. 1.

safety, transparency, comprehensibility, accountability and ethical standards.[22] To enable the EU to take the lead in AI governance they called out the need to establish an EU AI infrastructure and a detailed analysis of the in forced laws and regulations.[23] Finally, they stated their support of the ban on lethal autonomous weapons systems.[24] That was the first event of AI governance being handled as an independent topic from robotics.

On 29th September 2017 the 'Tallinn Digital Summit' took place[25], organized by the Estonian presidency of the Council of Europe in cooperation with the president of the European Council and the European Commission.[26] It served as a platform for discussing digital innovation plans to enable Europe to maintain its technological edge and take a leading role in the digital sector in the upcoming years.[27] The resulted outcome strongly presented the need for a stronger and more coherent Digital Europe.[28] Less than one month later on 19th October 2017 the leaders of the European Council discussed that approach among other topics.[29] Eight points to address got singled out where each of them was and still is relevant for the governance of AI: cybersecurity, a first-rate infrastructure and communications network(5G), a future-oriented regulatory framework, digitalization in the public sector and government, combating online crime, digital skills of the citizens, R&D investment efforts and addressing technological trends.[30] Although AI was not a main topic it was included in the last point due to its increasing hype.[31] Further, the European Council invited the Commission to put forward an European approach to AI by 2018.[32,33] The approach was introduced on 25th April 2018 which will be discussed in more detail later.[34]

The 'Joint Declaration' was published on 14th December 2017, addressing the upcoming year 2018-2019.[35] It gets published annually and includes the legislative priorities of the European Union for the upcoming year.[36] The Council, the Parliament and the Commission are the main actors to discuss and agree on these goals. In despite of the other documents discussed above the 'Joint Declaration' did not include AI nor any related term.[37] However, in addition it stated that it is important to follow up on ensuring

[22] EESC, *Opinion on AI*, p. 5.
[23] EESC, *Opinion on AI*, pp. 2, 5.
[24] EESC, *Opinion on AI*, p. 3.
[25] Council, *European Council meeting (19 October 2017) – Conclusions*, p. 5.
[26] European Parliament, *"Digitales Gipfeltreffen Tallinn", 29.09.2017, 29 September 2017*, accessed on 7.5.2024.
[27] European Parliament, *Digitales Gipfeltreffen Tallinn*, accessed on 7.5.2024.
[28] Council, *European Council meeting*, p. 5.
[29] Council, *European Council meeting*, pp. 1-10.
[30] Council, *European Council meeting*, pp. 6-8.
[31] Council, *European Council meeting*, p. 7.
[32] Council, *European Council meeting*, p. 7.
[33] European Commission, *Declaration on the Cooperation on Artificial Intelligence*, p. 3.
[34] European Commission, *Declaration - Cooperation on AI*, p. 3.
[35] Commission, *Joint Declaration on the EU's legislative priorities for 2018-19*, p. 1.
[36] Commission, *Joint Declaration*, p. 1.
[37] Commission, *Joint Declaration*, pp. 1-2.

a high level of data protection, digital rights and ethical standards for AI with a special interest in capturing its benefits and minimizing its risks.[38]

Another milestone in capturing Artificial Intelligence (AI) governance as an independent field from robotics was the Statement on 'Artificial Intelligence, Robotics and Autonomous Systems' by the European Group on Ethics in Science and New Technologies (EGE) published on 9th March 2018.[39] The EGE is an independent advisory body to the European Commission which is responsible for advising them on ethical, social and fundamental rights issues arising from science and new technologies.[40] They pointed out the need of establishing a framework directly regarded to AI in the EU with a strong emphasis on ethical aspects.[41] Ethical, legal as well as societal governance issues should be tackled while ensuring that AI is created with a human-centered approach.[42] Therefore, as the 'civil law rules on robotics' and the 'Opinion on AI' already proposed they support the idea of the development of ethical codes of conduct for AI with a special interest in protecting fundamental European values.[43] This idea was typically European due to its emphasizement on the importance of European values and additionally it came at the right time since ethical principles of AI were beginning to gain on importance in the broader international landscape.[44] It also aligns with the invitation of the Council to the Commission to put forward an European approach to AI.

5.2 EU members cooperating on AI governance

Afterwards the history of AI governance experienced major leaps. The first international document addressing AI as an independent encapsulated topic, the 'Digital Day Declaration on Cooperation on AI', was published by the European Commission on the 10th April 2018.[45,46] Twenty-three member states of the European Union as well as UK and Norway signed up on cooperating on AI.[47,48] In the same year later on Romania, Greece, Cyprus and Croatia joined as well.[49] All participants agreed to three key visions within this cooperation. First, they agreed on boosting the uptake of AI as well as its technological and industrial capacity.[50] That includes an essential condition in the development of AI, better access to public sector data.[51] Second, they agreed on addressing socio-economic challenges as the upcoming transformation of the labour

[38] Commission, *Joint Declaration*, p. 2.
[39] Stix (2022), "The Ghost of AI Governance", p.3.
[40] Stix (2022), "The Ghost of AI Governance", p.3.
[41] Stix (2022), "The Ghost of AI Governance", p.3.
[42] Stix (2022), "The Ghost of AI Governance", p.3.
[43] Stix (2022), "The Ghost of AI Governance", p.3.
[44] Stix (2022), "The Ghost of AI Governance", p.3.
[45] European Commission, *Declaration - Cooperation on AI*, p. 8.
[46] Stix (2022), "The Ghost of AI Governance", p. 4.
[47] European Commission, *Declaration - Cooperation on AI*, p. 8.
[48] European Commission, *EU Member States sign up to cooperate on Artificial Intelligence.*
[49] European Commission, *EU Member States sign up.*
[50] European Commission, *Declaration - Cooperation on AI*, p. 3.
[51] European Commission, *Declaration - Cooperation on AI*, p. 3.

market and the need of a modernised education system.[52] Third, they agreed on working towards a legal and ethical framework build on the fundamental rights and values of the European Union with a special interest on privacy, data protection, transparency and accountability.[53] To a certain extent the 'Cooperation' can be seen as the forerunner of the AI Act as it laid the foundation.[54] In June 2018 it was endorsed by the European Council.[55]

Not long after that, another further leap was reached on 25th April 2018 as the Commission published the strategy on AI[56] entitled 'Artificial Intelligence for Europe'.[57,58] It was the response to the request of the Council on putting forward an European approach to AI by 2018.[59,60] Back then the Council highlighted the urgency of addressing emerging trends as AI while ensuring a high-level of data protection, digital rights and ethical standards. As discussed above the Parliament, the EESC and the 'Joint Declaration' as well already recommended on following that path. The 'Strategy' then stated that the power of AI should be at the service of human progress while no one is left behind.[61] Based on that it stated the necessity to develop a strategy that ensures the competitiveness of the EU in the global landscape of AI.[62] Therefore the EU positioned itself as an international actor of AI governance that puts ethical considerations and fundamental rights at the core of AI governance.[63]

Therefore, the 'Cooperation' and the 'Strategy' encompassed the same three main elements: boost Europe's technological and industrial capacity, prepare Europe for socio-econimic changes accompanied by AI and ensure that Europe has an appropriate ethical and legal framework addressing the whole application cycle of AI.[64] To address these key visions the Commission will set up a 'Coordinated plan on AI' in cooperation with all member states by the end of 2018.[65] According to plan the Commission presented the 'Coordinated Plan on AI' on 7th December 2018.[66,67] It picked up where the 'Declaration'

[52] European Commission, *Declaration - Cooperation on AI*, p. 3.
[53] European Commission, *Declaration - Cooperation on AI*, p. 3.
[54] Stix (2022), "The Ghost of AI Governance", p. 4.
[55] Commission, *Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Comittee and the Commitee of the Regions Coordinated Plan on Artificial Intelligence*, p. 2.
[56] Commission, *Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Comittee and the Commitee of the Regions Artificial Intelligence for Europe*, pp. 5-16.
[57] Commission, *Communication Artificial Intelligence for Europe*.
[58] Commission (2023), *Commission welcomes political agreement*, p. 2.
[59] Commission, *Communication Artificial Intelligence for Europe*.
[60] Commission (2023), *Commission welcomes political agreement*, p. 2.
[61] Commission, *Communication Artificial Intelligence for Europe*, p. 19.
[62] Commission, *Communication Artificial Intelligence for Europe*, p. 1.
[63] Stix (2022), "The Ghost of AI Governance", p. 6.
[64] Commission, *Communication Artificial Intelligence for Europe*, pp. 5-16.
[65] Commission, *Communication Artificial Intelligence for Europe*, pp. 3, 18.
[66] Commission (2023), *Commission welcomes political agreement*, p. 2.
[67] Commission, *Communication Coordinated Plan on Artificial Intelligence*.

left off and especially highlighted the dependence of AI on the GDPR.[68]

5.3 AI Alliance and High-Level Expert Group on Artificial Intelligence

The 'Strategy' particularly emphasized the importance of all member states working together.[69] In order to achieve that, the Commission acknowledged the need for an exchange of knowledge between all relevant stakeholders in the field of AI.[70] Therefore the second aim of the 'Strategy' was to set up an AI Alliance by July 2018[71] which initially was planned as a framework for stakeholders and experts to develop ethical guidelines for AI.[72] However, in June 2018 the Commission initiated two separate groups which are closely working together: the AI Alliance and the High-Level Expert Group on Artificial Intelligence.[73,74] Up until now the AI Alliance serves a multi-stakeholder platform to provide input from all parts of society.[75] It enables an open discussion about all relevant aspects in the life cycle of AI as well as its social end economical impact.[76] In contrast, the AI HLEG only consists of experts which are supporting the implementation of the 'Strategy' by providing recommendations on future-oriented policy development and on ethical, legal and social issues, including socio-economic challenges.[77] That group steers the work of the AI Alliance and reflects their gathered views in their own analysis and reports.[78]

5.3.1 Important Deliveries of the AI HLEG

Ethics guidelines on trustworthy AI

In the scope of the 'Strategy' the Commission requested the High-Level Expert Group on Artificial Intelligence to establish ethical guidelines for AI.[79,80] As a response to that the AI HLEG presented the first draft of 'Ethics guidelines for trustworthy AI' on the 18th December 2018.[81] These guidelines still had to go through a piloting process starting with 26th June 2019 which was realized through an open consultation with the outcome of over 500 comments that could be collected. Finally on 8th April 2019 the final version

[68] Commission, *Communication Coordinated Plan on Artificial Intelligence*, p. 6.
[69] Commission, *Communication Artificial Intelligence for Europe*, p. 17.
[70] Commission, *Communication Artificial Intelligence for Europe*, p. 18.
[71] Commission, *Communication Artificial Intelligence for Europe*, p. 18.
[72] Commission, *Communication Artificial Intelligence for Europe*, p. 16.
[73] European Commission, *Die Europäische KI-Allianz*, accessed on 7.5.2024.
[74] European Commission, *Policy and investment recommendations for trustworthy Artificial Intelligence*, accessed on 7.5.2024.
[75] European Commission, *Die Europäische KI-Allianz*, accessed on 7.5.2024.
[76] European Commission, *Die Europäische KI-Allianz*, accessed on 7.5.2024.
[77] European Commission, *High-level expert group on artificial intelligence*, accessed in 7.5.2024.
[78] European Commission, *HLEG AI*, accessed in 7.5.2024.
[79] Commission, *Communication Coordinated Plan on Artificial Intelligence*, p. 8.
[80] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 4.
[81] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 4.

of the 'Ethics guidelines for trustworthy AI' was presented.[82] As discussed in *Ethics guidelines for trustworthy AI* within these guidelines, key requirements of trustworthy AI were singled out as well as technical and non-technical methods to realize trustworthy AI. Further, a first draft of the later on published 'Assessment List of Trustworthy AI' was included as well as a definition of AI to enable a common understanding for further deliverables.[83]

Policy and investment recommendations for trustworthy AI

Within this recommendations the focus is layed on humans and society; private sector; public sector; and research and academia to determine what policies are needed for trustworthy AI.[84] Further four key policy areas got singled out that will act as enablers for trustworthy AI: data and infrastructure; education and skills; governance and regulation; and funding and investment.[85] It was published on the 26th June 2019 without any preceding public discussions.[86]

Assessment List for Trustworthy Artificial Intelligence (ALTAI)

The Assessment List for Trustworthy AI (ALTAI) is a self-assessment tool that translated ethics guidelines into a checklist.[87] Developers and deployers of AI are advised to use this tool if they want to practically implement the key requirements.[88] It is available in PDF format as well as a web based tool.[89] The final version was published on the 17th July 2020 implemented next to the key requirements of the 'Ethics guidelines on trustworthy AI' results from public discussions that were conducted by the Commission between June and December 2019.[90]

Sectoral Considerations on the Policy and Investment Recommendations -

The base of this document was the previous deliverable 'Policy and Investment Recommendations for trustworthy AI'.[91] A series of workshops were held were these recommendations were systematically discussed and reviewed.[92] Representatives and stakeholders of three sectors that are considered essential for the development and deployment of AI were invited: health, the public and manufacturing/IoT.[93] The 'Sectoral Considerations on the

[82] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. ii.
[83] AI HLEG (2019), *Ethics guidelines for trustworthy AI*, p. 36.
[84] European Commission, *Policy and investment recommendations*, accessed on 7.5.2024.
[85] European Commission, *Policy and investment recommendations*, accessed on 7.5.2024.
[86] European Commission, *Policy and investment recommendations*, accessed on 7.5.2024.
[87] European Commission, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, accessed on 7.5.2024.
[88] European Commission, *ALTAI*, accessed on 7.5.2024.
[89] European Commission, *ALTAI*, accessed on 7.5.2024.
[90] European Commission, *ALTAI*, accessed on 7.5.2024.
[91] HLEG AI (2020), "Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence", p. 4.
[92] HLEG AI (2020), "Sectoral Considerations", p. 4.
[93] HLEG AI (2020), "Sectoral Considerations", pp. 3-4.

Policy and Investment Recommendations for trustworthy AI' compromises the outcome of these workshops and was published on the 23rd July 2020.[94]

5.4 Ethical Charter on the use of AI in judicial systems and their environment

On the 3rd December 2018 the European Commission for Efficiency of Justice (CEPEJ) introduced the 'EU Charter on the use of AI in judicial systems and their environment' to prevent violation of citizen's rights and freedoms by any intelligent tool introduced into any judicial system.[95] The CEPEJ is a judicial body of the European Council consisting of experts from all the 46 member states.[96] Their aim is to improve the efficiency and functioning of justice in the member States and to develop the implementation of the instruments adopted by the Council.[97] In the new introduced Charter they agreed on five fundamental principles for introducing intelligent tools into any judicial system:

1. Principle of respect for fundamental rights: It must be ensured that the design and implementation of any AI tool as well as services are complying with fundamental rights.[98]

2. Principle of non-discrimination: It must be ensured that the development or intensification of any discrimination between individuals or groups of individuals are prevented.[99]

3. Principle of quality and security: Quality and security must be ensured in the processing of judicial decisions and data. Therefore, certified sources and intangible data with models conceived in a multi-disciplinary manner must be used and a secure technological environment must be provided.[100]

4. Principle of transparency, impartiality and fairness: Accessible and understandable data processing methods must be ensured as well as the authorisation of external audits.[101]

5. Principle 'under user control'. It must be ensured that users are informed about the use of AI and that they have control over the resulting decisions.[102]

[94] HLEG AI (2020), "Sectoral Considerations", p. 4.
[95] CEPEJ (2018), "European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment", p. 5.
[96] CoE, *About the European Commission for the efficiency of justice (CEPEJ)*, accessed on 5.05.2024.
[97] CoE, *About the (CEPEJ)*, accessed on 5.05.2024.
[98] CEPEJ (2018), "Charter on AI in judicial systems", p. 7.
[99] CEPEJ (2018), "Charter on AI in judicial systems", p. 7.
[100] CEPEJ (2018), "Charter on AI in judicial systems", p. 7.
[101] CEPEJ (2018), "Charter on AI in judicial systems", p. 7.
[102] CEPEJ (2018), "Charter on AI in judicial systems", p. 7.

5.5 Report on liability for AI and other emerging technologies

In March 2018 the Commission set up the Expert Group on Liability and New Technologies, ordered to operate in two formations: the Product Liability Directive (PLF) formation and the New Technologies formation (NTF).[103] While the PLF was tasked to assess the product liability directive, the NTF was commissioned to assess whether the current liability regime was still suited to facilitate the uptake of new technologies, including AI.[104] As part of this task, theNTF were asked to make recommendations for amendments in the event of any found shortcoming, without limiting themselves to the existing national and EU legal instruments.[105] Ten meetings of the NTF were held from June 2018 to May 2019 to discuss that ordered task.[106] On the 27th November 2019 the expert group published the report 'Report on liability for AI and other emerging technologies' which layed out their gathered findings and recommendations.[107]

The report stated that the current legal framework at least provides a starting point for assessing liability by providing basic protection for victims suffering from harm caused by any emerging digital technology.[108] However, this framework is not well suited to the dynamic, complex and rapidly evolving field of AI due to the specific characteristics of the technology such as as complexity, opacity, openness, autonomy, predictability, data-driven and vulnerability.[109] These characteristics make it rather difficult to offer victims a claim for compensation in all justified cases as already the allocation of liability is unfair or inefficient in many cases.[110] Therefore the NTF made the following recommendations:

- **Legal personality:** The report argue that damages resulting from the new technologies could still be attributed to existing legal entities or categories and therefore opposes the idea of granting autonomous systems legal personality as considered in the 'Civil Law Rules on Robotics resolution'.[111]
- **Operator's liability:** The report suggested that operators in a non-private environment using any emerging technology that bears the potential of significant harm should continue to bear liability.[112] The experts argued that these operators bear the risks of operating such systems and that such a liability should be strict.[113] Furthermore, the report stated the necessity to replace the traditional concepts of

[103] NFT (2019), "Liability for artificial intelligence and other emerging digital technologies", p. 12.
[104] NFT (2019), "Liability for AI", p. 13.
[105] NFT (2019), "Liability for AI", p. 13.
[106] NFT (2019), "Liability for AI", p. 13.
[107] NFT (2019), "Liability for AI", p. 13.
[108] NFT (2019), "Liability for AI", p. 3.
[109] NFT (2019), "Liability for AI", pp. 3, 5.
[110] NFT (2019), "Liability for AI", p. 3.
[111] NFT (2019), "Liability for AI", p. 37.
[112] NFT (2019), "Liability for AI", p. 39.
[113] NFT (2019), "Liability for AI", p. 39.

the terms owner, user and keeper with a more flexible and broaden concept of the term operator.[114]

- **Producer's liability:** The report suggested that the producer should be held liable for a defect in any product or any incorporated digital content into an emerging digital technology even if it appeared after the product was already placed on the market.[115] Further, it suggested that if it was foreseeable that unforeseen developments might occur the development risk protection should not apply to producers anymore.[116]
- **Fault liability and the duties of care:** The report addressed the importance of both operators and producers of any new technology complying with a suited range of obligations.[117] Further, it considered producers that are incidentally also act as the operator to signal the expansion of the producer's responsibilities, whose functions can also partly merge into those of an operator.[118]
- **Vicarious liability:** The report stated the possibility of expanding vicarious liability regimes to harms caused by autonomous technologies.[119] They demonstrated that with a comparison of a conventional vehicle, where the operator explicitly has the control over the vehicle and therefore in control of potentially arising risks, to an autonomous vehicle, including a autopilot mode. If any harm is caused by the operator using the autopilot, the operator is acting on behalf of the producer and therefore the producer is vicariously liable for the caused damage.[120]
- **Logging by design:** The report considered the implementation of logging systems into AI technologies to simplify the process of identifying the source that caused the damage.[121] If such a logging system is not implemented it could further trigger a rebuttable presumption that the condition of proving the liability is therefore fulfilled by the lack of information.[122]
- **Burden of proof:** Due to the characteristics of AI as well as the interconnection that comes with new technologies, it is often extremely difficult to determine the cause of an error. Still, the report advises that the general burden of proof for the causation of the damage should remain on the shoulders of the victim.[123] However, if it is unreasonably difficult for the victim to prove important elements, the burden of proving causation should be alleviated.[124] Further, the report suggested that

[114] NFT (2019), "Liability for AI", p. 41.
[115] NFT (2019), "Liability for AI", p. 42.
[116] NFT (2019), "Liability for AI", p. 43.
[117] NFT (2019), "Liability for AI", p. 44.
[118] NFT (2019), "Liability for AI", p. 44.
[119] NFT (2019), "Liability for AI", p. 45.
[120] NFT (2019), "Liability for AI", p. 35.
[121] NFT (2019), "Liability for AI", p. 47.
[122] NFT (2019), "Liability for AI", p. 47.
[123] NFT (2019), "Liability for AI", p. 49.
[124] NFT (2019), "Liability for AI", p. 49.

in cases where it is proven that an emerging technology caused harm but it is almost impossible to prove the fault, the burden of proof should be reversed.[125] However, these recommendations might imply an amendment of the PLD where the evaluation and recommendations are lying on the shoulder of the PLF.

- **Insurance and compensation funds:** The report further emphasizes the idea of introducing a mandatory liability insurance for certain emerging technologies to simplify the way towards a successfully claim for compensation for victims.[126] Finally, the report stated the possibility to establish compensation funds for victims having a hard time to claim compensation due to difficulties that arise through the characteristics of AI or in the case of uninsured technologies.[127]

5.6 From the White Paper on AI to the proposal of the AI Act

Since the beginning of 2020 some European legislation on AI had been expected since Ursula von der Leyen pledged that within hundred days of taking on the role as the president of the European Commission (EC) she is going to propose some legislation on AI.[128,129] She took office on 1st December 2019 after her election on the 16th July 2019.[130] At the same time of the elections the High-Level Expert Group on Artificial Intelligence (AI HLEG) were establishing ethical guidelines as well as policy and investment recommendations for trustworthy AI.[131] That is why a member of this group stated that Leyen's strategy regarding of being a reasonable one was being unrealistic.[132] In his opinion the next steps would clearly be the translation of those newly created guidelines and recommendations into a legal framework.[133] Additionally the work that had been done by the AI HLEG showed that the road towards trustworthy AI is going to be tedious and therefore he concluded that it would take at least a year instead of three months to establish a regulatory framework.[134] That as well turned out to be an optimistic approach.[135]

On the 19th February 2020 the foundation for a European AI regulation was set.[136,137]

[125] NFT (2019), "Liability for AI", p. 52.
[126] NFT (2019), "Liability for AI", p. 62.
[127] NFT (2019), "Liability for AI", p. 62.
[128] Floridi (2021), "The European Legislation on AI: A Brief Analysis of its Philosophical Approach", p. 215.
[129] (2021), "The draft AI Act: a success story of strengthening Parliament's right of legislative initiative?", p. 620.
[130] Parliament (2019), *A Union that strives for more: the first 100 days.*
[131] Floridi (2021), "The European Legislation on AI", p. 215.
[132] Floridi (2021), "The European Legislation on AI", p. 1.
[133] Floridi (2021), "The European Legislation on AI", p. 215.
[134] Floridi (2021), "The European Legislation on AI", p. 215.
[135] Floridi (2021), "The European Legislation on AI", p. 215.
[136] Commission (2023), *Commission welcomes political agreement*, p. 2.
[137] Floridi (2021), "The European Legislation on AI", p. 215.

The European Commission published the 'White Paper on AI - A European Approach to Excellence and Trust' which emphasized the need for trustworthy, safe and ethical AI.[138,139,140] It set the clear vision for AI in the European Union to build an ecosystem of excellence and trust.[141] An ecosystem based on excellence must be created to promote the uptake of AI which needs to be supplemented by an ecosystem of trust to recognize and mitigate potential dangers of AI for society.[142]

To achieve that goal the Commission highlighted the issue of defining the scope of a future regulation in virtue of the absence of a globally accepted definition of AI itself.[143] Therefore adapting a definition is of great importance.[144] However it needs to be flexible enough to keep up with the technical progress while being precise enough to provide the necessary legal certainty.[145] Additionally, they highlighted the future goal of aligning policies across Europe as some countries already have taken the step of proposing intern regulations.[146] In the absence of an EU-wide approach, there is a significant risk of fragmentation in the internal market, which could impede the market uptake.[147] Further it was accompanied by a "report on the safety and liability implications of ai the iot and robotics" that highlighted existing gaps in the current product safety and liability legislation.[148] The liability part strongly referred to issues found in the report of the NTF.

The 'White Paper' was published and the tragic and disruptive COVID-19 pandemic begun to spread.[149] Nonetheless, on the 21st April 2021 the Commission managed to publish their proposal of a regulation regarding AI in accordance to the 'White Paper', the EU Artificial Intelligence Act (Proposal for a regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts).[150,151,152,153] It was supported by

[138] Commission (2023), *Commission welcomes political agreement*, p. 2.
[139] Natasa et al. (2023), "Artificial Intelligence: Friend or Foe? Experts' Concerns on European AI Act", p. 6.
[140] Floridi (2021), "The European Legislation on AI", p. 215.
[141] European Commission, *WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust*, p. 3.
[142] European Commission, *White Paper*, pp. 5-25.
[143] European Commission, *White Paper*, p. 16.
[144] European Commission, *White Paper*, p. 16.
[145] European Commission, *White Paper*, p. 16.
[146] European Commission, *White Paper*, p. 10.
[147] European Commission, *White Paper*, p. 10.
[148] Commission (2023), *Commission welcomes political agreement*, p. 2.
[149] Floridi (2021), "The European Legislation on AI", p. 215.
[150] Floridi (2021), "The European Legislation on AI", p. 1.
[151] Golpayegani et al. (2023), "Comparison and Analysis of 3 Key AI Documents: EU's Proposed AI Act, Assessment List for Trustworthy AI (ALTAI), and ISO/IEC 42001 AI Management System", p. 48.
[152] Ruschemeier (2023), "AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal", p. 363.
[153] 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.*

a revision of the 'Coordinated Plan on AI' that included required reforms to further enhance the leading position of the European Union for the development of reliable AI and stated the necessity to prioritize human well-being, reliability and fairness while respecting fundamental European values.[154,155]

5.7 Steps taken after the proposal of the AI Act

As the Commission demonstrated in the 'Report on Artificial Intelligence Liability' attached to the 'White Paper', AI was exposing specific challenges to existing liability rules. It pointed out the urgency to address these challenges with the aim of ensuring the same level of protection for any application of AI as it is currently provided for traditional technologies.[156] In response to the 'White Paper' the Parliament adopted a proposal on the 6th October 2020 in which the Commission was invited to submit a regulation regarding liability for the application of AI.[157,158,159] Almost two years later the commission responded to this request on the 28th September 2022 with the proposal for an European Artificial Intelligence Liability Directive (AILD).[160] Within this proposal the commission also included a revision of the current in force PLD as it is of great importance to ensure its compatibility with the proposed AILD as well as its alignment to the current digital age and AI.[161,162]

On the 6th December 2022 the Council formally adopted its common position on the AI Act, now waiting for the Parliament to adopt its position to further start trilogue negotiations.[163,164] As the Parliament adopted its negotiating position on the 14th June 2023 with 499 votes in favour, 28 against and 93 abstentions, the trilogue negotiations started.[165,166] Negotiation meetings took place in June, July, September, October and December 2023.[167] Finally, on the 9th December 2023 the Commission welcomed the reached

[154] Adamakis (2023), "A Comparative Analysis of the EU and US Artificial Intelligence (AI) Regulation Regimes", p. 16.
[155] European Commission, *Coordinated Plan on Artificial Intelligence*, accessed on 7.5.2024.
[156] European Commission, *White Paper*, p. 15.
[157] 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)*, p. 1.
[158] (2020), "Civil Liability Applicable to Artificial Intelligence: A Preliminary Critique of the European Parliament Resolution of 2020", p. 1.
[159] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 15.
[160] European Commission, *Proposal AILD*, p. 1.
[161] European Commission, *Proposal AILD*, p. 11.
[162] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 17.
[163] Legislative Train Schedule, *Artificial intelligence act In "A Europe Fit for the Digital Age"*, accessed on 6.5.2024.
[164] Pingen, *Council's Common Position on Artificial Intelligence Act*, accessed on 6.5.2024.
[165] FLI, *Timeline of Developments*, accessed on 6.5.2024.
[166] Legislative Train Schedule, *AI Act*, accessed on 6.5.2024.
[167] Legislative Train Schedule, *AI Act*, accessed on 6.5.2024.

political agreement between the Parliament and the Council on the AI Act.[168].[169,170] The final approval of the AI Act by the Parliament took place on 13th March 2023 with 523 votes in favour, 46 against and 49 abstentions.[171]

Further upcoming steps are the formal adaption of the AI Act by the Parliament which is currently taking place and the final formally endorsement of the Council.[172] Afterwards, the AI Act will be published in the Official Journal.[173] Measured from the day of publication, twenty days later the AI Act will officially enter into force and two years after its entry into force it will be fully applicable with three exceptions:[174]

- Prohibitions will already take effect six months after the AI Act entered into force.[175]
- Governance rules as well the obligations for GPAI models are already applicable twelve months after the AI Act entered into force .[176] art 101 not fines gpai
- Rules for embedded AI systems into regulated products will only apply three years after the AI Act entered into force.[177]

In order to ensure a smooth transition for the fulfillment of new obligations of the AI Act, the Commission introduced the AI Pact.[178] It is a voluntary initiative that gives companies the opportunity to demonstrate and share their commitment on layed out objectives of the AI Act as well as the opportunity to early implement its key obligations.[179]

[168] Commission (2023), *Commission welcomes political agreement*, p. 1.
[169] FLI, *Timeline of Developments*, accessed on 6.5.2024.
[170] Legislative Train Schedule, *AI Act*, accessed on 6.5.2024.
[171] Legislative Train Schedule, *AI Act*, accessed on 6.5.2024.
[172] Legislative Train Schedule, *AI Act*, accessed on 6.5.2024.
[173] Commission (2023), *Commission welcomes political agreement*, p. 2.
[174] Commission (2023), *Commission welcomes political agreement*, p. 2.
[175] EU Parliament, *European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*, art. 113(a).
[176] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 113(b).
[177] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 113(c).
[178] Europeam Commission, *AI Pact*, accessed on 6.5.2024.
[179] Europeam Commission, *AI Pact*, accessed on 6.5.2024.

CHAPTER 6

European Artificial Intelligence Act - a future proof solution?

Ups and downs in the history of AI layed out the exponentially increasing application of AI in society. Further, *The future goal of achieving trustworthy AI* stated the resulting necessity to establish an ethical and legal regulatory framework to address new arising challenges due to the opaque, autonomous and complex characteristics of the technology. In relation to that, *Applicable legal acts and concerns posed by AI* gave an introduction to important applicable legal sources and their concerns regarding AI to underpin the importance of establishing new regulatory framework. *European Union's journey towards trustworthy AI* then showed the long wired path of the European Union towards the new regulatory framework addressing AI, called AI Act.

As the journey shows, to counteract the emerging challenges as well as ensuring trustworthy AI the European Commission published the proposal of the AI Act. The establishment of a legal framework regarding AI should enable further use of the its provided advantages, especially to tackle urgent societal challenges in areas such as climate protection, sustainable infrastructure, health and well-being, quality education and digital transformation.[1] Worldwide, the AI Act is the first comprehensive law on AI that already has been agreed on.[2] It has the aim to establish harmonised rules in the European Union for the whole application cycle of AI.[3,4,5] That is realized by the approach of preventing an mitigating

[1] 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*, p. 1.
[2] European Parliament, *EU AI Act: first regulation on artificial intelligence*, accessed on 6.5.2024.
[3] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[4] Piachaud-Moustakis (2023), "The EU AI Act", p. 8.
[5] Mökander et al. (2022), "Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation", p. 244.

harmful behavior by the application of AI to further enable its acceptance in society and therefore its application and ongoing development.[6] Trustworthiness is ensured in this regulation by the upmost importance of respecting European Union's rights and values, particularly the CFR.[7]

Since it is the first regulation explicitly addressing AI it might serves as a benchmark for other countries[8] and might turn the EU into a hub for global trustworthy AI.[9] Therefore, when drafting this regulation particular attention was paid to ensure that it is future-proof.[10]

6.1 Definition of AI

> In the initial proposal of the Commission of 21st April 2021 an AI system was defined according to art. 3(1) as any software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.[11]

The definition approach of the Commission is closely linked to the history of AI. Over decades subareas of AI had been established following different approaches towards the implementation of a smaller part of the overall goal of AI. For the realization of these approaches different technologies and methods were used. As it can be seen, with Annex I the Commission tried to close the gap of a definition of AI by encapsulating all current established technologies and methods. To ensure that the definition is future-proof the Commission got the power to adapt Annex I by upcoming techniques and approaches.[12]

> In the legislative resolution of the European Parliament of 13th March 2024 an AI system was defined according to art. 3(1) as any machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.[13]

However, negotiations between the Council and the Parliament led to an aversion of the pre-defined list as they rather aimed to adopt a technology-neutral and uniformed

[6] Commission, *Proposal Artificial Intelligence Act*, p. 1.
[7] Mökander et al. (2022), p. 244.
[8] Schuett, "Risk Management in the Artificial Intelligence Act", p. 1.
[9] Piachaud-Moustakis (2023), p. 8.
[10] Commission, *Proposal Artificial Intelligence Act*, p. 15.
[11] Commission, *Proposal Artificial Intelligence Act*, art. 3(1).
[12] Commission, *Proposal Artificial Intelligence Act*, p. 12.
[13] EU Parliament, *European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*, art. 3(1).

approach.[14] The Council narrowed down the definition of AI to ensure that it is sufficiently clear and moved away from the approach of the pre-defined list.[15,16] Further, the Parliament required that the definition aligns to the definition that was agreed on by the Economic Co-operation and Development (OECD).[17,18] Key elements of this definition are the terms 'infers' and 'autonomy' as they clearly differentiate AI systems to other software.[19] To ensure that the definition is future-proof, it was deliberately formulated in broad terms.[20]

6.2 Legal Scope

Since the AI Act aims to address the whole application chain of AI, key participants play an important role and therefore are clearly defined within the regulation, including private and public operators.[21] Key participants to which the AI Act applies are enshrined in art. 2(1).[22]

According to art. 3(2) a provider can be anybody that develops an AI system or a GPAI, including anybody that has an AI system or GPAI developed. Whether it is against payment or for free, if they place them on the marked or puts it into service under its own name or trademark they are considered a provider and therefore must fulfil the requirements layed out by the AI Act.[23]

According to art. 3(4) a deployer is anybody using an AI system under its authority. Only the use in the context of a personal non-professional activity is excluded.[24]

Initially, the proposal only considered two key participants, the provider and deployer.[25] According to the proposal, art. 2(1)(a-c) states that the regulation applies to every provider and deployer of an AI system whether the actual system or the resulting outcome is used within the EU.[26,27] The legislative solution formally adapted that and expanded art. 2(1)(a) by providers that place any general-purpose AI model on the market.[28]

[14] Fernhout et al., *The EU Artificial Intelligence Act: our 16 key takeaways*, accessed on 6.5.2024.
[15] EU 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*, art. 3(1).
[16] Legislative Train Schedule, *Artificial intelligence act In "A Europe Fit for the Digital Age"*, accessed on 6.5.2024.
[17] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[18] Legislative Train Schedule, *AI Act*, accessed on 6.5.2024.
[19] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[20] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[21] Commission, *Proposal Artificial Intelligence Act*, p. 12.
[22] Commission, *Proposal Artificial Intelligence Act*, art. 2(1).
[23] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 3(2).
[24] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 3(4).
[25] Commission, *Proposal Artificial Intelligence Act*, art. 2(1).
[26] Arzt et al. (2022), "Artificial Intelligence and Data Protection: How to Reconcile Both Areas from the European Law Perspective", p. 49.
[27] Commission, *Proposal Artificial Intelligence Act*, art. 2(1)(a-c).
[28] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2(1)(a).

Summed up art. 2(1)(a-c) states that the AI Act applies to: (1)EU deployers of AI systems, (2)EU and non-EU providers that place any AI system or GPAI models in the EU's market and (3)providers and deployers of non-EU AI systems if the output of the system is used within the EU.[29]

Compared to the proposal, the legislative resolution of the parliament also expanded the scope of application by other key participants as importers, distributors and product manufacturer of any AI system.[30] Further, it also applies to authorised representatives of providers that are not established in the Union as well as any affected person who lives in the European Union.[31]

Further, exceptions where obligations of the regulatory do not apply to are listed.[32] These are first activities of research, development and prototyping that are preceding the release on the market of an AI system.[33] Second AI systems that are exclusively for military, defence or national security purposes irrespective of the type of institution that carries out these activities are also excluded from the scope of this regulation.[34]

6.3 Risk Categories

The AI Act implements a risk-based approach to either prohibit or regulate specific applications of AI systems regarding their potential of harm.[35] Four risk categories were considered for the establishment of these risk categories: unacceptable risk, high risk, limited risk and minimal risk.[36,37]

1. Unacceptable risk: Specific practices of AI systems bear the potential of unacceptable risks to safety, security and fundamental rights and are therefore categorised here.[38] In example such practices are social scoring, facial recognition and manipulation.[39] Within the regulatory framework of the AI Act any AI systems that are including such practices are prohibited.[40] Thus, no general rules on detecting a prohibited AI system had been established, only specific AI practices had been banned.

2. High risk: Specific application areas of AI systems are bearing high-risks for the

[29] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2(1)(a-c).
[30] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2(1)(d-e).
[31] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2(1)(f-g).
[32] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2(3-8).
[33] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2(6).
[34] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2(3).
[35] EU Parliament, *AI Act - European Parliament legislative resolution*, rectical 26.
[36] Commission, *Proposal Artificial Intelligence Act*, p. 12.
[37] Mökander et al. (2022), p. 245.
[38] Steinkjer et al., *4 – Artificial Intelligence Act: Safe, reliable and human-centred artificial intelligence*, accessed on 8.5.2024.
[39] Steinkjer et al., *AI Act: Safe, reliable and human-centred AI*, accessed on 8.5.2024.
[40] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 5.

personal health, security or fundamental rights and are therefore categorised here.[41] In example such areas with the potential of high-risks are areas of critical infrastructure, education, law enforcement and medical devices.[42] Within the regulatory framework of the AI Act they are so called high-risk AI systems and are the heart of the regulation. Providers of any AI system falling under this category have to comply with specific requirements and obligations.[43]

3. Limited risk: Specific usages of AI systems are only bearing the potential of limited risks and are therefore categorised here.[44] In example such usages are including chatbots or deep fakes.[45] Within the regulatory framework of the AI Act the usage of such AI systems are regulated by the requirement of specific transparency obligations.[46] Thus, no general rules on detecting such usages had been established, only specific applications of AI systems had been listed.

4. Minimal risk: Specific AI systems are only bearing the potential of minimal risks and are therefore categorised here.[47] In example the majority of AI system available in the European Union in 2021 belong to that categorization such as AI-enabled video games or spam filters.[48,49] Within the initial regulatory framework provided by the proposal of the AI Act they were largely left unregulated. That changed with the regulation of GPAI models enshrined in the legislative resolution of the European Parliament.

The necessity of establishing a precise definition of the terms risk, high-risk and low-risk was already stated by stakeholder as it anchored in the initial proposal of the AI Act.[50] Nevertheless, the definition of risk only had been formally adopted after the final agreement on the regulation.

According to art. 3(2) in the legislative resolution of the European Parliament, risk is defined as the combination of the probability of an occurrence of harm and the severity of that harm.[51]

[41] Steinkjer et al., *AI Act: Safe, reliable and human-centred AI*, accessed on 8.5.2024.
[42] Steinkjer et al., *AI Act: Safe, reliable and human-centred AI*, accessed on 8.5.2024.
[43] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 9-17.
[44] Steinkjer et al., *AI Act: Safe, reliable and human-centred AI*, accessed on 8.5.2024.
[45] Steinkjer et al., *AI Act: Safe, reliable and human-centred AI*, accessed on 8.5.2024.
[46] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 50.
[47] Steinkjer et al., *AI Act: Safe, reliable and human-centred AI*, accessed on 8.5.2024.
[48] Piachaud-Moustakis (2023), p. 9.
[49] FLI, *High-level summary of the AI Act*, accessed on 8.5.2024.
[50] Commission, *Proposal Artificial Intelligence Act*, p. 8.
[51] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 2.

6.3.1 Prohibited AI practices

The AI Act prohibits all AI systems that make use of practices bearing the potential of posing a major risks to the safety of people.[52,53,54,55,56] If any other Union law already prohibits an AI practice the AI act does not contradict with that.[57] All prohibited practices are listed in table 6.1.

One major point of discussion during the negotiations about the AI Act was the ban on real-time biometric identification for law enforcement purposes as well as its exceptions.[58] In the legislative resolution of the Parliament they excluded the ban of real-time biometric identification for law enforcement purposes in the case of searching for victims of human trafficking or sexual exploitation, or for the prevention of terrorist attacks.[59] However, relying on such exception requires thorough assessments, technical and organizational measures, notifications and an arrest warrant.[60]

6.3.2 High-risk AI systems

The heart of the AI Act is the strict and extensive regulation of high-risk AI systems.[61,62] Therefore companies developing, importing, distributing or deploying any AI system should determine whether it constitutes as a high-risk AI system. According to art. 6 in the legislative resolution of the European Parliament, following AI systems are classified as high-risk:[63]

1. Any AI system that is used as a safety component of a product or itself is a product covered by European Union's laws in Annex I and further is required to undergo a third-party conformity assessment under those laws on Annex I.

2. Any AI system that is listed in Annex III.

2.1. Excluded from 2. are AI systems that fulfil one or more of the following conditions:

2.1.1. If the AI system performs a narrow procedural task.

2.1.2. If the AI system improves the result of a previously completed human activity.

[52] Piachaud-Moustakis (2023), p. 8.
[53] FLI, *The AI Act*.
[54] Commission (2023), *Commission welcomes political agreement on Artificial Intelligence Act*, p. 1.
[55] Commission, *Proposal Artificial Intelligence Act*, p. 12.
[56] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 5.
[57] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 5(1a).
[58] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[59] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[60] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[61] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[62] FLI, *High-level summary of the AI Act*, accessed on 6.5.2024.
[63] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 6.

Art. 5(1)	prohibited practice
(a)	AI systems that deploy subliminal, manipulative, or deceptive techniques in order to distort a person's behaviour and impair the person's ability to make an informed decision-making, causing significant harm to anybody.
(b)	AI systems that exploit any vulnerability of a person or a specific group of persons related to age, disability, or socio-economic circumstances to distort behaviour, causing significant harm to anybody.
(ba)	The use of biometric categorisation systems inferring with sensitive data like race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. Except if lawfully acquired biometric datasets are labelled or filtered.
(c)	AI systems that includes an evaluation or classification of anybody based on their social behavior or personal traits which further causes detrimental or unfavourable treatment of those people, also known as social scoring.
(da)	AI system that practice risk assesment of an individual that committed criminal offenses solely on the basis on profiling or personality traits. Except: If it is used to support the human assessments based on objective, verifiable facts directly linked to a criminal activity.
(db)	AI systems that comply facial recognition databases by untargeted scraping of facial images from the internet or CCTV footage.
(dc)	AI systems that are inferring emotions of an individual in the area of workplaces or educational institutions. Except if it is needed for a medical or safety reasons.
(d)	The use of 'real-time' remote biometric identification in publicly accessible spaces for the purpose of law enforcement. Except for targeted searches of victims, the prevention of substantial and imminent threats to life or a foreseeable terrorist attack, and if a suspects of a serious crime get identified.

Table 6.1: Prohibited AI practices in the AI Act

2.1.3. If the AI system detects decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment without proper human review.

2.1.4. If the AI system performs a preparatory task to an assessment relevant for the purpose of the use cases listed in Annex III.

2.3. Regardless of exceptions mentioned in 2.1, an AI system that is listed in Annex III and performs profiling of natural persons shall always be considered to be high-risk.

The exceptions of high-risk AI systems that are given in 2.3 are going to be very relevant if the AI Act enters into force, as many providers will try to argue that their provided system does not pose any high-risk.[64] Reason for that is that they will try to circumvent the high regulatory burden and costs that are accompanied with the qualification of a high-risk AI system.[65] If there is an objection that an AI system does not have a high risk potential, its provider must document its assessment before the system is placed on the market or put into operation.[66] Nevertheless, if their objection goes through the AI system will still be registered in the European Union's database of high-risk AI systems, a database gathering information about any existing high risk AI system that falls under this regulation, before their market placement or operation.[67] To ensure accessibility and transparency the referred database will be publicly accessible and shall serve as a central repository that gathers detailed information about high risk AI systems.[68]

Requirements and obligations

The AI Act imposes strict obligations when it comes to the application of AI systems that are categorised as high-risk. They do not only apply to providers and deployer but also to importers and distributors of such systems.

However, most of the established obligations are directed to the provider of an high-risk AI system.[69] They must comply with very strict requirements in order to ensure their trustworthiness, transparency and accountability.[70] Before they place their high-risk AI system on the market they must test their systems according to the layed out rules and register their systems in the EU database of high-risk AI systems.[71] That database is publicly accessible.[72] Among many other obligations following requirements must be ensured:

[64] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[65] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[66] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[67] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[68] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 71.
[69] FLI, *High-level summary of the AI Act*, accessed on 6.5.2024.
[70] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[71] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[72] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.

1. A risk management system throughout the whole life cycle of any high-risk AI system must be established, implemented, documented and maintained.[73]

2. High-risk AI systems that are using data as the base of their training process must conduct data governance, ensuring that training, validation and testing datasets are relevant, sufficiently representative and, to the best extent possible, free of errors and complete according to the intended purpose.[74]

3. Draw up a technical documentation before its market placement to demonstrate compliance and provide authorities with the information to assess that compliance. That documentation must be kept up-to-date.[75]

4. Design high risk AI system in a manner that technically allows an automated record keeping of events that are relevant for identifying national level risks and substantial modifications throughout the system's life cycle.[76]

5. Provide instructions for the use of an high-risk AI system as a supply for deployers to further interpret the system's output and enable its correct application.[77]

6. Design their high-risk AI system in a manner that allows deployers to implement human oversight and allows for intervention.[78]

7. Design their high-risk AI system to achieve an appropriate level of accuracy, robustness and cybersecurity.[79]

8. Establish a quality management system to ensure compliance with the AI Act.[80]

Obligations to the deployer are mainly requiring them to use any high-risk AI system only within their intended purpose of use and according to the instructions of use that must be provided by the provider.[81,82] That obligation will be the foundation of any possible discussion about liability, as providers will certainly will bring the argument of deployers using the system divergent to the instructions of use.[83] As the provider is subject to the obligation of designing their system in a manner that allows automated record keeping, the deployer further has the obligation to monitor the input data and operation of the system and keep that gathered data for at least six months.[84] Similar to that, the provider is subject to the obligation of designing their system in a manner

[73] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 9.
[74] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 10.
[75] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 11.
[76] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 12.
[77] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 13.
[78] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 14.
[79] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 15.
[80] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 17.
[81] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 8.
[82] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 26(1).
[83] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[84] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 26(5-6).

that allows to implement human oversight. To complement this the deployer must install human oversight to the greatest possible extent.[85]

Finally, obligations to importers and distributors are mainly related to verifying whether their imported or distributed high-risk AI system is in compliance with the regulatory framework.[86,87] The only difference is that importers need to check the systems compliance through the verification of various documentations while distributors are only required to check the compliance of their systems.[88,89,90]

Despite all that, the AI Act implemented a mechanism that shifts responsibilities towards other members of the supply chain.[91] Art. 25 of the legislative resolution of the European Parliament lays out the rule that importers, distributors, deployers or other third parties can be considered a provider of an high-risk AI system if one of the following three conditions are met: (1) they put their name or trademark on the system after its market placing or operation, (2) they made substantial modifications after its market placing or operation assuming the system remains high-risk or (3) they modified the systems intended purpose which makes the system high-risk.[92]

6.3.3 Specific transparency obligations of certain AI systems

Some AI systems are only allowed if specific transparency obligations are provided. Most of these obligations do not apply to some circumstances for law enforcement[93,94] or when the system is used for creative, satirical or similar purposes. Whether the obligation is required by the provider or deployer it shall be provided at latest at the first interaction or exposure.[95] Additionally all transparency obligations shall not affect other requirements and obligations set out by the AI Act for high-risk AI systems or Union or national law.[96]

AI system that are directly interacting with a natural person only providers are the subject to obligations.[97] They need to ensure that the person interacting with the system is aware of interacting with an AI system, if not aleady obvious.[98] Further, the use of biometric categorisation or emotion recognition systems, if not already prohibited, only requires obligations for deployers. They must inform persons exposed to the system about

[85] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 26(2).
[86] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 23.
[87] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 24.
[88] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 23.
[89] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 24.
[90] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[91] Fernhout et al., *The EU AI Act: our 16 key takeaways*, accessed on 6.5.2024.
[92] 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*, art. 25.
[93] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 50(1).
[94] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 50(2).
[95] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 50(5).
[96] EU Parliament, *AI Act - European Parliament legislative resolution*, art 50(6).
[97] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 50(1).
[98] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 50(1).

its usage and ensure a safe processing of involved personal data according to regulation (EU) 2016/679 and (EU) 2016/1725 as well as directive (EU) 2016/280.[99]

AI systems generating audio, image, video or text content are requiring obligations of the provider as well as of the deployer. Deployers of AI systems generating deep fakes must disclose that the content has been artificially generated or manipulated.[100] The same applies to generated or manipulated text for the use of communicating matters of public interest to the public.[101] Further, obligations for providers are applying to the use of AI systems as well as GPAI systems.[102] Providers of such systems that are used for generating audio, image, video or text content must ensure that the output of these systems is marked machine-readable as artificially generated or manipulated unless it is used for standard editing and does not alter input data or its semantic.[103] They are also requested to ensure that their technical solutions are effective, interoperable, robust and reliable as far as feasible.[104]

6.3.4 General Purpose AI Models

> According to art. 3(44b) in the legislative resolution of the European Parliament the definition of general purpose AI model is any AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are released on the market.[105]

General purpose AI models (GPAI models) are specifically regulated within the provided framework of the AI Act.[106,107] Reason for that is that a model must be somehow regulated as a model will never be categorized as a (high-risk) AI system, as it simply is not an AI system.[108,109] The only exception here are GPAI systems that are build on top of a GPAI models as they might be categorized as a (high-risk) AI system.[110,111]

[99] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 50(3).
[100] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 50(4).
[101] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 50(4).
[102] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 50(2).
[103] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 50(2).
[104] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 50(2).
[105] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 3(63).
[106] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 51.
[107] European Parliament, *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*, accessed on 6.5.2024.
[108] EU Council, *AI Act - General Approach*, p. 6.
[109] Ruschemeier (2023), "AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal", p. 369.
[110] EU Council, *AI Act - General Approach*, p. 6.
[111] Ruschemeier (2023), "AI as a challenge for legal regulation", p. 369.

> According to art. 3(44e) in the legislative resolution of the European Parliament the definition of general purpose AI system is any AI system which is based on a general purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.[112]

Obligations for GPAI models are distinguishing between obligations that apply to all GPAI models and additional obligations that apply only to GPAI models that carry a systematic risk.[113,114]

> According to art. 3(65) in the legislative resolution of the European Parliament the definition systematic risk is any risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain.[115]

All providers of GPAI models must comply with the following obligations[116], with the exception of providers of GPAI models with free and open licenses, which only have to comply with the latter two obligations, provided they do not pose systematic risks[117]:

1. Create and maintain technical documentation, including training and testing process and evaluation results.

2. Supply downstream providers that intend to integrate the GPAI models into their own AI system with information and documentation in order to provide a common understanding of capabilities and limitations of the GPAI models.

3. Establish a policy to respect the Copyright Directive.

4. Create and publish a detailed summary about the used content for training the GPAI models.

In addition to the obligations above, providers of GPAI models with systemic risk must also comply to following obligations:[118]

1. Perform model evaluations, including conducting and documenting adversarial testing to identify and mitigate systemic risk.

2. Assess and mitigate possible systemic risks, including their sources.

3. Track, document and report serious incidents and possible corrective measures to the AI Office and relevant national competent authorities without undue delay.

4. Ensure an adequate level of cybersecurity protection.

[112] EU Parliament, *AI Act - European Parliament legislative resolution*, Art. 3(66).
[113] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 53.
[114] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 55.
[115] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 3(65).
[116] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 53(1).
[117] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 53(2).
[118] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 55(1).

Until harmonized standards are adopted to ensure these above listed obligations, providers of GPAI models with systemic risk may relay on codes of practice.[119] Institutions such as the AI Office will support affected companies in the development of codes of practice based on a dialog with stakeholders.[120]

6.4 Further important provisions

Already in the 'White Paper', from which the proposal of the AI Act arised, the necessity of respecting fundamental rights in the application of AI was stated.[121] Given that, the assumption that a suitable mechanism would be implemented in the new regulation is be justified, however, the proposal of the AI Act only formally described its intention of ensuring a high level protection of fundamental rights in the rectials rather than in the articles.[122] While the council started to incorporate some obligations considering the protection of fundamental rights[123], it was only in the legislative resolution of the parliament that enshrined the obligation to the deployer to carry out a Fundamental Rights Impact Assessment (FRIA) to ensure the protection of fundamental rights.[124] Nevertheless, it is only required for high-risk AI systems.[125]

The approach of filing complaints that the AI Act follows is rather unusual, as there is practically no requirement of any legal standing.[126] Any citizen who has any reason to believe that the AI Act has been infringed has the right to lodge a complaint with a market supervisory authority.[127] Further any citizen has the right to receive explanations about resulting decisions created through the use of AI and the main elements of their decision process, however, that only applies to high-risk AI systems that are listed in Annex III.[128]

The governance structure of the AI Act is rather complex and layered. It involves multiple entities as notifying and notified bodies, conformity assessment bodies, an AI Board[129],

[119] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 55(2).
[120] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 56(1).
[121] Bertaina et al. (2024), "Fundamental Rights and Artificial Intelligence Impact Assessment: A New Quantitative Methodology in the Upcoming Era of Ai Act", p. 7.
[122] Bertaina et al. (2024), "FRAIA: A New Quantitative Methodology in the Upcoming Era of AI Act", p. 7.
[123] Bertaina et al. (2024), "FRAIA: A New Quantitative Methodology in the Upcoming Era of AI Act", p. 8.
[124] Bertaina et al. (2024), "FRAIA: A New Quantitative Methodology in the Upcoming Era of AI Act", p. 8.
[125] Bertaina et al. (2024), "FRAIA: A New Quantitative Methodology in the Upcoming Era of AI Act", p. 9.
[126] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 85.
[127] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 85.
[128] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 86.
[129] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 65.

an AI Office[130] as well as national competent[131] and market surveillance authorities[132]. These entities play a crucial role in the various measures in supporting innovations as the AI regulatory sandboxes, as the legislative resolution of the parliament included the promotion of them as well as real-world-testing set up by the national authorities to develop and train innovative AI before it is launched on the market.[133] Further, it promotes measures for SMEs and start-ups, as the legislative resolution of the parliament stated the necessity to promote businesses, particularly SMEs in their AI development to avoid undue pressure from industry giants that control the value chain.[134]

Finally, violations of the rules laid out by the AI Act are leading to fines that are ranging from 35 million euros or 7% of the global turnover of a company to 7.5 million euros or 1.5% of their global turnover.[135] Therefore, fines resulting from non-compliance with the AI Act are depending on the infringement as well as the size of the company.[136]

6.5 The interplay of AI Act, AILD and revised PLD

Partially, the AI Act is a product safety legislation, as it seeks to prevent and reduce harm from AI related risks by specifying safety standards.[137,138] It has its focus on preventing risks before its deployment and placement on the market. However, it will not stop the use of AI in society and due to its unpredictability, risks can never be ruled out completely.[139] That is why a safety legislation is usually complemented by a liability legislation.[140] In case a risk nevertheless gets materialised, the liability legislation seeks to ensure that harmed parties can be adequately compensated.[141] The current existing law is addressing this liability gap with the PLD which also applies to AI systems but unfortunately it does not add up to its autonomy, complexity and opacity.[142]

After various different parties pointed out that existing compensation gap, the Commission stated the necessity to ensure the same level of protection for the application of AI as it is currently provided for other technologies.[143] That is why the scope of the Commission's

[130] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 64(1).
[131] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 70.
[132] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 74(2).
[133] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 57-61.
[134] EU Parliament, *AI Act - European Parliament legislative resolution*, art. 62-63.
[135] European Parliament, *AI Act: deal on comprehensive rules for trustworthy AI*, accessed on 6.5.2024.
[136] European Parliament, *AI Act: deal on comprehensive rules for trustworthy AI*, accessed on 6.5.2024.
[137] Ziosi et al. (2023), "The EU AI Liability Directive (AILD): Bridging Information Gaps", p. 2.
[138] Adamakis (2023), "A Comparative Analysis of the EU and US Artificial Intelligence (AI) Regulation Regimes", p. 16.
[139] Ziosi et al. (2023), "The EU AILD: Bridging Information Gaps", p. 2.
[140] Launders, *Beyond the AI Act: The AI Liability Directive the Product Liability Directive*, accessed on 10.5.2024.
[141] Launders, *Beyond the AI Act*, accessed on 10.5.2024.
[142] Ziosi et al. (2023), "The EU AILD: Bridging Information Gaps", p. 2.
[143] European Commission, *WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust*, p. 15.

approach included a revamp of the existing PLD and the proposal of the AILD.[144,145] These two directives are building the foundation of regulating AI in the European Union and are closely interconnected with the AI Act.[146] As the AI Act is laying out rules and obligations for research and the whole life cycle of AI these two directives are laying out rights of individuals harmed by AI.[147,148] As a result the AI Act is not a standalone piece of legislation and should be seen in the wider context of the Commission's approach of ensuring an effective regulation of AI.[149]

6.5.1 Key changes in the revised PLD

Revising the PLD had the ambition of modernising the current no-fault-based product liability regime in order to keep up with the digital age.[150] Still, it continues to pursue a strict liability regime.[151] One of the upmost important changes considering AI is the expansion of its scope, which includes a more comprehensive definition of both the liable persons and the products.[152] The previous notion of producer is now broadened by encompassing economic operators as the manufacturer of a product or a component, the provider of a related service, the authorized representative, the importer, the fulfillment service provider or the distributor.[153] Further, the term product is going to include both software as well as software updates, regardless of whether it is an embedded or a standalone solution, including AI.[154,155] Only non-commercial open source software is excluded to promote innovation and broaden access to software.[156]

Another important change is the redefinition of defect and damage. Within the revised version of the PLD the term defect is also considering any effect on the product arisen through any ability to continue to learn after deployment.[157] In particular that expansion was introduced due to the ability of AI to learn and develop after its deployment. Further the term damages was expanded to encompass the loss and corruption of data unless exclusively used for professional purposes.[158] However, if the flaw of the product could

[144] 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)*, pp. 1, 11.
[145] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 17.
[146] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 17.
[147] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 17.
[148] Ziosi et al. (2023), "The EU AILD: Bridging Information Gaps", p. 2.
[149] Launders, *Beyond the AI Act*, accessed on 10.5.2024.
[150] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 17.
[151] Ziosi et al. (2023), "The EU AILD: Bridging Information Gaps", p. 2.
[152] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", pp. 32-33.
[153] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 33.
[154] European Commission, *Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on liability for defective products*, art. 4(1).
[155] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 32.
[156] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 32.
[157] European Commission, *Proposal revised PLD*, art. 6(1)(c).
[158] European Commission, *Proposal revised PLD*, art. 4(6)(c).

not be detected due to the state of scientific and technical knowledge at the relevant time the accused producer is granted with protection.[159,160]

Finally, an approach of lightening the burden of proof was implemented.[161] If a claimant is unable to specify the cause of the damage due to the technical or scientific complexity of the damage, it can be presumed on the basis of sufficiently relevant evidence.[162] Further, a claimant asserting a plausible claim may demand that the defendant disclose relevant evidence.[163] If he fails to do so, a defect is presumed.[164]

6.5.2 Key provisions in the AILD

Compared to the strict liability approach of the PLD, the AILD is concerned with a fault-based liability for damages caused by AI.[165] The directive aims to compensate for damage caused intentionally or negligently and applies to any application of AI, regardless of most provisions only applying to high-risk AI systems.[166] AI as well as high-risk AI is defined according to the definitions enshrined in the AI Act.[167] Similar to that, the AILD encompasses both providers and deployer as defined in the AI Act and therefore deviates from the approach of the PLD of defining producer and users[168].

Under the AILD a presumption of causality is enshrined that enables claimants to seek compensation on any damage caused by an AI system with a broader approach of burden of proof to raise the chance of a successful claim. A rebuttable presumption of causality is enshrined in art. 4 by establishing a casual link between the violation of a duty of care under Union or national law and the output of an AI system or the inability of the AI system to produce an output that caused the damage.[169] That article would apply if all of the following conditions are met:

(1) The claimant showed the presence of a violation of a certain EU or national obligation that is relevant to the accused harm of an AI system and therefore caused the damage.[170]

(2) It must be reasonably likely that the output of an AI system or the inability of the AI system to produce an output that caused the damage was a result of a negligent behavior of the defendant, considering the circumstances of the individual case.[171]

[159] European Commission, *Proposal revised PLD*, art. 10(1)(e).
[160] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 36.
[161] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 34.
[162] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 34.
[163] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 34.
[164] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 34.
[165] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 28.
[166] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", pp. 27-28.
[167] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 27.
[168] Adamakis (2023), "Comparative Analysis: EU and US AI Regulation", p. 28.
[169] European Commission, *Proposal AILD*, art. 4.
[170] European Commission, *Proposal AILD*, art. 4(1)(a).
[171] European Commission, *Proposal AILD*, art. 4(1)(b).

> (3) The claimant demonstrated that the damage was causes by any produced output of an AI system or as a result of the inability of the AI system to produce an output.[172]

Therefore the burden of proof does not exclusively rely on the complainant as the person liable has to prove that the conditions of liability are not fulfilled. However, the defendant is enabled to rebut this given presumption of causality if it is possible to prove that its fault could not have been the cause of damage.[173]

Further, the high number of involved parties in the life cycle of AI systems often makes it hard for a claimant to identify the person that is potentially liable for a caused damage. Therefore art. 3(1) of the AILD enshrined that national courts have the power to order disclosure of evidence of an high-risk AI system if a suspect exists that it has caused any damage.[174] That provision should help claimants to identify the liable person more easily. Further, art. 3(3) enshrined that a claimant can further request disclosure of evidence from other persons that are not the defendants but only in cases where all other attempts to obtain the evidence from the defendant have been unsuccessful.[175]

[172] European Commission, *Proposal AILD*, art. 4(1)(c).
[173] European Commission, *Proposal AILD*, rectical 30.
[174] European Commission, *Proposal AILD*, art. 3(1).
[175] European Commission, *Proposal AILD*, art. 3(3).

CHAPTER 7

Discussion

AI is a hotly discussed topic worldwide, especially as the technology was faced with a new hype in 2011. Reason for that were advantages taken in subareas of AI that started to show the real power behind this technology. Associated with that, its autonomy, opacity and complexity started to grow which also arose risks next to the newly given opportunities. While these risks might lead to harmful behaviour, the positive aspects have the ability to crucially benefit environment and society. Hence, it is important to push the ongoing development of AI by establishing a legal regulatory framework to prevent its potential risks. With that goal in mind the European Union introduced the AI Act. In order to analyze the future-proof approach of the AI Act the following research questions were singled out and will be discussed.

What conclusions can be drawn from comparing the history of AI with the journey of the European Union towards achieving trustworthy AI?

Before the term AI was coined its preceding concepts were based on the idea of replicating human intelligence within a technical environment. As it developed into an independent field of research, this idea was retained as the goal of AI and was further defined in more detail under the term AGI. Later on, it was extended to the goal of achieving ASI, machines with capabilities that go far beyond those of human beings. This leads to the conclusion that the heart of AI has always been and still is the human being and how they function. Due to different manifestations of moral values as well as the autonomy and free-will of each individual, human beings are relying on the legal system to establish a well-functioning, harmonious society. It is therefore obvious that the technical replication of a human being also requires to be regulated by the legal system in order to enable a frictionless embedding into society.

As early as 1942, a science fiction author recognized this necessity and established three laws of robotics in one of his short stories. These laws also applied to AI, as it was considered part of robotics. However, it was not until 2017 that the government of the EU made use of them in their 'Civil Law Rules on Robotics' (2018/C 252/25). It can be argued that early limitations of the technology were the reason that not enough attention was paid to its possible autonomy which further led to disregarding its legal aspects. Instead, the focus relied on scaling up AI systems to more complex problems and making them act independently. Their limitations gave the impression of the technology not being capable of delivering its expected performance, reinforced by optimistic predictions of experts in the field of AI that did not materialize. That is why its possible autonomy was not even considered at that period of time and the main concern of governments worldwide was the decision of whether or not to continue funding the research field of AI.

Therefore, even if the original goal of AI already encompassed the extent of its potential autonomy, its actual dimension only became tangible once the earlier limitations of the technology had been overcome. With advantages taken in subareas as ML and DL as well as with its enabled progress by the invention and application of BD, the research field of AI faced another boom in 2011 and its real autonomy, complexity and opacity started to show. That boom also marked the start of the current ongoing hype about AI. Nevertheless, as stated above, it was not until 2017 that the topic AI was included into discussions and upcoming plans of the government of the EU. In the same year AI started to turn into an independent field of governance and discussions about specifically regulating that technology started. Although, new liability issues arising from AI were already pointed out, the focus then mainly relied on ethical rather than on legal aspects. That was underpinned by the argument that all legal rights and obligations remain binding and must continue to be complied with. Finally, in 2019 an analysis of the NTF stated the importance of a legal regulatory framework as it pointed out the issue of the current in forced law not being able to deal with the autonomy, opacity and complexity of AI. From this date on discussions about an implementation of a regulatory framework regarding AI started.

Based on that it can be said that the need of a regulatory framework for the application of AI could have been recognized earlier than it actually was. Nevertheless, it is important to highlight that the EU still managed to be the first lawmaker worldwide to establish a regulatory framework regarding AI. Further, the trigger of the AI Act, the 'White Paper', was published in 2020 shortly before the COVID pandemic broke out. Therefore, as a worldwide shift of priority took place for the government, the Commission nevertheless managed to publish their proposal in 2021.

Further, the history of AI has shown the importance of giving realistic rather than optimistic and exaggerated outlooks in order to prevent a loss of interest and trust in society. Nevertheless, enthusiastic forecasts and promises were made throughout the history of AI governance in the European Union. Within this thesis two cases were singled out.

First, Ursula von der Leyen pledged that within hundred days of taking on the role as

the president of the European Commission she would propose some legislation on AI. That statement was given while the AI HLEG were establishing 'Ethics guidelines for trustworthy AI' and other important documents related to AI governance. Therefore, it was foreseeable that the promise was a bit overachieved. If Ursula von der Leyen would have taken into account the current work of the expert group that euphoric outlook could have been formulated in a way more realistic way. Fortunately, it did not restricted the progress of AI governance as experts were already onto establishing a regulatory framework regarding AI. Still, it caused frustration among experts that were already involved into the topic as they knew that these promises could not be met. Unfortunately, within the scope of this thesis it was not possible to analyze the impact it had on society and therefore future research could invest into this topic.

Second, the European Union stated that the AI Act will be the first regulatory framework addressing AI and further might become an international standard as they already achieved that with the enforcement of the GDPR in terms of data protection. The first part of the statement was met as the final agreement on the AI Act was reached on the 9th December 2023. Still, it is questionable if this regulation is able to keep up with the dynamic and fast development of AI. Failure to do so could negatively impact the trust of society, which could be a long-lasting consequence. Within the analysis of the second research question the future-proof approach of the AI Act will be discussed in more detail. In terms of the promise that the AI Act will follow the path of success that had been reached with the GDPR in becoming an international standard it is important to note that the AI Act is dealing with the most challenging technology that has ever existed. In contrast to other technologies, AI is outstanding due to its complexity, opacity and autonomy. However, these characteristics are also the reason why the current legal scope is not able to cover the entire application of AI. Compared to that the GDPR is a technology-neutral regulation that is dealing with the usage of data, which only evolves in its volume and velocity. Therefore, it is important to take a closer look at the future development and application of the AI Act to verify the fulfillment of the promise that the regulation is going to evolve to an international standard. As the regulation is not enforced yet this analysis will be up to future research.

Besides that, the history of AI had shown the difficulty of finding a unique definition due to its extensive scope and constant development. Because of that, subdomains have evolved over time where each has its focus on one of its different characteristics. One of the most important subdomains is learning which encompasses a variety of AI practices like ML, DL and NN. That issue of agreeing on a unique definition of AI was also reflected in the journey of the EU towards trustworthy AI. Even before the proposal of the AI Act, many experts delivered a variety of possible definitions of AI. Within the proposal of the AI Act the Commission followed an approach that can be closely linked to the history of AI. They encapsulated all current established technologies and methods that are considered as AI under the term AI system. To align with the constant development of this technology the Commission got the power to adapt that list continuously. Positive aspects of this approach are its easy decision process of whether a technology is considered

as AI and its easy possibility to adapt in order to keep up with the future development of AI. However, the history of AI has also shown that many subareas of AI have already been invented and applied way before the related subarea has been established. Adapting this list could therefore prove to be a very difficult and complex task as it is closely intertwined with the development of its broad research field. Further, that approach bears the risk that the adaption process might not pace up with the development of AI which might result into a constant lag between AI governance and its technological state of the art.

Deviating from the proposed approach of the Commission, in the final agreement of the AI Act a rather technology-neutral and uniform approach was implemented. Within this context the definition of AI system is no longer extended by a list of practices instead it was aligned with the definition that was agreed on by the OECD. Special attention was paid to include the terms 'infers' and 'autonomy' as they clearly differentiate AI systems to other software. To further ensure that the definition is future-proof, it was deliberately formulated in broad terms. The major benefit of this approach is that it does not require to be continuously adapted as it already encompasses future developments in the field of AI by incorporating the major component of autonomy. However, definitions that are deliberately formulated in broad terms are leaving plenty of room for interpretation and discussion.

In summary, it can be said that both approaches have advantages and disadvantages that are exactly opposite to each other. While one leaves plenty of room for interpretation and discussion, the other is very precisely defined. Still, the precise approach has a weakness in its continuous need for adaptation, which the other closes with its deliberately formulated definition. That leads to the conclusion that the combination of both approaches would have been even more future-proof as they would compensate for each other's weaknesses. A deliberately formulated definition supported by a list of practices and methods that are considered as AI could be an improvement on the current agreement of the definition of AI system. As an example, if the applied practice or method is already listed in the collection there would be no need to discuss whether any system applying them falls under the deliberately formulated definition of AI. Conversely, if a practice is not covered by the collection of practices and methods, the deliberately formulated definition of AI would cover that. That improvement is further underpinned by the fact that it is already necessary to keep track of the development of AI, as the revised PLD enshrines that if any flaw of a product could not be detected due to the state of scientific and technical knowledge at the relevant time the accused producer is granted with protection.

Another aspect that emerged throughout the history of AI was the importance of categorizations which were established due to the lack of a unique definition in order to define its scope more precisely. Through that it was feasible to gain a better understanding of the possibilities of this technology. Despite the fact that the EU was able to agree on a definition of AI systems to approach the term AI the mechanism of establishing categorizations was applied as well. The AI Act implements a risk-based approach where AI systems get categorized by their potential threat to fundamental rights of human

beings and based on that they are either banned or allowed if specific regulations and obligations are met. These categories are following an approach of a predefined collection of practices or application areas of AI. Based on that it can be said that mechanism to narrow down the functionality of AI throughout its history had also been used in the process of the EU in establishing a regulatory framework for this technology.

Despite that, the EU did not made use of any of the already existing categorizations. Including them could have been an improvement to ensure the future-proof concept of the AI Act. As an example, they could have incorporated the classification that describes the capability of AI of emulating human beings. As AI systems that would belong to the category ANI would only be able to perform a simple task and cannot grow in their ability, these systems could have been regulated by the approach that is now implemented in the AI Act. Further, AI systems belonging to the category AGI would be more autonomous and could be applied to more application areas. Therefore they could have been handled similarly as in the current approach but more stringent. In example, the execution of the FRIA could have been a mandatory task for any AI system under this category. Finally, by including the concept of ASI it would have been ensured to cover the possible state of AI being more powerful than any human being. Although these improvements would strengthen the future-proof approach of the AI Act it would be difficult to deliberately formulate them in a way that they are legally secure, leavening little room for interpretation.

At last, the significance of interrelated disciplines and funding emerged throughout the history of AI. It was not possible to gain any insights into the investment of interrelated disciplines of AI by the EU government within this thesis. Therefore the analysis of the provided support of all research overlaps of AI will be up to future research. Similarly to that, through the journey of the EU towards trustworthy AI not many insights into the investment of funding could be gained. However, many discussion of the EU AI governance at least included the topic and further two reports of the AI HLEG were solely focusing on investment suggestions. Within the scope of this thesis it was not possible to further analyse the suggested plan and their implementation and effectiveness thus it will be up to future research. Important to note is that during the research of this thesis some sources stated the current shift towards private investments as Google, etc. rather than funding by government. Within the scope of this thesis it was not possible to further analyse the current state of the art about private investments and its efficacy thus it will be up to future research.

How does the AI Act ensures that it is future-proof and what existing legal concerns were taken into account during its development?

In order to analyse if the AI Act was implemented as a future-proof approach, prior loopholes in the existing legal system and their consideration in the new regulation are being discussed. Although many existing regulations were taken into account during the development process of the AI Act, it was only possible to focus on a small selection of regulations within the scope of this thesis: the CFR, the GDPR and the PLD.

Due to its biased decision process, AI is one of the greatest threats to fundamental rights of human beings. That is why the EU has decided to put the protection of the CFR at the heart of the AI Act. The established risk categories of this regulation are based on the possible threats of an AI system and its application area and purpose regarding the CFR. AI systems that are bearing an unacceptable risk to either safety, security or fundamental rights are strictly banned. These restrictions are formulated as a collection of forbidden practices and use cases. Although this approach offers a basic network of protecting fundamental rights of any human being, it is missing a mechanism of easily detecting unacceptable threats of any system that is not already banned. However, it is most likely that these excluded systems will be categorized as high-risk which further are required to adhere to a large number of regulations and obligations. That also includes a FRIA to ensure the protection of fundamental rights. It is important to note that the FRIA was not included into the AI Act before its final agreement even though it always had its heart in the protection of fundamental rights. The proposal of the AI Act only formally enshrined that protection in its recticals and its classification of high-risk AI systems. Despite it being incorporated into the final agreement of the AI Act it only applies to high-risk AI systems rather than any AI system. That is why the protection of fundamental rights is not completely ensured as the FRIA is only obligatory for high-risk AI systems. Thus, if any system is not already prohibited or classified as high-risk, the FRIA might not be considered as it is not mandatory. Based on that it can be said that the AI Act serves with a base protection of fundamental rights, thus, extending the scope of the FRIA to any AI system that is not already prohibited would be a possible improvement. Nevertheless, within the scope of this thesis the effectiveness of the FRIA could not be analysed and is therefore an open topic to future research in order to discuss the real dimension of the protection of fundamental right.

Further, data is the fuel of AI and therefore the GDPR is one of the most important regulations that must be complied with the application of AI. That is why existing loopholes of the GDPR must be covered by the AI Act. Automated decision-making and profiling are two AI practices that were posing major challenges to the GDPR as they were only partially covered within that regulation. To cover automated decision-making processes, the AI Act enshrines that any citizen has the right to receive explanations about resulting decisions created through the use of AI and main elements of their decision process. However, that regulation only applies to systems categorized as high-risk that are listed in Annex III. Invariably classified as high-risk AI systems are those including the practice of profiling. Systems under the high-risk category are required to fulfill a large amount of regulations and obligations within the AI Act. Among other things, these are leading to a better risk management, data management, system documentation as well as the possibility of human oversight and the ability to intervene. Additionally, these regulations and obligations are providing a foundation to monitor and prevent personal data re-identification. However, that cannot be prevented completely due to the ability of AI to connect non-identified data to the related individual. It only enables the possibility to eventually determine the event which gives the possibility to cancel or reverse the current task. Still, that process must be implemented within the AI system

itself as it cannot be executed by any human being due to its autonomous way of working. Finally, within the AI Act many requirements for the applied data of an AI system were enshrined, whether it is used for the training or application of any system. Nevertheless, within the scope of this thesis it was not possible to go into more detail of the provided coverage by the AI Act of the loopholes that arose from the application of AI in the GDPR and therefore that analysis is up to future research.

In the event of damage caused by the use of AI, liability is regulated by the PLD which follows a no-fault-based approach. In regards to this directive, most of its challenges that were arising through the application of AI were mainly about its definitions. As its origin layed in the area of mass-production it was feasible to implement precise definitions of the terms product, producer and defect. However, they are not suitable for the broad field of AI and its complexity. An analysis has exposed these challenges and as a consequence of that the Commission proposed the AILD as well as a revision of the PLD. Their main goal was to ensure the same level of protection for any application of AI as it is currently provided for traditional technologies. Within the revised PLD definitions have been adapted to suit the technological state of the art and the burden of proof procedure has become less strict. Further, the PLD is now accompanied by the AILD which is concerned with fault-based liability for damages caused by AI. Based on that it can be said that challenges of the old PLD in regards to AI previously found within the scope of this thesis were considered within the proposal of these two directives. Despite that, it is up to future research to analyse its implementation as well as the effectiveness of this approach.

The main goal of the AI Act is to promote trustworthy AI by supporting its innovation while accompanied risks are being prevented or at least mitigated. Therefore it is important to analyse the actual extent of balance between these two goals that the new regulation is going to provide. Looking at the goal of preventing or at least mitigating the risks of the application of AI it can be said that this requirement is the fundament of the AI Act. It classifies any AI system into one of four distinct categories based on their potential threat to safety, security and fundamental rights: either unacceptable, high-risk, limited risk or minimal risk. Depending on their category they get either prohibited or regulated whereas the degree of regulation also varies. Furthermore, as the previous discussion has shown, this risk-based approach is able to cope with many of the existing legal gaps. Therefore it s definitely an improvement of the current legal situation. Based on that. it can be said that the AI Act is suitable as a protection of preventing risks arising from the application of AI. Although it leaves room for improvement, the possibility of enhancement is given always and everywhere as things are constantly growing. Especially in regards to AI as it is the most challenging technology that ever existed. It is also important to note that the AI Act is the first legal framework regarding AI worldwide and therefore could not be created on the basis of any other law. With its establishment an important milestone in the history of international AI governance was taken.

On the other side it is important to analyze the extend to which the innovation of AI is supported by the enforcement of the AI Act. First and foremost the EU explicitly

excluded the field of research in the scope of the regulatory framework which enables continuous and unrestricted research and development of the technology. However, the application of AI is limited due to its established exclusionary system. If the use of AI systems is limited, future development is suffering as well, as its application provides for a broader test environment. Thus, it is essential for obtaining improvement and new ideas. Further, regulations and obligation required by the AI Act are placing many demands on AI systems in order to bring them to market. Thereby, the time to marked might be prolonged as it is time-consuming to implement and fulfill them. This entails the risk that AI systems developed in other countries will overtake or undercut EU partners in the development of AI solutions. Finally, the implementation of these regulations and obligations is not only time-consuming, it also entails costs. To avoid undue pressure from industry giants that are controlling the value chain the EU governance enshrined specific support for SMEs and startups within the AI Act.

Besides that, AI companies operating in the EU are also benefiting from the enforcement of the AI Act. If their systems are complying with the regulation they have the advantage of marketing their product as trustworthy. This might boost the demand of their products and services. Additionally, companies do not have to think about certain critical areas and their implementation themselves due to prescribed rules of the legal act. This leaves more time that can be directly invested into innovation. Furthermore, the determined standardized system that comes with the AI Act creates a level playing field that ensures fair competition. Finally, as the EU has positioned itself as a pioneer in AI governance by enforcing that regulatory framework, so do the companies that adhere to it. In addition to companies benefiting from its enforcement, the EU itself gains advantages. First, they did not had to adapt to any other legal framework when they were enacting the AI Act. Second, their legal framework could set a precedent for other countries to follow. This is underpinned by the fact that all AI companies operating within the EU must comply with the rules of the regulation.

Within the research of this work, one regulation of the AI Act stood out in particular due to its rather unusual approach. Art. 85 enshrines that practically no requirement of any legal standing is required in order to file complaints. Anybody in believe of any infringement of the AI Act has the right to lodge complaint with a market supervisory authority. As a result, an unmanageable number of these could arise, which could lead to major problems. Critical cases could be processed too late or, in the worst case, not be dealt with at all.

Reflection

In conclusion, the enforcement of the AI Act provides a clear framework for the application of AI and is an important milestone worldwide. Despite the fact that the need for a regulatory framework could have been recognized earlier, the EU still managed to be the first mover in the field of AI governance. Important aspects that emerged throughout the history of AI were partially met. Both processes have applied similar mechanisms as the unfolding of subdomains and categorizations. Still, already gained knowledge of these mechanisms could have been utilized within the AI Act. This applies in particular to the difficulty of finding a unique definition of AI as envisaged in the approach of the Commission. Further, within processes of the EU governance almost no exaggerated outlooks or time schedules were given. One exception was the promise of the AI Act becoming an international standard as it was already achieved with the establishment of the GDPR. Second, an external statement by Ursula von der Leyen was critical as well, as it was obvious that it was rather unrealistic.

As AI created many loopholes in the current enforced law, it was mandatory to cover them within the enforcement of the AI Act. Research taken within this thesis has shown that they were at least considered within the scope of the new regulatory. However, the effectiveness remains subject to future research. Finally, the future-proof approach of the AI Act had the aim of supporting its innovation while preventing and mitigating accompanied risks. The second part is covered as risk prevention is the heart of the AI Act. Despite potential improvements it definitely improves the current legal situation and is suitable as a protection of preventing risks arising from the application of AI. Nevertheless, innovation might face a setback due to the required time and costs of the implementation phase of the AI Act. As a fact that is not indispensable, as neither the introduction of a regulatory framework regarding AI is. Any approach to AI regulation would have been time-consuming and costly. Therefore, the implementation phase as well as the provided support during this period is crucial to ensure that the AI Act is future-proof.

CHAPTER 8

Conclusion

AI had, still has, and is going to have an enormous influence on our daily lives. Compared to other technologies it is outstanding due to its complexity, opacity, and autonomy. That technology provides with crucial competitive advantages to benefit the environment and society which might take companies and countries to the next level. Despite that, AI also bears the potential to harm society or an individual and thereby its trustworthiness is negatively impacted. Nonetheless, for AI to bring advantages to society, trust in decisions made by these systems is required. Otherwise, society might avoid or even refuse its usage.

That is why the European Union had already been following the path of achieving trustworthy AI for many years with the final result of the proposal and agreement of the AI Act. To keep pace with the constant development and increasing capabilities of AI it follows a future-proof approach to ensure two important aspects. First, maintaining the technological leadership of the EU. Second, that newly developed technologies and their functioning are in line with the values, fundamental rights, and principles of the Union.

Nevertheless, the enforcement of a regulatory framework regarding AI bears many challenges. Due to the unique characteristic of AI it is the most challenging technology that ever existed. Its complexity, autonomy and opacity makes it almost impossible to deliberately formulate a legal framework that covers every aspect of the technology. Despite that, the implementation phase for a new regulation is time-consuming and costly. Therefore, the enforcement of the AI Act bears the risk of AI providers from countries overtaking or undercutting EU partners in the development of AI solutions as they do not have to comply with the regulation.

Despite these challenges, the AI Act still serves as a crucial foundation in the field of AI governance, not only in the European Union even internationally. AI is and is going to be one of the most interrupting and powerful technologies that has ever existed. To benefit from this technology it is upmost important to prevent and mitigate its potential

risks, thus, establishing a legal regulatory framework regarding AI was an essential step that has now been taken with the establishment of the AI Act. Due to its emphasis on innovation as well as ethical considerations it reflects the need of benefiting from AI while mitigating its potential harm. However, moving forward it is essential to closely monitor the enforcement of the AI Act to ensure that the right balance is given between fostering innovation and protecting societal values and rights. Further, future AI governance must take greater account of the development history of AI, as this could speed up its processes and avoid repeated errors.

At last, it is important to note that the AI Act is not a standalone piece of legislation in the field of AI governance. It must be seen in the wider context of the EU law. Of particular note are the GDPR, the accompanied proposal of the AILD and the revised PLD.

List of Figures

List of Tables

Bibliography

Articles

Adamakis, Eleftherios. “A Comparative Analysis of the EU and US Artificial Intelligence (AI) Regulation Regimes”. In: *International Hellenic University* (2023). URL: http://hdl.handle.net/11544/30384.

AI, High-Level Expert Group on. “Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence”. In: *EU Publications* KK-02-20-527-EN-N (July 2020). DOI: 10.2759/943666.

Aizenberg, Evgeni and Jeroen van den Hoven. “Designing for human rights in AI”. In: *Big Data Society* 7.2 (Aug. 2020), pp. 1–14. ISSN: 2053-9517. DOI: 10.1177/2053951720949566.

Akinrinola, Olatunji et al. “Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability”. In: *GSC Advanced Research and Reviews (GSCARR)* 18.03 (Feb. 2024), pp. 50–58. ISSN: 2582-4597. DOI: 10.30574/gscarr.2024.18.3.0088.

Antunes, Henrique Sousa. “Civil Liability Applicable to Artificial Intelligence: A Preliminary Critique of the European Parliament Resolution of 2020”. In: *Monograph Book* (Dec. 2020). DOI: 10.1007/s12027-023-00751-y.

Artzt, Matthias and Tran Viet Dung. “Artificial Intelligence and Data Protection: How to Reconcile Both Areas from the European Law Perspective”. In: *Vietnamese Journal of Legal Sciences* 7.2 (2022), pp. 39–58. ISSN: 2719-3004. DOI: 10.2478/vjls-2022-0007.

Bertaina, Samuele et al. “Fundamental Rights and Artificial Intelligence Impact Assessment: A New Quantitative Methodology in the Upcoming Era of Ai Act”. In: (Jan. 2024), pp. 1–42. DOI: 10.2139/ssrn.4698609.

Boire, Richard. “Understanding AI in a world of big data”. In: *Big data & information analytics* 2.5 (2017), pp. 23–43. ISSN: 2380-6974. DOI: 10.3934/bdia.2018001.

Chen, Weiru et al. “Systematic analysis of artificial intelligence in the era of industry 4.0”. In: *Journal of Management Analytics* 10.1 (2023), pp. 89–108. ISSN: 2327-0012. DOI: 10.1080/23270012.2023.2180676.

Couch, James R. “Artificial Intelligence: Past, Present and Future”. In: *Journal of the South Carolina Academy of Science* 21.1 (2023), pp. 1–2. ISSN: 1553-5975. URL: https://scholarcommons.sc.edu/jscas/vol21/iss1/2.

Deng, Li. “Artificial Intelligence in the Rising Wave of Deep Learning: The Historical Path and Future Outlook”. In: *IEEE Signal Processing Magazine* 35.1 (Jan. 2018), pp. 180–177. ISSN: 1558-0792. DOI: 10.1109/MSP.2017.2762725.

Donahoe, Eileen and MacDuffee M. Megan. “Artificial Intelligence and Human Rights”. In: *Journal of Democracy* 30.2 (Apr. 2019), pp. 115–126. ISSN: 1086-3214. DOI: 10.1353/jod.2019.0029.

Dorr, Laura. “Types of Artificial Intelligence, Explained”. In: *Dental Products Report* 56.12 (Dec. 2022), pp. 38–39. URL: https://www.proquest.com/trade-journals/types-artificial-intelligence-explained/docview/2760891917/se-2.

Efficiency of Justice (CEPEJ), European Commission for the. “European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment”. In: *Council of Europe* (Dec. 2018). URL: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

Elvin, Rob, Nicola Smith, and Francesca Puttock. “Reconciling Artificial Intelligence (AI) With Product Safety Laws”. In: *Squire Patton Boggs* (Dec. 2023). URL: https://www.squirepattonboggs.com/en/insights/publications/2023/12/reconciling-artificial-intelligence-ai-with-product-safety-laws.

Floridi, Luciano. “The European Legislation on AI: A Brief Analysis of its Philosophical Approach”. In: *Philosophy Technology* 34.2 (June 2021), pp. 215–222. ISSN: 2210-5441. DOI: 10.1007/s13347-021-00460-9.

Fradkov, Alexander L. “Early History of Machine Learning”. In: *IFAC-PapersOnLine* 53.2 (2020), pp. 1385–1390. ISSN: 2405-8963. DOI: https://doi.org/10.1016/j.ifacol.2020.12.1888.

Gundugurti, Prasad Rao et al. “Ethics and Law”. In: *Indian J Psychiatry* 64.Suppl 1 (Mar. 2022), pp. 7–15. DOI: 10.4103/indianjpsychiatry.indianjpsychiatry_726_21.

Haenlein, Michael and Andreas Kaplan. "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence". In: *California Management Review* 61.4 (2019), pp. 5–14. DOI: 10.1177/0008125619864925.

Haigh, Thomas. "How the AI Boom Went Bust". In: *Commun. ACM* 67.2 (Jan. 2024), pp. 22–26. ISSN: 0001-0782. DOI: 10.1145/3634901.

Hilchenbar, Leonore and Vitorio Dimov. "AI and the GDPR". In: *Härting* (Dec. 2023). URL: https://haerting.de/en/insights/ai-and-the-gdpr/.

Janssen, H, M.S.A Lee, and J Singh. "Practical fundamental rights impact assessments". In: *International journal of law and information technology* 30.2 (2022), pp. 200–232. ISSN: 0967-0769. DOI: 10.1093/ijlit/eaac018.

Kalota, Faisal. "A Primer on Generative Artificial Intelligence". In: *Education Sciences* 14.2 (2024). ISSN: 2227-7102. DOI: 10.3390/educsci14020172.

Kaur, Davinder et al. "Trustworthy Artificial Intelligence: A Review". In: *ACM Comput. Surv.* 55.2 (Jan. 2022), pp. 1–38. ISSN: 0360-0300. DOI: 10.1145/3491209.

Khan, Fatima Hameed, Muhammad Adeel Pasha, and Shahid Masud. "Advancements in Microprocessor Architecture for Ubiquitous AI - An Overview on History, Evolution, and Upcoming Challenges in AI Implementation". In: *Micromachines (Basel)* 12.6 (2021), p. 665. ISSN: 2072-666X. DOI: 10.3390/mi12060665.

Kusak, Martyna. "Quality of data sets that feed AI and big data applications for law enforcement". In: *ERA-Forum* 23.2 (2022), pp. 209–219. ISSN: 1612-3093. DOI: 10.1007/s12027-022-00719-4.

Leong, Yee Rock. "Rethinking Human Motivation Psychology: The Hierarchy of Human Fear Model". In: (Nov. 2023). DOI: 10.31234/osf.io/ef3vs.

Leslie, David et al. "Artificial intelligence, human rights, democracy, and the rule of law: a primer". In: *The Council of Europe* (2021). DOI: 10.5281/zenodo.4639743.

Li, Shu, Michael Faure, and Katri Havu. "Liability Rules for AI-Related Harm: Law and Economics Lessons for a European Approach". In: *European journal of risk regulation* 13.4 (2022), pp. 618–634. ISSN: 1867-299X. DOI: 10.1017/err.2022.26.

Liability, Expert Group on and New Technologies – New Technologies Formation. "Liability for artificial intelligence and other emerging digital technologies". In: *Publications Office of the European Union* DS-03-19-742-EN-N (Nov. 2019). DOI: 10.2838/25362.

Liu, Haochen et al. "Trustworthy AI: A Computational Perspective". In: *ACM Trans. Intell. Syst. Technol.* 14.1 (Nov. 2022), pp. 1–59. ISSN: 2157-6904. DOI: 10.1145/3546872.

Matos Pinto, Inês de. “The draft AI Act: a success story of strengthening Parliament’s right of legislative initiative?” eng. In: *ERA-Forum* 22.4 (2021), pp. 619–641. ISSN: 1612-3093.

Mayor, Adrienne. “What Pandora’s Box tells us about AI”. In: *World Economic Forum* (Oct. 2018). URL: https://www.weforum.org/agenda/2018/10/an-ai-wake-up-call-from-ancient-greece/#:~:text=For%20her%20part%2C%20Pandora%20was,she%20ever%20age%20or%20die..

Melanie, Mitchell. “Why AI is Harder Than We Think”. In: *CoRR* abs/2104.12871.v2 (Apr. 2021), pp. 1–12. DOI: 10.48550/arXiv.2104.12871.

Mökander, Jakob et al. “Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation”. In: *Minds and machines* 32.2 (2022), pp. 241–268. ISSN: 0924-6495. DOI: 10.1007/s11023-021-09577-4.

Natasa, Anamaria, Monica Mihaela Maer Matei, and Cristina Monacu. “Artificial Intelligence: Friend or Foe? Experts’ Concerns on European AI Act”. In: *Economic computation and economic cybernetics studies and research* 57.3/2023 (2023), pp. 5–22. ISSN: 0424-267X. DOI: 10.24818/18423264/57.3.23.01.

Navas, Susana. “Producer Liability for AI-Based Technologies in the European Union”. In: *Canadian Center of Science and Education* 9.1 (Aug. 2020), pp. 77–84. ISSN: 1927-5234. DOI: 10.5539/ilr.v9n1p77.

Piachaud-Moustakis, Bianca. “The EU AI Act”. In: *Pharmaceutical Technology Europe* 35.11 (Nov. 2023), pp. 8–9. ISSN: 1753-7967. URL: https://www.proquest.com/scholarly-journals/eu-ai-act/docview/2889704236/se-2.

Pizzi, Michael, Mila Romanoff, and Tim Engelhardt. “AI for humanitarian action: Human rights and ethics”. In: *International review of the Red Cross (2005)* 102.913 (2020), pp. 145–180. ISSN: 1816-3831. DOI: 10.1017/S1816383121000011.

Pollmaecher, Thomas. “The DGPPN congress 2022: Ethics, law and mental health”. In: *Nervenarzt* 93.11 (2022), pp. 1091–1092. ISSN: 0028-2804. DOI: 10.1007/s00115-022-01392-1.

Reusch, Benedikt. “Handlungsfähigkeit durch, trotz und gegenüber (Big) Data und KI: Eine Bestandsaufnahme mit Hilfe des Frankfurt-Dreiecks”. In: *Ludwigsburger Beiträge zur Medienpädagogik* 23 (Oct. 2023), pp. 1–28. ISSN: 2190-4790. DOI: 10.21240/lbzm/23/18.

Robles Carrillo, Margarita. “Artificial intelligence: From ethics to law”. In: *Telecommunications Policy* 44.6 (2020), p. 101937. ISSN: 0308-5961. DOI: https://doi.org/10.1016/j.telpol.2020.101937.

Rodríguez de las Heras Ballell, Terese. “The revision of the product liability directive: a key piece in the artificial intelligence liability puzzle”. In: *ERA Forum* 24.2 (July 2023), pp. 247–259. DOI: 10.1007/s12027-023-00751-y.

Ruschemeier, Hannah. “AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal”. In: *ERA-Forum* 23.3 (2023), pp. 361–376. ISSN: 1612-3093. DOI: 10.1007/s12027-022-00725-6.

Samoili, S. et al. “AI Watch Defining Artificial Intelligence”. In: *Publications Office of the European Union* EUR 30117 EN (Feb. 2020). ISSN: 1831-9424. DOI: 10.2760/382730.

Sartor, Giovanni. “Artificial intelligence and human rights: Between law and ethics”. In: *Maastricht journal of European and comparative law* 27.6 (2020), pp. 705–719. ISSN: 1023-263X.

Schuchmann, Sebastian. “Analyzing the Prospect of an Approaching AI Winter”. In: (May 2019). DOI: 10.13140/RG.2.2.10932.91524.

Schuett, Jonas. “A Legal Definition of AI”. In: *SSRN Electronic Journal* abs/1909.01095 (2019). DOI: 10.2139/ssrn.3453632.

— “Risk Management in the Artificial Intelligence Act”. In: *European journal of risk regulation* (), pp. 1–19. ISSN: 1867-299X. DOI: 10.1017/err.2023.1.

Secinaro, Silvana et al. “The role of artificial intelligence in healthcare: a structured literature review”. In: *BMC medical informatics and decision making* 21.1 (2021), pp. 125–125. ISSN: 1472-6947. DOI: 10.1186/s12911-021-01488-9.

Stergiou, Konstantinos D. et al. “A Machine Learning-Based Model for Epidemic Forecasting and Faster Drug Discovery”. eng. In: *Applied sciences* 12.21 (2022), p. 10766. ISSN: 2076-3417.

Tan, Haocheng. “A brief history and technical review of the expert system research”. In: *IOP Conference Series: Materials Science and Engineering* 242.1 (Sept. 2017), p. 012111. DOI: 10.1088/1757-899X/242/1/012111.

Taye, Mohammad Mustafa. “Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions”. In: *Computers (Basel)* 12.5 (2023), p. 91. ISSN: 2073-431X. DOI: 10.3390/computers12050091.

Thiebes, Scott, Sebastian Lins, and Ali Sunyaev. “Trustworthy artificial intelligence”. In: *Electronic Markets* 31.2 (June 2021), pp. 447–464. ISSN: 1422-8890. DOI: https://doi.org/10.1007/s12525-020-00441-4.

Toosi, Amirhosein et al. “A Brief History of AI: How to Prevent Another Winter (A Critical Review)”. In: *PET Clinics* 16.4 (Oct. 2021), pp. 449–469. ISSN: 1556-8598. DOI: 10.1016/j.cpet.2021.07.001.

Turing, Alan M. “I.—COMPUTING MACHINERY AND INTELLIGENCE”. In: *Mind* LIX.236 (Oct. 1950), pp. 433–460. ISSN: 0026-4423. DOI: 10.1093/mind/LIX.236.433.

Tzafestas, Spyros G. “Ethics and Law in the Internet of Things World”. In: *Smart Cities* 1.1 (2018), pp. 98–120. ISSN: 2624-6511. DOI: 10.3390/smartcities1010006.

Walsh, Kenneth R., Sathiadev Mahesh, and Cherie C. Trumbach. “Autonomy in AI Systems: Rationalizing the Fears”. In: *The Journal of technology studies* 47.1 (2021), pp. 38–47. ISSN: 1071-6084. URL: https://www.jstor.org/stable/48657934.

Wróbel, Izabela. “Artificial intelligence systems and the right to good administration”. In: *Review of European an Comparative Law* 49.2 (May 2022), pp. 203–223. ISSN: 2545-384X (online). DOI: 10.31743/recl.13616.

Zamora-Cárdenas, W., M. Zumbado, and Trejos-Zelaya. “McCulloch-Pitts Artificial Neuron and Rosenblatt ” s Perceptron : An abstract specification in Z”. In: *Technology Inside by CPIC* 5.5 (Apr. 2020), pp. 16–29. ISSN: 2215-5392. URL: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://towardsdatascience.com/perceptron-the-artificial-neuron-4d8c70d5cc8d&ved=2ahUKEwiIyfD0_sGFAxVQxQIHHcorCZkQFnoECBIQAQ&usg=AOvVaw0DR8cuYM5BSR6xIGgSyipy.

Završnik, Aleš. “Criminal justice, artificial intelligence systems, and human rights”. In: *ERA-Forum* 20.4 (2020), pp. 567–583. ISSN: 1612-3093.

Ziosi, Marta et al. “The EU AI Liability Directive (AILD): Bridging Information Gaps”. In: *European Journal of Law and Technology* 14.3 (June 2023). DOI: https://dx.doi.org/10.2139/ssrn.4470725.

Books

Artificial Intelligence, High-Level Expert Group on. *Ethics guidelines for trustworthy AI*. Publications Office, 2019. DOI: doi/10.2759/346720.

Gerards, J.H. et al. *Getting the future right – Artificial intelligence and fundamental rights – Report*. European Union Agency for Fundamental Rights, June 2020. ISBN: 978-92-9474-860-7. DOI: 10.2811/774118.

Hassanien, Aboul-Ella, Mohamed Hamed N Taha, and Nour Eldeen M Khalifa. *Enabling AI Applications in Data Science*. 1st ed. Studies in Computational Intelligence. Cham:

Springer, Sept. 2021. ISBN: 978-3-030-52067-0. DOI: 10.1007/978-3-030-52067-0.

Parliament, European et al. *The impact of the General Data Protection Regulation (GDPR) on artificial*. PE 641.530. Publications Office, June 2020. ISBN: 978-92-846-6771-0. DOI: 10.2861/293.

Russell, Stuart J., Peter Norvig, and Ernest Davis. *Artificial Intelligence: A Modern Approach*. Third edition, Global edwiition. Prentice Hall Series in Artificial Intelligence. Upper Saddle River, NJ: Pearson, 2010. ISBN: 0-13-604259-7. URL: https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf (accessed: Jan. 27, 2024).

— *Artificial Intelligence: A Modern Approach*. 4th ed. Pearson Series in Artificial Intelligence. Upper Saddle River, NJ: Pearson Education, 2010. ISBN: 1-292-40113-3.

Santosh, KC and Casey Wall. *AI, Ethical Issues and Explainability—Applied Biometrics*. eng. 1st ed. 2022. SpringerBriefs in Computational Intelligence. Singapore: Springer Nature Singapore Imprint: Springer, 2022. ISBN: 9811939357. URL: 10.1007/978-981-19-3935-8.

Book Chapters

Kanal, Laveen N. "Perceptron". In: *Encyclopedia of Computer Science*. Ed. by Anthony Ralston, Edwin D. Reilly, and David Hemmendinger. GBR: John Wiley and Sons Ltd., 2003, pp. 1383–1385. ISBN: 0470864125.

Kim, Haesik. "Historical Sketch of Artificial Intelligence". In: *Artificial Intelligence for 6G*. Cham: Springer International Publishing, Feb. 2022. Chap. 1, pp. 3–14. ISBN: 978-3-030-95041-5. DOI: 10.1007/978-3-030-95041-5_1.

Kunkel, Carsten and Juliana Schoewe. "Zur Zulässigkeit automatisierter Entscheidungen im Einzelfall einschließlich Profiling im Sinne des Art. 22 DSGVO – Praxisrelevanz und Wirksamkeit der Norm in Zeiten von Big Data und KI". In: *Künstliche Intelligenz in der Anwendung: Rechtliche Aspekte, Anwendungspotenziale und Einsatzszenarien*. Ed. by Thomas Barton and Christian Müller. Angewandte Wirtschaftsinformatik. Wiesbaden: Springer Fachmedien Wiesbaden, Feb. 2021. Chap. 02, pp. 9–23. ISBN: 978-3-658-30936-7. DOI: 10.1007/978-3-658-30936-7_2.

Panesar, Arjun. "What is Artificial Intelligence?" In: *Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes*. Apress, 2020. Chap. 1, pp. 1–19. ISBN: 978-1-4842-3799-1. DOI: 10.1007/978-1-4842-3799-1.

Taulli, Tom. "KI-Grundlagen". In: *Grundlagen der Künstlichen Intelligenz: Eine nicht-technische Einführung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2022, pp. 1–

19. ISBN: 978-3-662-66283-0. DOI: 10.1007/978-3-662-66283-0_1. URL: https://doi.org/10.1007/978-3-662-66283-0_1.

Governance

Commission, European. *Evaluation of Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0157&from=EN. May 2018.

— *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. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496.

— *Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on liability for defective products*. COM(2022) 495 final. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0495.

— *WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust*. COM(2020) 65 final. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0065.

Committee, European Economic Social. *Opinion of the European Economic and Social Committee on 'Artificial intelligence — The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society'*. (2017/C 288/01). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016IE5369.

Council, European. *European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*. P9$_T$A(2024)0138. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.

— *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*. 2021/0106(COD). URL: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf.

European Commission. *Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Comittee and the Commitee*

of the Regions Artificial Intelligence for Europe. COM (2018) 237 final. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237.

— *Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Comittee and the Commitee of the Regions Coordinated Plan on Artificial Intelligence*. COM (2018) 795 final. URL: https://eur-lex.europa.eu/resource.html?uri=cellar:22ee84bb-fa04-11e8-a96d-01aa75ed71a1.0002.02/DOC_1&format=PDF.

— *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. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1706317681006&uri=CELEX:52021PC0206.

— *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. URL: https://artificialintelligenceact.eu/wp-content/uploads/2024/01/AI-Act-FullText.pdf.

European Union, Official Journal of the. *Amendment of the Product Liability Directive*. (85/ 374/ EEC). URL: https://eur-lex.europa.eu/eli/dir/1999/34/oj.

— *Charter of Fundamental Rights of the European Union*. 2012/C 326/02. URL: http://data.europa.eu/eli/treaty/char_2012/oj.

— *Product Liability Directive*. (85/ 374/ EEC). URL: http://data.europa.eu/eli/dir/1985/374/oj.

— *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. (EU) 2016/679. URL: http://data.europa.eu/eli/reg/2016/679/oj.

Parliament, European. *European Parliament resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*. (2018/C 252/25). URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52017IP0051.

Press releases

Commission, European. *Commission welcomes political agreement on Artificial Intelligence Act*. Dec. 9, 2023. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473 (accessed: Jan. 27, 2024).

Parliament, European. *A Union that strives for more: the first 100 days*. July 16, 2019. URL: https://www.europarl.europa.eu/news/de/press-room/20190711IPR56824/parlament-wahlt-ursula-von-der-leyen-zur-prasidentin-der-eu-kommission (accessed: Jan. 27, 2024).

Webpages

BioStrand. *AI, ML, DL, and NLP: An Overview*. URL: https://www.ibm.com/blog/supervised-vs-unsupervised-learning/ (accessed: May 5, 2024).

Commission, Europeam. *AI Pact*. URL: https://digital-strategy.ec.europa.eu/en/policies/ai-pact (accessed: May 6, 2024).

Commission, European. *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. URL: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment (accessed: May 6, 2024).

— *Coordinated Plan on Artificial Intelligence*. URL: https://digital-strategy.ec.europa.eu/en/policies/plan-ai (accessed: May 6, 2024).

— *Datenschutz in der EU*. URL: https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_de (accessed: Feb. 8, 2024).

— *Declaration on the Cooperation on Artificial Intelligence*. URL: https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence#:~:text=On%2010%20April%2025%20European,European%20approach%20to%20deal%20therewith. (accessed: May 7, 2024).

— *Die Europäische KI-Allianz*. URL: https://digital-strategy.ec.europa.eu/de/policies/european-ai-alliance (accessed: May 6, 2024).

— *EU Member States sign up to cooperate on Artificial Intelligence*. URL: https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence (accessed: Feb. 11, 2024).

— *High-level expert group on artificial intelligence*. URL: https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai (accessed: May 6, 2024).

— *Joint Declaration on the EU's legislative priorities for 2018-19*. URL: https://commission.europa.eu/document/download/a2cf5d97-6d0b-474e-8a23-df6c676d406b_en?filename=joint-declaration-eu-legislative-priorities-2018_en.pdf (accessed: Feb. 11, 2024).

— *Policy and investment recommendations for trustworthy Artificial Intelligence*. URL: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence (accessed: May 6, 2024).

Council, European. *European Council meeting (19 October 2017) – Conclusions*. URL: https://www.consilium.europa.eu/media/21620/19-euco-final-conclusions-en.pdf (accessed: Feb. 11, 2024).

Differences, Key. *Difference between Law and Ethics*. URL: https://keydifferences.com/difference-between-law-and-ethics.html#Conclusion (accessed: Apr. 28, 2024).

Europe, Council of. *About the European Commission for the efficiency of justice (CEPEJ)*. URL: https://www.coe.int/en/web/cepej/about-cepej (accessed: May 5, 2024).

Fernhout, Frederieck and Thibau Duquin. *The EU Artificial Intelligence Act: our 16 key takeaways*. URL: https://www.stibbe.com/publications-and-insights/the-eu-artificial-intelligence-act-our-16-key-takeaways (accessed: May 6, 2024).

Launders, Julia. *Beyond the AI Act: The AI Liability Directive the Product Liability Directive*. URL: https://artificialintelligenceact.eu/high-level-summary/#:~:text=The%20AI%20Act%20classifies%20AI%20according%20to%20its%20risk%3A&text=Minimal%20risk%20is%20unregulated%20(including,is%20changing%20with%20generative%20AI). (accessed: May 10, 2024).

Life Institute, Future of. *High-level summary of the AI Act*. URL: https://artificialintelligenceact.eu/high-level-summary/#:~:text=The%20AI%20Act%20classifies%20AI%20according%20to%20its%20risk%3A&text=Minimal%20risk%20is%20unregulated%20(including,is%20changing%20with%20generative%20AI). (accessed: May 6, 2024).

— *Timeline of Developments*. URL: https://artificialintelligenceact.eu/developments/ (accessed: May 6, 2024).

Life Institute (FLI), Future of. *The AI Act*. URL: https://artificialintelligenceact.eu (accessed: Jan. 27, 2024).

National Human Rights Institution, European Network of. *Implementation of the EU Charter of Fundamental Rights*. URL: https://ennhri.org/wp-content/uploads/2019/11/Implementation-of-the-EU-Charter-of-Fundamental-Rights-Activities-of-NHRIs.pdf (accessed: May 6, 2024).

Parliament, Europeam. *Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI*. URL: https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai (accessed: May 6, 2024).

Parliament, European. *"Digitales Gipfeltreffen Tallinn", 29.09.2017, 29 September 2017*. URL: https://www.consilium.europa.eu/de/meetings/eu-council-presidency-meetings/2017/09/29/ (accessed: May 7, 2024).

— *EU AI Act: first regulation on artificial intelligence*. URL: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (accessed: May 6, 2024).

Pingen, Dr. Anna. *Council's Common Position on Artificial Intelligence Act*. URL: https://eucrim.eu/news/councils-common-position-on-artificial-intelligence-act/ (accessed: May 6, 2024).

Priyam, Utkarsh. *The Evolution of Parallel Distributed Processing*. URL: https://www.linkedin.com/pulse/evolution-parallel-distributed-processing-utkarsh-priyam/ (accessed: May 6, 2024).

Schedule, Legislative Train. *Artificial intelligence act In "A Europe Fit for the Digital Age"*. URL: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence (accessed: May 6, 2024).

Steinkjer, Lars Erik, Gry Hvidsten, and Ekin Ince Ersvaer. *4 – Artificial Intelligence Act: Safe, reliable and human-centred artificial intelligence*. URL: https://www.wr.no/aktuelt/4-artificial-intelligence-act-safe-reliable-and-human-centred-artificial-intelligence (accessed: May 6, 2024).

Wartner, Sandra. *Vertrauen in die Künstliche Intelligenz*. URL: https://www.risc-software.at/fachbeitraege/fachbeitrag-vertrauen-in-die-kuenstliche-intelligenz/#:~:text=Und%20fehlt%20das%20Vertrauen%2C%20fehlt,und%20Grenzen%20kennen%20und%20verstehen. (accessed: Feb. 28, 2024).