AI: A Dilemma For European Policymakers: Are European Fundamental Rights, Law And Values Safe In Times of AI?

Julian Alexander Samol (S2639882)
Management Society and Technology
University of Twente, Enschede
Wordcount: 11951
First Supervisor: Dr. Shawn Donnelly
Second Supervisor: Dr. Caroline Fischer

Summary

Artificial Intelligence (AI) is being employed in virtually every aspect of day-to-day life from text auto-correction to crime prediction. Its potential applications are unimaginable, opening room for social and economic opportunities, but also setting the stage for the emergence of unexpected problems, raising the question of how this technological diffusion affects European law, rights and values. This phenomenon is explored in this thesis by investigating the dilemma European policy makers are faced with when envisioning a market-wide implementation of AI solutions while safeguarding European fundamental rights, law and Union values. By analysing AI-related scientific literature and policy documents it was found that the issues of lacking algorithmic transparency as well as discriminatory and biassed AI challenge rights, values and the General Data Protection Regulation. Furthermore, it was found that policy makers are challenged when attempting to combat this issue in the form of the Artificial Intelligence Act proposal as well as the adoption of a framework for Trustworthy AI, whereas an assessment found that the issues AI poses are not sufficiently catered to as a result of several loopholes and deficient mechanisms.

Abbreviations: Artificial Intelligence (AI), Artificial Intelligence Act (AIA), Assessment List for Trustworthy AI (ALTAI), Charter of Fundamental Rights by the European Union (CFR), European Artificial Intelligence Board (EAIB), European Union (EU), General Data Protection Regulation (GDPR), High-Level Expert Group (HLEG), Trustworthy AI (TAI)

Table of Contents

1 Introduction

1.1 Background

In 2017, the president of the Russian Federation Vladimir Putin stated that artificial intelligence (AI) "(...) comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world." (The Verge, 2017). While at the time, AI's colossal opportunities may not have been as evident for most people, current developments reveal how AI is applicable in virtually every aspect of day-to-day life and steadily growing in relevancy. Academic articles written by AI are passing peer review (Harrison, 2023), numerous mayors of European capitals were tricked into having phone calls with a deep fake of Kyiv's major (Oltermann, 2022) and more notably, the European Union (EU) believes that AI can bring benefits to multiple sectors, such as energy, transport and health (European Parliament, 2019). With AI's social relevance noticeably increasing, research in this field is ultimately becoming increasingly relevant as well. This ongoing development is motivating political bodies including the EU to implement legislation with the aim of safeguarding citizens' rights, laws and values. Four years after the initial statement by the Russian president, the EU introduced the first version of the "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" in 2021, also known as the AI Act (AIA). The AIA addresses issues attributed to AI and lays the groundwork to "preserve the EU's technological leadership" while also ensuring that "Europeans can benefit from new technologies developed and functioning according to Union values, fundamental rights and principles" (European Commission, 2021, p.1).
Thereby, the EU is working on a legal framework to ensure that AI systems within the Union market are safe and respect existing laws such as the GDPR, as well as rights and values laid down in the Charter of Fundamental Rights of the European Union (CFR). Simultaneously, it aims at harmonising AI-related law of the Member States at the EU level, with the intention of the resulting legal certainty attracting AI providers to settle down within the European single market. In order to achieve these objectives, the AIA not only proposes several instruments and mechanisms, but also repeatedly stresses the importance of safeguarding fundamental rights and values; however, by use of AI-related literature this thesis will show that the AIA takes on a very ambitious task. For one, the EU wants to encourage the use and marketwide implementation of AI (European Commission, 2021), however, an improperly regulated roll-out of AI runs the risk of undermining existing laws, rights and values. Contrarily, heavily regulating AI and prioritising citizens' rights may diminish the appeal and advantages AI technology can bring. Furthermore, when regulating AI and aiming to facilitate the development of Trustworthy AI (TAI), companies need to be able and willing to meet requirements set out by the EU. Exploring this dilemma, this thesis aims at first, identifying AI challenges relevant in the adoption of the AIA, with the results then being able to aid in pointing out how AI can undermine EU law, rights and values, encouraging the EU to address these challenges before finally passing the AIA. With increased attention towards AI-related problems, possible damages may be prevented by influencing the policy-making process and ultimately safeguarding EU law, values and citizens' rights, or even motivating future amendments to the AIA, with a larger focus on safety measures.

1.2 Knowledge Gap

By exploring how AI challenges EU law, rights and values and assessing the EU's response to this dilemma in the form of the AIA along with its attempt to facilitate the development of TAI, this thesis aims to reduce a knowledge gap of recurring elements in both scientific literature and policy documents by identify the root of AI related challenges and explore how these are addressed or reflected in EU related (policy) documents. Furthermore, this research will extend research evaluating the framework for TAI relevant for this thesis. By doing so, insights from this thesis can be then used

to inform policymakers and aid in addressing the dilemma they are confronted with when aiming to facilitate the rollout of AI in the European single market.

1.3 Research Questions

Based on the provided aims, this research will answer the research question:

"How does Artificial Intelligence challenge the European Union's efforts of maintaining its fundamental rights, values and EU law while developing a market for trustworthy artificial intelligence?"

In order to establish a full picture of the dilemma EU policymakers are faced with, this research will also address the following sub questions:

(a) What are the existing and relevant fundamental rights and internal market regulatory standards applicable in the adoption of the AIA and how are they challenged by AI?

(b) What is Trustworthy AI and how are these AI related challenges reflected in the guidelines for Trustworthy AI?

(c) What are the currently proposed instruments of the AIA?

(d) Are these able to address the identified AI related problems and are they correctly placed?

Having laid down the research questions this thesis aims to answer, the following chapter will provide the theoretical backdrop used to do so, followed by a discussion of its methodological approach.

2 Theory

2.1 Introduction: AI: As A Challenge For Policy Makers

This section will focus on the relevant laws, rights and values for the AI discussion in the European context. It will identify potential clashes between AI and law, fundamental rights and values as well as introduce the EU's response to these challenges in the form of the AIA and the TAI framework as discussed in academic discourse. This theoretical framework will be used for an in-depth analysis in chapter four. Besides posing a challenge in itself, some of the identified challenges are not only attributable to, but also exacerbated by the complex nature of AI. As some systems involve calculations "beyond human cognitive comprehension", the complexity of AI calculations, as well as the "language of AI" are difficult to understand due to their high technicality, which is why these systems are also referred to as "black box technology" (Greeinstein, 2022). The complexity of understanding AI processes also brings challenges in defining as well as developing AI risk prognoses (Ruschemeier, 2023).

2.2 European Fundamental Rights & Foundational Values Affected By AI

As the AIA attempts to safeguard fundamental rights and union values (AIA Objective 1), the for this discussion relevant fundamental rights and values laid down in the CFR, as well as potential AI related challenges have to be identified. Generally, AI is seen as having the potential to enhance human well-being, one of the EU's main visions for AI, but on the other hand also posing serious threats to fundamental rights (Schippers, 2020). (Brkan et al., 2020) stresses that the use of AI for legal analysis purposes in a judicial setting can infringe on the right to a fair trial (Article 47), for example when legal analytics tools are used and decisions made by the systems are not explainable attributed to black-box phenomenon (Brkan et al., 2020). Closely related to this is the right to liberty and security, (Article 6). When AI systems are used for predictive policing, a person's right to liberty may be violated when a person is falsely classified as high-risk by a system due to data correlation with previously arrested people, instead of causal evidence. (Aizenberg & van den Hoven, 2020) Article 21 of the CFR entails the right to non-discrimination (European Union, 2012), which can be violated by AI in various ways. For example, data used to train AI systems can contain biases which can result in discrimination when the systems are finally used (Stahl et al., 2022). Not only listing the EU's fundamental rights, the CFR also addresses its foundational values in the preamble, "the Union is founded on the indivisible, universal values of human dignity, freedom, equality and solidarity; it is based on the principles of democracy and the rule of law. It places the individual at the heart of its activities, by establishing the citizenship of the Union and by creating an area of freedom, security and justice." (European Union, 2012, p.2). How "values embodied by the Charter's human rights provision relate to an support each other" (Aizenberg & van den Hoven, 2020, p.5) is described by (Aizenberg & van den Hoven, 2020), by highlighting how relevant AI aspects may be at odds with these values and the rights they are embodied by. For instance, human dignity is seen as the "overarching human value at stake in AI" (Aizenberg & van den Hoven, 2020, p.5) which can be violated by either discrimination or unjustified actions, privacy violations or as a result of job market transformations. Furthermore, humans can perceive their dignity as being violated through humiliation, in terms of being put in a state of helplessness or losing autonomy over one's representation, instrumentalisation, when one is treated as means to an end, or when one is made superfluous.

2.3 European Law Relevant to AI

In the European context, AI systems have to adhere to a wide array of laws and regulations, be it "EU primary law (the Treaties of the European Union and its CFR)," or "EU secondary law (such as the General Data Protection Regulation, the Product Liability Directive, the Regulation on the Free Flow of Non-Personal Data, anti-discrimination Directives, consumer law and Safety and Health at Work Directives)" (HLEG, 2019, p.6). As an in-depth analysis of AI conflicts concerning all applicable laws and regulations relevant to AI and the AIA is beyond the limitations of this research, this thesis will strictly focus on the General Data Protection Regulation (GDPR) for the reflection on AI challenges for European law. The GDPR is particularly interesting and relevant for the AI discussion as it lays down the treatment of personal data, consequently "the GDPR always applies itself to AI when AI techniques process personal data, perform profiling, as well as make automated decisions based on personal data and/or that affect the data subjects" (Lorè et al., 2023, p.6) and is therefore integral to this discussion. Relevant aspects and therefore potential points of conflict between AI technology and the GDPR are the: "principles of transparency and non-discrimination, and, especially, the right to information, the right to erasure, the right to human intervention in cases of automated decision-making and profiling" (Kesa & Kerikmäe, 2020, p.70), as well as the right to explanation (Lorè et al., 2023). Furthermore, data shall only be processed for "specific, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes" unless consent is given (Artzt & Dung, 2023, p.52), with more potential clashes arising in regards to the principles of accuracy and data minimisation (Artzt & Dung, 2023).

2.4 European Response to Governing AI: AIA

The EU lays down four objectives in the AIA and aims to:

> *(1) ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;*
> *(2) ensure legal certainty to facilitate investment and innovation in AI;*
> *(3) enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;*
> *(4) facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation." (European Commission, 2021, p.3).*

While the EU follows the GDPR in the sense of taking a risk-based approach in the AIA (Raposo, 2022) and attempts to address challenges such as the ones reflected in this thesis, scholars fear that the AIA may fall short of achieving its aims and requires improvement in many areas (Ebers et al., 2021). One cause for concern raised is the definition of AI systems within the AIA. For one, the AIA regulates 'AI systems', and leaves open what or if there is a difference between AI and AI systems (Ruschemeier, 2023), secondly, the definition for AI systems provided in the AIA is criticised as being too broad (Ebers et al., 2021) and said to be covering "almost every computer programme", instead of just AI systems (Ruschemeier 2023, p.368). As a result, the AIA runs the risk of failing to protect fundamental rights in some cases, while also (over) regulating non-AI systems (Ruschemeier, 2023). Further concerns are raised regarding missing or insufficient mechanisms such as a lack of mechanisms to boost innovation with there being just two measures within the AIA designed to promote innovation, as well as there being no legal mechanism for AI in research (Raposo, 2022), raising doubt on whether the second objective of the AIA can be realised and missing a mechanism for liability cases (Ebers et al., 2021). Some of the exceptions for certain risk-laden AI systems, such as for manipulative systems, are also seen critically with questionable or potentially unenforceable measures (Raposo, 2022) allowing for loopholes (Ebers et al., 2021). Additionally, to mitigate risks, the AIA foresees procedures to assess the risk of AI systems to determine the amount of necessary regulation for a system based on its risk profile, however, leaves out relevant parameters for how the decision regarding these risk assessments are made (Raposo, 2022) similarly the AIA provides some

high-risk system providers with a high degree of discretion when a assessing their systems conformity to AIA obligations which runs the risk of providers potentially bypassing them (Ebers et al., 2021).

2.5 Lawful, Safe & Trustworthy AI

"It is through Trustworthy AI that we, as European citizens, will seek to reap its benefits in a way that is aligned with our foundational values of respect for human rights, democracy and the rule of law." (HLEG, 2019, p.4). What constitutes lawful, safe and trustworthy AI likely differ based on cultural, legal and other contexts; generally, as raised in scientific literature, "trust requires that we have reason to believe both that AI is reliable and acting in our behalf" (von Eschenbach, 2021, p.1620). As for the European context, the EU adopted a definition provided by a High-Level Expert Group (HLEG) on AI, which was incorporated into the AIA, where systems categorised as high risk "will have to comply with a set of horizontal mandatory requirements for trustworthy AI and follow conformity assessment procedures before those systems can be placed on the Union market." (European Commission, 2021, p.3). In the HLEG's "framework for achieving Trustworthy AI" it it is explained that TAI has three components. It has to be (1) lawful, complying with all applicable laws and regulations, (2) ethical, ensuring adherence to ethical principles and values and it should be (3) robust from a technical and social perspective (HLEG, 2019). The principles for TAI cover the three aspects set out in the aims of the AIA, with the aspect of safety being covered by the HLEGs robustness principle. Derived from these three principles, the HLEG formulated the Assessment List for Trustworthy AI (ALTAI) with seven key requirements that have to be met to realise TAI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) societal and environmental well-being and finally, (7) accountability. Ideally, such a framework could be used to make sure that conflicts with European law, fundamental rights and values are addressed before a system is entering the market, however, scientific discussion criticises the general approach of this framework as leaving questions and concerns unanswered (Hickman & Petrin, 2021), raising the question whether the provisions laid out by the HLEG are potentially unattainable, thereby, against the backdrop of AI related challenges, a question is raised whether the approach for TAI chosen does not overlook critical aspects in regard to algorithmic transparency, with some systems (black box systems) posing a major challenges for TAI (Procter et al., 2023), whereas some scholars go as far as noting that "AI is not a thing to be trusted" (Ryan, 2020, p.17), which would ultimately beg to question whether goals if the HLEG framework and AIA proposal are achievable.

2.6 Preliminary Conclusion

Based on a first review of AI literature it is evident that if left unchecked, AI can clash with European law, fundamental rights and values. Furthermore, scientific discussion raises several concerns regarding the AIA itself. Oriented by the theoretical framework, a coding scheme (3.4) will be used to guide the analysis in chapter four. The relevant concepts for this thesis are, (1) *algorithmic transparency, (2) black-box technology, (3) discriminatory & biassed AI and (4) Trustworthy AI.* (1) Algorithmic transparency refers to the transparency of AI processes and the reasoning behind decisions (Kesa & Kerikmäe, 2020). (2) Whereas black box technology refers to systems where "due to the complexity of these systems and the amount of data manipulation carried out by them, it can be impossible to find out how exactly a particular decision was made by the system" (Kesa & Kerikmäe, 2020, p.70). (3) Discriminatory and biassed AI will refer to cases "when the output of a machine-learning model can lead to the discrimination against specific groups or individuals" (Belenguer, 2022, p.773). (4) Lastly, *TAI refers to* AI that is "lawful, complying with all applicable laws and regulations, (...) ethical, ensuring adherence to ethical principles and values and it should be robust from a technical and social perspective (HLEG, 2019, p.2). Having introduced the theoretical framework of this thesis along with its main concepts, the following chapter will explain the methodological approach of this thesis.

3 Methods

3.1 Introduction

As this thesis aims at highlighting the conflicts between AI and European law, fundamental rights and values as well as the resulting dilemma EU policymakers are posed with when trying to encourage the diffusion of AI technology in the European single market by implementing the AIA and envisioning the facilitation of TAI, a literature review was found to be a suitable research method. This approach allows for an in-depth analysis of the relevant aspects that can be directly applied to European law and other documents, in order to identify points of conflict.

3.2 Data Collection

In order to answer the research- and subquestions, a theoretical framework for AI challenges, fundamental rights and values, relevant GDPR provisions, TAI as well as the AIA was developed. The framework for AI related challenges only encompasses publications of the past five years in an attempt to firstly, avoid technological outdatedness as AI is continuously being developed and secondly, thereby only analysing challenges currently and therefore for the AIA relevant challenges. For this framework and the analysis in chapter four, which will answer the research question "How does Artificial Intelligence challenge the European Union's efforts of maintaining its fundamental rights, values and EU law while developing a market for trustworthy artificial intelligence?" the relevant data, in the form of documents, had to be collected. As this research heavily relies on policy-related documents by the EU, the EUR-Lex service was utilised, which provides access to EU documents. Additionally, to develop an understanding of how AI may clash with EU law, fundamental rights and values, literature on problematic aspects of AI contributing to this dilemma, literature on TAI and the AIA proposal was collected using the FINDUT function of the library of the University of Twente as well as Google Scholar. Furthermore, the bibliographies of already collected papers were scanned for additional relevant papers to include in this thesis. Finally, as of 14th of June 2023 the Commission, Council and Parliament have published their positions on the AIA, the analysis of this thesis however will strictly focus on the initial proposal of the AIA and documents published until this date.

3.3 Research Design

Section 4.1 of the document analysis will first explain the relationship between European fundamental rights and values as the CFR lays down both rights and values which have to be adhered to within the Union and thereby are relevant for the AIA and this thesis. This is followed by an explanation of GDPR provisions relevant for AI. The GDPR is of great importance, as AI relies on vast amounts of data and depending on the AI system's purpose, sensitive data of individuals. For the handling of such data, the GDPR provides multiple metrics that have to be adhered to when handling data. Moreover, the GDPR will be the focus of how the formulated AI concepts challenge policymakers when it comes to upholding the law. Section 4.2 will begin with a description of the relevant challenges identified in scientific literature, followed by an analysis of how these conflict with European fundamental rights, values and the GDPR. The findings of section 4.1 and 4.2 will serve as a basis to answer the first sub question of "What are the existing and relevant fundamental rights and internal market regulatory standards applicable in the adoption of the AIA and how are they challenged by AI?". Section 4.3 will introduce the EU's response to the dilemma presented in this thesis. For this, the sub questions "What is Trustworthy AI and how are these AI related challenges reflected in the guidelines for Trustworthy AI?" and "What are the currently proposed instruments of the AIA?" will be answered. This is done by first describing the requirements system providers have to adhere to for their systems to be deemed TAI, whereas this description will begin with the requirements that reflect the AI related challenges identified in scientific literature before describing the remaining requirements, followed by a summary of the AIA instruments. The TAI framework is of relevance as it not only serves as the

foundation for some of the AIA's mechanisms but is also part of the AIA's objectives and envisioned to contribute to the effort of maintaining law, rights and values. Finally section 4.4 will answer the last sub question of "Are these able to address the identified AI related problems and are they correctly placed?" by critically assessing the AIA and the TAI framework using literature discussing these. Moreover, in the analysis, it will be assessed whether conflicts identified in the AI-related literature are accounted for by the instruments in the AIA. By answering all sub questions, this thesis will provide an answer whether the dilemma EU policymakers are faced with would be solved by implementing the AIA proposal or whether amendments should be made. In order to answer these subquestions , a content analysis was conducted on the basis of the codes which have been formulated using the literature described in this section.

3.4 Data Analysis

For this, Atlas.ti was utilised, which ultimately allowed for keeping records of important sections, relevant quotes or information and aids in quickly refinding relevant sections. To underline the finding of this research, the Code Co-Occurrence feature was used to highlight the relationship of codes and their corresponding themes. Inspired by the theoretical framework, the relevant codes for the analysis are listed in the table below (examples for each code can be found in Appendix B):

Category	**Codes**
AI Challenges	AI Challenges, Algorithmic Transparency & Black Box Technology, Discriminatory & Biassed AI, Equality, Freedom, Human Dignity, Solidarity
Assessments (GDPR, HLEG Framework)	Transparency Provisions, Non-Discrimination & Bias Provisions
GDPR Assessment	GDPR AI-Provisions, GDPR Conflicts
Trustworthy AI	HLEG Requirements, TAI Conflicts
AIA Instruments	Risk Categorisation, Innovation, High Risk Obligations, Transparency Obligation Governance Codes of Conduct
AIA Assessment	AIA Shortcomings Ambiguities, Incomplete Scopes, Problematic Definitions Missing Mechanisms Faulty Instruments

For the analysis of how fundamental rights, values and the GDPR are challenged by AI, documents on fundamental rights and values challenges were coded regarding whether a challenge is attributable to problems identified in scientific literature. For this the codes ‘Algorithmic Transparency (Black-Box Technology)’ or ‘Discrimination & Biassed AI’ were utilised, with the aim of seeing how these challenges can actually contribute to violation of rights and values and how these might look like. Before categorising a challenge into either of those codes, the code ‘AI Challenges’ was used to code challenges for fundamental rights or values. If applicable, the ‘GDPR Conflicts’ was used to code any discussed challenges regarding ‘GDPR AI-Provision’, whereas the latter code was used to develop an understanding of what provisions of the GDPR directly address AI.
As per (Aizenberg & van den Hoven, 2020), specific rights violations constitute a violation of a foundational value (4.1.1), the codes ‘Human Dignity’, ’Freedom’, ’Equality’ and ’Solidarity’ were utilised to code sections that directly indicated a value violation or or when a corresponding right was violated.
For the analyses of the GDPR and the HLEG framework, the overarching codes ‘Transparency Provisions‘ and ‘Non-Discrimination & Bias Provisions’ were used to code provisions that prescribe insights into for instance how an algorithm works, how or why decision was made, or the data it uses and provisions that aim to prevent discriminatory, erroneous or biassed AI decisions. Together, these codes serve to highlight how the two main challenges discussed in this thesis are reflected in the GDPR and HLEG framework, whereas next to the GDPR codes, in the analyses of the two documents, the code ‘TAI conflicts‘ was used to code sections that entail information that pose a challenge to the HLEG framework’s provisions.
The codes used to gather information of the AIA Instruments are inspired by the titles of the AIA proposal (European Commission, 2021) and were used to code sections that discuss the AIA’s instruments and mechanisms, wherein of these coded sections, the codes were used to gather information how system are categorised based on their risk, what innovation mechanisms, transparency obligations and obligations for high risk systems are proposed, as well as how the AIA foresees the governance aspect of the AIA including the European Artificial Intelligence Board (EAIB) and its tasks, and finally what kind of codes of conduct are foreseen. For the AIA assessment, any challenge discussed regarding the AIA instruments and mechanisms were coded as ‘AIA Shortcomings’, whereas the shortcoming will then be categorised into one of the corresponding codes. ‘Faulty Instruments' refers to instruments or mechanisms that are unattainable or insufficient in safeguarding rights and values. Here, literature was coded using the code ‘Ambiguities’ regarding instruments not specifying important aspects, ‘Incomplete Scopes’ regarding prohibitions of certain systems that leave loopholes running the risk of violating rights and values. The code ‘Problematic Definitions’ was used to code literature discussing erroneous definitions of instruments and mechanisms, lastly, the code ‘Missing Mechanisms’ was used to code literature discussing specific expected mechanisms.

3.5 Preliminary Conclusion

By method of document analysis and utilising a coding scheme, recurring themes in European policy documents and scientific literature, potential challenges for European lawmakers when trying to regulate AI in the form of the AIA will be explored.

4 Analysis

This chapter will analyse various AI-related challenges for fundamental rights, GDPR provisions and foundational values, with a focus on the two overarching challenges presented in this thesis. This is followed by an assessment of the European approach to addressing the dilemma at hand, by investigating potential shortcomings that may hinder the achievement of its goals. Section 4.1 will first highlight the interplay of fundamental rights and values at stake in certain AI contexts, before continuing with the particularly relevant GDPR provisions thereby providing the relevant backdrop to point out how AI clashes with these in section 4.2. Together, these sections will answer the first subquestion. Section 4.3 will provide an introduction on the for this thesis relevant provisions of the HLEG's TAI framework to answer the second sub question, followed by an introduction on the AIA's instruments to answer the third subquestion. Finally, section 4.4 will provide an assessment of these instruments and the relevant TAI guidelines, in order to answer the fourth and final subquestion.

4.1 Laws, Rights & Values

4.1.1 Fundamental Rights and Values Relationship

Within the CFR, some provisions are especially relevant in AI contexts, meanwhile, some fundamental rights serve as a basis for the values of equality, freedom, human dignity and solidarity. Human dignity relies on CFR Article 1, human dignity and Article 2, the right to life. Freedom relies on Article 6, the right to liberty and security, Article 8, protection of personal data and lastly Article 11, freedom of expression and information. Equality is based on Article 21, non-discrimination, while solidarity is based on Article 34, social security and social assistance. Therefore a violation of a fundamental right may come at the detriment of freedom, equality, solidarity or human dignity, notably, a violation of any of the values also constitutes a violation of human dignity (Aizenberg & van den Hoven, 2020).

4.1.2 Relevant GDPR AI Provisions

In order to comply with the GDPR and for instance fulfil the TAI requirement of being lawful, systems in the EU have to adhere to numerous provisions. This section will focus on the provisions identified in the literature as particularly relevant to this discussion. In the analysis, out of the 21 instances the code 'GDPR AI-Provisions' was used to identify provisions relevant to AI, ten instances sections were also coded as representing 'Transparency Provisions' and three instances as 'Non-Discrimination/Bias Provision' (Appendix C, Graphic 1), underlining the prevalence of these challenges in literature on AI and the GDPR, especially in the case of algorithmic transparency.
For instance, it was found that the GDPR prohibits decisions solely based on automated decision making and profiling that has 'legal or similarly significant effects', with a legal effect constituting any processing activity that has an impact on someone's legal rights, whereas a significant effect entails that the system decision must have the potential to significantly influence their circumstances, behaviour or their choices. This prohibition however, has three exceptions. These decisions can be performed if either, (a) it is necessary for performance of or entering into a contract(Artzt & Dung, 2023), (b) is authorised by union or member state law to which the controller is subject and which also lays down suitable measures to safeguard the data subjects rights and freedom and legitimate interest, or (c) the data subject consents to it (Data Protection Working Party, 2018). The GDPR provides several rights that controllers of a system need to enable. Besides the rights already addressed, data subjects have the right to be informed[1] about being subjected to automated decision making (Article 13), receive meaningful information about the logic involved and need to be given an explanation of the significance and envisaged consequences of this processing (Data Protection Working Party, 2018). Relevant here is that as per principle of transparency, this information has to be

[1] Referred as the right of- or right to explanation in (Kesa & Kerikmäe, 2020) & (Artzt & Dung, 2023)

concise, easily accessible and easy to understand, using clear and plain language, regardless of a system's complexity.
Additionally this principle provides that data must be processed lawfully, fairly and transparently in relation to the data subject, whereas part of the GDPR's fairness requirement provides non-discrimination of individuals. Another legal requirement is the right to obtain human intervention, providing the right to challenge a system's decision (Artzt & Dung, 2023). When collecting personal data, the principle of purpose limitation entailed in Article 5 provides that data should only be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes" (Artzt & Dung, 2023, p.52), with further processing requiring consent of the data subject. Article 5 also entails the principle of data minimisation which provides that data not necessary to meet a system's purpose shall not be collected and the principle of data accuracy, providing that data controllers keep their data accurate and up to date (Artzt & Dung, 2023). Finally, the GDPR's Article 17 provides data subjects with the right to have data controllers erase their personal data 'without undue delay' (Kesa & Kerikmäe, 2020).

4.2 AI: Challenges

Having identified the relevant standards AI has to meet, the following section will analyse the AI-related concepts which have been identified and their respective conflicts which should be addressed when striving for a market-wide implementation of AI.

4.2.1 AI Challenges For European Rights and Values

In the document analysis the code for AI challenges for EU fundamental rights and values, the 'AI Challenges' code was utilised 31 times, on 15 instances these challenges were also coded as attributable to discriminatory and biassed AI, whereas of these eleven were attributable to (a lack of) algorithmic transparency, or the black box technology. Moreover, in regard to challenges to foundational values, 31 out of 31 challenges were coded as violating human dignity which is explained by the fact that any challenge to EU values constitutes a violation of human dignity (Aizenberg & van den Hoven, 2020). In regard to the remaining value codes, the 'Equality' code has been used 13 times, the Freedom code 12 times and the Solidarity code four times (Appendix C, Graphic 2).
Summarised, when systems are opaque, a challenge persists that "results are difficult, if not impossible to dispute or appeal. The harms that they can potentially perpetrate often have no remedy, and those who suffer these harms consequently lack recourse to address them" (von Eschenbach, 2021, p.1612). In relation to EU values, an individual's human dignity can be challenged by a confrontation with an opaque decision, resulting in humiliation or helplessness. Otherwise, human dignity is also challenged when AI leads to job losses and the affected workers become exchangeable, due to AI posing as an efficient alternative to human workers, or in the case of large-scale data collection for online advertisement, with individuals being reduced to data points and therefore means to garner profits for a company (Aizenberg & van den Hoven, 2020).
For the issue of discriminatory and biassed AI, the literature discusses several ways of how bias can infiltrate systems. Firstly, as humans are developing algorithms for AI systems, their biases can be introduced into data, or lead to biassed interpretations of AI system results (Gerrards, 2019). Moreover, these errors can remain undetected in testing phases until issues start materialising when a system is finally deployed (Artzt & Dung, 2023). (Training) data can also reflect historical and societal biases, notably through profiling by AI systems, existing stereotypes may be exacerbated (Data Protection Working Party, 2018). This can ultimately give rise to a new type of discrimination based on how individuals are represented in data (Aizenberg & van den Hoven, 2020). In regard to historical or societal biases, when AI is used to determine a person fit for a high-position job, historical data likely correlates success in a high position with male employees, ignoring factors such as historical discrimination and prejudices (Gerrards, 2019), ultimately resulting in unfair discrimination, violating Article 21 CFR, at the cost of equality.

Moreover, by systematically disadvantaging a (member of a) group over another (member), AI challenges the value of equality (Aizenberg & van den Hoven, 2020). AI discrimination also poses a challenge to the value of solidarity, as the increasing use of algorithms, gives rise to moral hazards by neglecting important factors when decisions are made by AI (von Eschenbach, 2021). Here, for the notion of solidarity, it is crucial that "individuals' ability to exercise their rights, and therefore uphold their dignity can be compromised" (p.8) in vulnerable circumstances (Aizenberg & van den Hoven, 2020). However, with AI's lack of a 'human factor' (Gerards, 2019, p. 206),"common sense" or "ethical override" (Artzt & Dung, 2023, p.42), factors such as historical discrimination are neglected, prioritising statistical correlations for AI decisions (Aizenberg & van den Hoven, 2020). The value of freedom, including the right to liberty and security (Article 6 CFR) protecting against the arbitrary deprivation of physical freedom, is challenged when for instance a person is falsely classified as high-risk by a system due to a statistical correlation such as the demographics of another arrested person, undermining the individual's ability to "enforce their autonomy" (Aizenberg & van den Hoven, 2020, p.7).

The analysis also found that the challenges of a lack of algorithmic transparency and discriminatory and biassed AI are also often discussed in relation to each other. The interplay of the challenges highlights that issue of AI biases can be exacerbated by lacking clarity into, firstly, what parameters lead to a decision being made by a system (von Eschenbach, 2021), secondly there being increased difficulty to comprehend 'the inside' of a system in general, making the system a black box (Gerards, 2019). The codes were utilised together on six instances (Appendix C, Graphic 4) in the context of rights and values. If used for decisions such as the job market or credit scores, the prevention of insight into how a decision was made (Rodrigues, 2020) can prevent the detection of violation of the equality value and Article 21 CFR. Moreover, if AI is deployed in a judicial context, for instance as a legal analytics tools the right to a fair trial (Article 47 CFR) challenged by a lack of algorithmic transparency preventing insight into whether a decision is justly made (Brkan et al., 2021), with an unjust decision coming at the cost of the freedom value, especially if a system is used to determine prison sentences (Gerards, 2019). A prominent example where such an issue was given is the Dutch court case surrounding the SyRi system, which was employed "to predict the likelihood of an individual committing benefit or tax fraud or violating labour laws" (Kesa & Kerikmäe, p.78), however the lack of transparency into the system's logic was part of the ruling to terminate its use (Greenstein, 2021). Additionally, by being deployed in poor neighbourhoods, the system ran the risk of disproportionately affecting vulnerable populations (Kesa & Kerikmäe, 2020) and therefore violating the equality value.

In the data analysis some challenges were identified that were not directly attributed to the two overarching challenges discussed in this thesis, whereas, AI can not only directly challenge values and rights, but also contribute to conflicts between them. In these cases AI serves as a means of violating rights and values. For example, the use of AI to track diseases or prevent crime by collecting and processing personal data, whereas the rights to privacy and data protection (Article 7 and 8 CFR) are at odds with national security (Sartor, 2020), for instance as part of Article 6 CFR. When AI is used to regulate speech and information, censoring specific information or political opinions would violate Article 11 of the CFR (Sartor, 2020), constituting a violation of the freedom value. A similar point is raised by (Brkan et al., 2021), whereas Article 39(2) of the CFR, the freedom of elections, is challenged by purposely deploying "partial information" (p.699) to specific voters. Gerrards also raises that the right to privacy is challenged by the growing amount of devices collecting data being rolled out, including the use of AI technology like facial recognition in public spaces. With the underlying presence of technological surveillance, challenging the value of freedom. Whereas this value can also be challenged when Article 11 CFR, the freedom of expression and information is undermined by the generation of filter bubbles with the result of providing individuals with biassed or manipulated information (Gerrards, 2019). Having underlined AI's capability of challenging fundamental rights and values, the next section will explore the legal context of this dilemma.

4.2.2 AI GDPR Compliance

Having introduced the relevant GDPR provisions and AI challenges, this section will analyse these challenges in light of the two main challenges at hand by use of the 'GDPR Conflicts' code. This code was utilised nine times in the analysis, of these, five were also coded as being attributable to the issue of algorithmic transparency (Appendix C, Graphic 3), with black box systems for instance, being at odds with the principle of transparency (Kesa & Kerikmäe, 2020). Moreover, its is raised whether the right to challenge a decision based solely on automated decision making can be realised or is even attainable when it comes to opaque systems, as highly complex systems make it hard to provide any information concerning the logic they apply at all (Kesa & Kerikmäe, 2020). Another transparency provision that is challenged is the right to be informed, whereas for providing the right to be informed about the logic behind a system's decision and the envisaged consequences, a major challenge is again the logic behind some systems being highly complex. Furthermore, in some instances it may even be "too difficult to explain in terms that people can understand" (Kesa & Kerikmäe, 2020, p.76), or "might not even be understandable in the first place" (Artzt & Dung, 2023, p.43). This problem was addressed by the Commission, whereas instead of a full explanation of an algorithm, a description about the data used in a decision making process and the main factors behind decision, the information source as well as its relevance should be provided. Nonetheless, a challenge prevails when dealing with systems lacking transparency and especially systems with high degrees of autonomy (Artzt & Dung, 2023).

Regarding challenges for non-discrimination provisions, which represent two of the ten coded sections (Appendix C, Graphic 3), one instance is also attributable to a lack of transparency. The ways of how AI can discriminate have been described in 4.2.1, for the GDPR AI discrimination would result in a breach of the fairness requirement. Notably, the issue of non-discrimination can be exacerbated when for instance black box systems increase the possibility that discriminatory decisions remain undetected (Artzt & Dung, 2023).

The remaining challenges for AI to be compliant with the GDPR are attributed to data principles, for one, by continuous learning and exploitation of data by AI systems, the right to erasure is challenged by the fact that even after deleting data from a system, the system has learned from that data subject's data would have to be removed somehow, which at least is very burdensome, if not impossible (Kesa & Kerikmäe, 2020). Secondly, the data minimisation principle is challenged by the very "way an AI system works" (Artzt & Dung, 2023, p.53), as the principle requires a limit on the amount and time of data being processed, however AI relies on vast amounts of data (big data) (Artzt & Dung, 2023). Lastly, the principle of purpose limitation is challenged as an AI system's purpose is difficult to determine a system's purpose in early development stages (Artzt & Dung, 2023). Concluding the GDPR discussion, the compliance of AI with the GDPR can be severely challenged by a lack of algorithmic transparency, which also attributes to challenges related to non-discrimination and biases.

4.3 European Response to Governing AI: AIA

4.3.1 Challenges Reflected in Trustworthy AI Framework (HLEG)

The data collection found that research on the HLEG TAI framework is highly limited[2], for this reason the HLEG guidelines for TAI will serve as the main focus of this analysis and the assessment in 4.4.4. This section will analyse the HLEG TAI framework with a focus on transparency and non-discrimination provisions. To be deemed trustworthy and meet the three components of being lawful, ethical and robust, the HLEG provides seven requirements that should be met, with these requirements being based on the ethical pillars of respect for human autonomy, the prevention of harm, fairness and explicability (Hickman & Petrin, 2021). Each of the the seven requirements, human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being and finally, accountability, entail provision whereas each provision was coded using the 'HLEG Requirements' code, which was utilised 24 times, whereas of these six transparency-, and five non-discrimination or bias provisions were identified and coded (Appendix C, Graphic 5). Of these, primarily the requirement of transparency, the requirement of diversity, non-discrimination and fairness, furthermore the quality and integrity of data section of the privacy and data governance requirement and lastly of the fundamental rights and human agency sections of the human agency and oversight requirement contain provisions addressing the identified main challenges.

As for human agency and oversight, similar to the GDPR, users of AI systems should be provided with knowledge and tools to understand a system's decision-making "to a satisfactory degree" (HLEG, 2019, p.16). Moreover, oversight for a public enforcer must be enabled and systems with little possible oversight are to be governed stricter. Additionally, as the first step in a system's production cycle, when risks to fundamental rights are likely, an assessment of whether these are reducible should be conducted, as well as creating feedback mechanisms for infringement on fundamental rights. For the transparency requirement, data sets, processes of data gathering-, labelling, any algorithms that are used and decisions made by AI systems should be well documented in order to allow for the identification of erroneous decisions. When problems arise due to black box technology, the framework foresees that AI system processes and their relations to human decisions need to be explainable, also when using AI systems, users should be informed that they are communicating with an AI system and not a real human being, furthermore, systems should not portray themselves as human, instead, they have to be identifiable as an AI system. The systems' capabilities and limitations should also be communicated, and the opportunity to interact with a human instead of an AI system, if required for compliance with fundamental rights, should be provided.

Regarding diversity, non-discrimination and fairness, discriminatory biases are supposed to be minimised or removed to the greatest extent possible in the data collection process; which is to be tackled using oversight processes to analyse data to identify any problems. Furthermore, AI services have to be accessible to everybody regardless of their background. Also for the development of AI systems, it is advised to consult with affected stakeholders and incorporate feedback. Particularly relevant for this research and the eventual assessment in chapter four, user data shall also not be used unlawfully or unfairly to discriminate against them. Discriminatory biases, inaccuracies or errors are to be averted before training data, while for data integrity, the testing of data should be accompanied by documentation of each step in the development of a system, in order to prevent the feeding of malicious data into the system. Finally concluding the provision relevant for the assessment in 4.4.4, data protocols regarding who can access data under which circumstances should be created, with only competent persons that must access an individual's data being allowed to do so.

The remaining 13 identified provisions mostly concern assessing the impact a system has, they also include, a possibility for human intervention in every decision cycle (the framework notes that this is rarely possible), human intervention in the design and monitoring phase, as well as during the

[2] Not only is the amount of research limited, but of the existing papers, many refer to an older draft version of the framework, not the version referenced in the AIA.

system's operation and oversight over the resulting economic, societal, legal and ethical impact to finally decide whether the use of said system is adequate in a given situation (part of human agency and oversight requirement). For technical robustness and safety, systems should be protected against cyberattacks, to prevent altering or stealing of data and preventing system corruption. Providers need to take steps to prevent unintended applications or abuse of their systems, which can be achieved by measures such as, an AI system stopping operation until a human intervenes and a general awareness of potential risks by providers. For the requirement of privacy and data governance, privacy and data protection of the information users provide, the data that is generated for users and their responses should be met throughout a system's life cycle. In order to prevent harm and increase fairness, social and environmental well-being have to be aimed for. All processes regarding the AI system and its use should be critically examined regarding its resource use, the HLEG also encourages the use of environmentally friendly techniques. Moreover, the social impact of a system shall be monitored to assess and prevent negative impacts, this also applies to influences on societal and democratic processes. Lastly, for accountability, actions and decisions of AI systems, as well as any concerns regarding them must be reported on, which should be further enabled by the protection of whistle-blowers, NGOs and other institutions that report on negative impacts. Additional impact assessments before and during the use of AI, proportional to the risk a system poses should be conducted. Finally, during their implementation, some of the seven requirements may clash, in these cases, the trade-offs should be properly addressed. The development or implementation of an AI system should not take form if there are no "ethically acceptable trade-offs" (HLEG, 2019, p.20), decisions regarding this shall be documented with the person making decisions being held accountable. For cases of adverse effects, redress mechanisms should be established (HLEG, 2019). Having identified the provisions required for TAI, the next section will present how the EU adopted these in the AIA.

4.3.2 AIA: Instruments

Before assessing the instruments proposed in the AIA against the backdrop of the dilemma presented in this thesis, an overview of these instruments is provided. In an attempt to achieve its four objectives, the AIA relies on a number of instruments and a risk-based approach where the level of regulation of a system is proportional to its level of risk. The main instruments and mechanisms are, a classification system categorising systems based on their risk, the establishment of the European Artificial Intelligence Board with subordinate national authorities, a database for high-risk AI systems to verify whether these systems adhere to the regulations (Article 60), a regulatory sandbox where new AI systems can be tested before they are placed on the market with the aim of boosting innovation (Article 53) and a conformity assessment mechanism. In case of not complying with the AIA, the involved parties are also set to receive expensive fines (European Commission, 2021). The overarching instrument of the AIA is the classification mechanism which determines the level of regulation an AI system is subjected to proportional to its assigned risk level. The proposal classifies systems as either having low or minimal risk, high risk or unacceptable risk (European Commission, 2021). Systems are deemed to be of unacceptable risk if they pose a "clear threat to the safety, livelihoods and rights of people and includes systems that manipulate human behaviour or allow 'social scoring' by governments" (McFadden et al., 2021, p.7). Article 6 of the AIA provides an extensive yet loose definition for high-risk AI systems, nonetheless, they can be summarised as systems that pose a significant risk to health, safety and fundamental rights (Townsend, 2021). Systems fitting neither of the definitions above are assumed to be either of low or minimal risk (as the AIA does not provide a definition for this category) and encouraged to follow voluntary codes of conduct with the AIA not adding additional obligations for these systems, with some exceptions (Townsend, 2021). These exceptions are for instance, systems producing deep fakes, emotion recognition and biometric categorisation systems and have to adhere to transparency obligations. For instance, systems need to be designed so that they inform people they are communicating with an AI system, emotion recognition and biometric categorisation need to disclose what they are doing, whereas deep fakes must be branded as deep fakes (Varošanec, 2022). The most regulation falls upon

high-risk systems, the AIA's second chapter, ranging from Article 9 to 15 provides the requirements these systems have to adhere to (European Commission, 2021). Summarised, heavily inspired by the HLEG framework, for high risk systems, providers have to establish a risk management system including data governance for training, validation and testing data, including traceability by means of logging, ensuring transparency for users regarding system characteristics and imitations, human oversight during system use and designing systems in a way that accuracy, robustness and cybersecurity is always ensured (Schaake, 2021).
For the governance of the AIA, the AIA proposes that the EAIB will enforce the act and is composed of representatives from member states and the Commission. This board takes up advisory tasks and issues opinions and recommendations when it comes to the implementation of the AIA, including technical existing standards regarding requirements established in the AIA. Furthermore, it provides advice and assists the commission of specific AI related questions (European Commission, 2021). In regard to proposed innovation mechanisms, the regulatory sandbox is envisioned to foster innovation by providing a controlled environment for the development and testing of new AI systems (European Commission, 2021) and can be compared to clinical trials for pharmaceutical products (Raposo, 2022). These sandboxes would have to be set up by the proposed national competent authorities. Lastly, the EU database for stand-alone high-risk AI systems will be accessible to the public and is envisioned to allow verification and oversight of whether high-risk systems follow the AIA requirements. Moreover, providers are required to provide information about their AI system and its conformity assessment and in the case of malfunctions, inform the national authority (European Commission, 2021).

4.4 AIA: Assessment

Despite the AIA focusing on combating risks to rights, laws and values there are numerous issues raised in scientific discussion that are either improperly or not addressed at all. This includes ambiguities, loopholes, incomplete instrument scopes, shortcomings attributed to, for example, a problematic definition of "AI systems" and entirely missing mechanisms. This analysis will focus on the previously mentioned aspects, beginning with the analysis of faulty instruments proposed in the AIA.

4.4.1 AIA: Shortcomings

In the analysis of AIA related documents the code 'AIA Shortcomings' was utilised 42 times, whereas the most coded (17) shortcoming was attributed to faulty instruments (Appendix C, Graphic 6). Of these, the most striking deficiency raised is the self-assessment procedure as part of the transparency obligations for high-risk systems, which is raised in multiple papers. While, as mentioned in the previous section, a limited number of high-risk systems need to cooperate with market surveillance authorities for this assessment, most system providers can conduct this assessment themselves (Raposo, 2022). Ideally an assessment of whether systems adhere to standards before entering the market would prevent risk-laden systems to cause damages and violate laws, rights and values, by allowing providers to independently conduct the assessment however, the possibility that companies will bypass obligations is plausible (Varosanec, 2022) as an independent assessment allows providers to determine whether their system is in fact an AI system at all or is likely to cause harm (Ebers et al., 2021). Against the backdrop of the two overarching challenges investigated in this thesis, not only faulty, but likely impossible to attain is the requirement as part of the instruments regulating high risk systems, whereas as per Article 10, high risk systems are required to use error-free data sets for testing, however error-free data sets are argued to be impossible to attain (Varosanec, 2022). Furthermore, Article 14 foresees that humans can fully understand a high-risk system's capacities and limitations, however (Ebers et al., 2021) stresses the impossibility of all systems meeting this requirement. Lastly, in the case of for instance biometric systems undergoing assessments by notified bodies, if only conducted by private business ideas, it may be the case that

these bodies are merely checking for the formal requirements of the AIA, instead of trying to actively protect fundamental rights.
Besides some AIA instruments evidently being deficient, the proposal also entirely lacks some expected instruments and mechanisms. This shortcoming was coded ten times, being the second most coded challenge (Appendix C, Graphic 6) with a lack of treatments for liabilities in case of damages (Raposo, 2022) at the forefront. Similarly, a right that enables individuals affected by AI systems to issue complaints to market surveillance authorities or sue a provider under the AIA is missing from the proposal (Ebers et al., 2021). Against the backdrop of AIA's second objective aiming to ensure innovation, the proposal lacks mechanisms for this objective, for one, the AIA only contains two mechanisms of this manner and more importantly, it misses a mechanism for the development of AI systems in research, lifting or adjusting restrictions for academic settings. Also, in regard to the risk categorisation, a mechanism is missing for possible cases of false categorisation and an eventual correction (Raposo, 2020). Moreover, the entire proposal fails to address general purpose systems, which are systems that can be deployed in a variety of contexts (Ruschemeier, 2023). Finally, while the AIA provides the Commission with the ability to amend the list of high risk systems, it does not include a mechanism to amend the list of prohibited systems, which may lead to problems when a system's problems are detected after a period of time, as well as misses a mechanism preventing systems from undergoing two assessments (Ebers et al., 2021).
Coded on eight instances, were shortcomings attributed to ambiguities (Appendix C, Graphic 6), with the AIA providing partly unspecified instruments leaving risk for rights and value violations and undermining legal certainty. For instance, developers are instructed to make systems transparent to a degree that enables users to interpret their output and use it appropriately, which requires concise, complete and correct information, however, neither is it made clear what info this might be, or who decides whether a user is able to interpret and use an output appropriately (Varosanec, 2022). Another aspect of ambiguity results from the AIA not addressing how member states could deviate from its requirements, such as extending the list of prohibited systems (Ebers et al., 2021). More ambiguity results concerning the risk assessment of systems, where the AIA fails to first, specifically lay out who is responsible for this assessment and at what point in time it is to be conducted. A problem resulting from this is that a system's threats may not be detected at the time of an early assessment (Raposo, 2022). Similarly questionable aspects are raised when it comes to social scoring and remote biometric identification systems. Regarding social scoring, for a prohibition systems have to operate "over a certain period of time" (European Commission, 2021, p.43) and be operated by a public body, yet the AIA does not specify how long this period of time is. Another issue is raised when it comes to the or a manipulative system to be deemed as in-fact manipulative for instance, manipulative intention by the developer or publisher is required, however how such an intention may be proven is not addressed, furthermore it being unlikely that a developer or publisher would admit manipulative intent (Raposo, 2022).
The fourth most coded AIA shortcoming is attributed to problematic definitions, which was coded seven times (Appendix C, Graphic 6), whereas for the most raised issue in this regard is that, AI systems are defined as "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with", (European Commission, 2021, p.39), with Annex I listing "(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods." (Ruschemeier, 2021, p.367). As a result of such a broad definition, "almost every computer programme" (Ruschemeier, 2023, p.368) is addressed by the AIA, regulating systems outside of its scope (non-AI systems) by for example applying requirements for high-risk systems to non-AI systems (Ruschemeier, 2023). This, accompanied by not specifying differences between AI and AI systems (Ebers et al., 2021), can diminish the legal certainty and confidence in AI the EU envisions to achieve with the AIA.

Not only running the risk of over regulation in some instances, other areas are arguably under-regulated such as biometric systems, whereas these shortcomings were coded on six instances and as attributable to incomplete scopes (Appendix C, Graphic 6). This includes AI developed exclusively for military purposes being exempt from the AIA's regulations. An issue here is that what constitutes exclusiveness is not elaborated (again an issue of ambiguity), furthermore a specific system can be developed or used for military purposes but also be used in a civil setting (Ruschemeier, 2023). The issue of a scope of exemptions being a cause of concern is also relevant for facial recognition technology, as the proposal permits real-time facial recognition technology in public spaces under specific circumstances. However, provides relatively vague parameters and includes situations like fraud, corruption, unauthorised entry and residency. The AIA states that a judicial authority needs to permit the use of facial recognition technology in public spaces by law enforcement, however, there is an exception where this judicial decision can be postponed in case of urgency. This exception allows the potential abuse of facial recognition technology in the given context is only noticed after a violation. As a result of the issues concerning social scoring systems in regard to ambiguity issues, periodic scoring or scoring systems by private entities are exempt from a ban. Similarly, when it comes to remote biometric identification systems, if not carrying out identification in real time, public spaces or is used for identity confirmation instead of identifying individuals, or used by private bodies, systems can circumvent the prohibition (Raposo, 2022). Lastly, among the systems listed posing significant risks that are used by law enforcement authorities, an issue raised is, the AIA listing systems merely focussing on individuals, leaving out those aimed at groups including predictive policing, opening the risk for over-policing and systemic discrimination (Ebers et al., 2021) ultimately undermining fundamental rights and values as described in section 4.2.1.

4.4.2 Trustworthy AI Assessment

Despite the lack of research assessing the HLEG framework as stated in 4.3.1, the findings of the already analysed literature serve as feasible means to create a rough assessment of potential challenges for AI system providers who attempt to meet the HLEG requirements, as well as policy makers when translating these into instruments as in the case of the AIA. This feasibility is underlined by the finding that the 'TAI Conflicts' code was utilised 22 times (Appendix D, Graphic 1).
"Trust requires that we have reason to believe both that AI is reliable and acting in our behalf" (p.1620), from which emerges that transparency into how an AI system works and how an outcome was made are required to deem a system as trustworthy (von Eschenbach, 2021). However the previous sections made clear that this notion is hard to come by for AI systems, especially entirely opaque systems. To meet the requirement of human agency and oversight, users should be able to understand a system's decision making "to a satisfying degree" (HLEG, 2019, p.16) however as pointed out by for example (Artzt & Dung, 2023) and (Kesa & Kerikmäe, 2020) in section 4.2.2 in regard to the GDPR's right to be informed, such transparency can not be provided for every system, furthermore it is questionable how meaningful the proposed alternative transparency measures for black box systems can be, especially considering that of these measures, the communication provision which was adopted into the AIA Article 14, foresees the communication of a systems limitations and capabilities, whereas the impossibility in some cases of this was already addressed in the previous section, however more research into these alternatives may yield answers. Also, it is left ambiguous what constitutes a satisfactory degree, as well as how public enforcers can deal with opaque systems. This issue also challenges the requirements of Transparency and Privacy & Data Governance, whereas for one, system processes should be documented to allow for the detection of erroneous decisions and more relevantly, to combat the effects of discriminatory AI, the HLEG framework requires that "socially constructed biases, inaccuracies, errors and mistakes (...)" have "to be addressed prior to training with any given data set." (HLEG, 2019, p.17), including the prevention of unlawful discrimination by AI. (HLEG, 2019). Not only is it unclear how these aspects can be properly addressed, but in the case that it is foreseen that data is for instance free of errors, as in the case of Article 10 of the AIA, as (Varošanec, 2022) raised that error free data sets are impossible to

attain or as (Raposo, 2022) puts it, “utopian” (p.108). Therefore, it seems improbable that discriminatory decisions can completely be averted and therefore improbable to meet the requirement Finally, the HLEG misses guidance on the lawful component of TAI, the AIA assessment underlined how several ambiguities and conflicts may undermine the goal of legal certainty, whereas further guidance on how to meet this criteria seems adequate. Concluding the TAI assessment, stressing that if it is unclear how a decision was reached, systems can not be deemed trustworthy (von Eschenbach, 2021), the question arises how trustworthy AI can really be. Considering that some systems, at least for the time being, are incomprehensible and lack transparency, from which results a risk of not detecting and counteracting issues such as discriminatory and biassed AI and ultimately raising doubts whether the framework sufficiently serves its aim to safeguard law, rights and values.

4.4.3 Concluding Analysis

Having analysed how AI challenges fundamental rights, values, law as well as European policy makers in the context of the AIA, the final section of this thesis will formulate answers to the research questions and draw a final conclusion to this thesis.

5 Conclusion

5.1 Answering The Research Question

The purpose of this research was to answer the following research question: “How does Artificial Intelligence challenge the European Union’s efforts of maintaining its fundamental rights, values and EU law while developing a market for trustworthy artificial intelligence?”. The document analysis guided by use of a coding scheme found that especially a lack of algorithmic transparency and AI discrimination pose challenges to fundamental rights, values and the GDPR, whereas the challenge of lacking algorithmic transparency can exacerbate the issue of algorithmic discrimination underlined by the concurrent use of their respective codes on six instances.
By answering the first subquestion, it was shown that there are a multitude of fundamental rights, values and GDPR provisions at odds with AI, which are challenged by systems with impaired or completely lacking algorithmic transparency, as well as discriminatory and biassed systems. Furthermore, it was shown that rights violations come at the cost of also going against the values of equality, freedom, human dignity and solidarity.
By answering the second sub question, it was shown that one of the first major European attempts to develop a framework with the aim of safeguarding rights, law and values reflects the identified challenges and aims to address these.
Furthermore, by answering the third subquestion was shown how the EU approaches this dilemma in the form of the AIA.
Finally, by answering the fourth subquestion it was shown that firstly the AIA proposal does not sufficiently address these challenges and in itself contains several different flaws, which are reflected in the different codes used in the AIA assessment. Ultimately, for this reason it was shown that the analysed AIA proposal requires further amendments to safeguard fundamental rights, values and European law. Also, it was found that while the HLEG framework seems like a promising approach to address AI related challenges, it is nonetheless likely that the framework contains unattainable provisions such as when aiming to provide transparency into systems, raising doubts of the limited an attainability of the framework as provided by the HLEG.
Similarly, this thesis was also limited in various ways. For instance, a strict time and word- limit allowed for a limited exploration of the issues of algorithmic transparency and discriminatory & biassed , as well as keeping the legal challenge analysis to only the GDPR. Here, for instance liability related laws and how AI applications such as autonomous vehicles or weapons challenge these could also have been explored. Furthermore, not all conflicts and shortcomings identified in the research process were able to be highlighted in this thesis and more generally as this thesis aimed to highlight various fields that are challenged by AI, only more general overviews of challenges for each field (rights, laws and values) could have been provided. Also, for this reason the analysis of transparency and non-discrimination provisions was restricted to the HLEG framework and the GDPR, excluding the AIA, which is partly mitigated by many of the HLEG provisions being adopted into the AIA. Finally, as this thesis was being written while the AIA was still being amendment and a final version has yet to be agreed on, some of the findings of the AIA assessment might be outdated if amendments such as the ones that will be highlighted in the final section (5.3) will be adopted into the final version of the AIA.

5.2 Knowledge Gap

This analysis found that algorithmic transparency as well as discriminatory and biassed AI are recurring themes in both the literature and for instance the GDPR and HLEG framework. Moreover, by applying the findings of literature examining AI challenges, the AIA and GDPR to the HLEG framework, a knowledge gap on research assessing the TAI framework was narrowed. Moreover, by merging the findings of several different papers and focuses, a grander picture of AI’s capability to challenge these was provided, therefore contributing to the discussion of highlighting the potential of AI to violate rights, values and law.

5.3 Practical Implications

As of 14th of June 2023, all involved bodies published their position on the AIA with the European Parliament having finalised its negotiating position for the AIA. This version of the AIA renews its definition for AI systems and adopts the definition provided by the Organisation for Economic Cooperation and Development. Furthermore, this position addresses some issues raised in literature such as not entirely banning biometric identification systems and exempting non-real-time use systems from a ban. These systems would also be prohibited in this version of the AIA.
Additionally, the Parliament's position addresses the raised concerns of biometric categorisation systems, prohibiting systems which use sensitive data including predictive policing systems and emotion recognition systems. It also specifies that for systems to be deemed high risk, they have to pose a significant risk, whereas systems influencing voters on social media to the high risk category. The position also adopts more provision from the HLEG framework, whereas a mandatory fundamental rights impact assessment for high risk systems is added. Notably, general purpose systems are finally added in the AIA, whereas it imposes obligations on these such as ensuring protection of fundamental rights. In regard to the governance aspect, the national authorities will also be provided with the ability to request access to training models of AI systems and the creation of a new body, the AI Office, is foreseen. This office is meant to conduct investigations across borders with the aim of harmonising the application of the AIA, including the possibility for citizens to file complaints and receive explanations of a system's decision. Lastly, an exception for research settings and open source systems is added (Madiega, 2023).
Notably, as this described version of the AIA is not a final version of the act, it remains unclear whether these amendments will be adopted. Moreover, while this thesis and the research it relied on can be used to guide the process of adjusting the proposal to encompass unaddressed risks to law, fundamental rights and values, due to the limitations of this thesis, future research should investigate further AI challenges relevant to this discussion. Also, an analysis similar to this thesis should be repeated with a focus on the position by the European Parliament and any new versions allowing to draw attention to possible shortcomings. Finally, research into how to open the black box and allow for required algorithmic transparency, including assessing whether the TAI framework helps in doing so or is attainable for AI system providers should be conducted in order to address challenges such as the two overarching challenges presented in this thesis.

References

Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. In *Big Data & Society* (2nd ed., Vol. 7, pp. 1-14). https://journals.sagepub.com/doi/full/10.1177/2053951720949566

Artzt, M., & Dung, T. V. (2023). Artificial Intelligence and Data Protection: How to Reconcile Both Areas from the European Law Perspective. In *Vietnamese Journal of Legal Sciences* (2nd ed., Vol. 7, pp. 39-58). sciendo. https://sciendo.com/article/10.2478/vjls-2022-0007?tab=pdf-preview

Belenguer, L. (2022). AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. In *AI and Ethics* (Vol. 2, pp. 771-787). Springer. https://link.springer.com/article/10.1007/s43681-022-00138-8

Brkan, M., Claes, M., & Rauchegger, C. (2021). European fundamental rights and digitalization. In *Maastricht Journal of European and Comparative Law* (6th ed., Vol. 27, pp. 697–704). https://journals.sagepub.com/doi/full/10.1177/1023263X20983778

Dafoe, A. (2018). *AI Governance: A Research Agenda*. https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf

Data Protection Working Party. (2018, August 22). *ARTICLE29 - Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01)*. European Commission. https://ec.europa.eu/newsroom/article29/items/612053

Ebers, M., Hoch, V. R. S., Rosenkranz, F., Ruschemeier, H., & Steinrötter, B. (2021). The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). In *Multidisciplinary Scientific Journal* (4th ed., Vol. 4, pp. 589–603). MDPI. https://www.mdpi.com/2571-8800/4/4/43

European Commission. (2021, April 21). 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. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

European Parliament. (2019, January 30). REPORT on a comprehensive European industrial policy on artificial intelligence and robotics. https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_EN.html

European Union. (2012, October 26). Charter of Fundamental Rights of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT

Gerards, J. (2019). The fundamental rights challenges of algorithms. In *Netherlands Quarterly of Human Rights* (3rd ed., Vol. 37, pp. 205-209). Sage. https://journals.sagepub.com/doi/full/10.1177/0924051919861773

Greenstein, S. (2021). Preserving the rule of law in the era of artificial intelligence (AI). In *Artificial Intelligence and Law* (Vol. 30, pp. 291-323). Springer. https://link.springer.com/article/10.1007/s10506-021-09294-4

Harrison, M. (2023, March 21). Researchers Reveal That Paper About Academic Cheating Was Generated Using ChatGPT. *Futurism*. https://futurism.com/the-byte/paper-passed-peer-review-generated-chatgpt

Hickman, E., & Petrin, M. (2021). Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective. In *European Business Organization Law Review volume* (Vol. 22, pp. 593-625). Springer. https://link.springer.com/article/10.1007/s40804-021-00224-0

HLEG. (2019, April 8). Ethics Guidelines for Trustworthy AI. *Independent High-Level Expert Group On Artificial Intelligence, set up by the European Commission*.

Kesa, A., & Kerikmäe, T. (2020). Artificial Intelligence and the GDPR: Inevitable Nemeses? In *TalTech Journal of European Studies* (3rd ed., Vol. 10, pp. 68-90). sciendo. https://sciendo.com/article/10.1515/bjes-2020-0022

Lorè, F., Basile, P., Appice, A., Gemmis, M. d., Malerba, D., & Semeraro, G. (2023). An AI framework to support decisions on GDPR compliance. In *Journal of Intelligent Information Systems*. Springer. https://doi.org/10.1007/s10844-023-00782-4

Madiega, A. (2023, June 7). *Parliament's negotiating position on the artificial intelligence act | Think Tank*. European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2023)747926

McFadden, M., Jones, K., Taylor, E., & Osborn, G. (2021). Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation. In *Oxford Information Labs*. Oxford Internet Institute. https://oxil.uk/publications/2021-12-02-oxford-internet-institute-oxil-harmonising-ai/

Oltermann, P. (2022, June 25). European politicians duped into deepfake video calls with mayor of Kyiv. *The Guardian*. https://www.theguardian.com/world/2022/jun/25/european-leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko

Procter, R., Tolmie, P., & Rouncefield, M. (2023). Holding AI to Account: Challenges for the Delivery of Trustworthy AI in Healthcare. In *ACM Trans. Comput.-Hum. Interact* (2nd ed., Vol. 30, p. 34). dl.acm.org/doi/abs/10.1145/3577009

Raposo, V. L. (2022). Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence. In *International Journal of Law and Information Technology* (1st ed., Vol. 30, pp. 88-109). Oxford. https://doi.org/10.1093/ijlit/eaac007

Ruschemeier, H. (2023). AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal. In *ERA Forum : Journal of the Academy of European Law* (Vol. 23, pp. 361-376). Springer. https://link.springer.com/article/10.1007/s12027-022-00725-6

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. In *Science and Engineering Ethics* (Vol. 26, pp. 2749–2767). Springer. https://link.springer.com/article/10.1007/s11948-020-00228-y

Sartor, G. (2020). Artificial intelligence and human rights: Between law and ethics. In *Maastricht Journal of European and Comparative Law* (6th ed., Vol. 27, pp. 705-719). Sage. https://journals.sagepub.com/doi/abs/10.1177/1023263X20981566

Schaake, M. (2021, June). *The European Commission's Artificial Intelligence Act*. Stanford HAI. https://hai.stanford.edu/sites/default/files/2021-06/HAI_Issue-Brief_The-European-Commissions-Artificial-Intelligence-Act.pdf

Schippers, B. (2020). Artificial Intelligence and Democratic Politics. In *Political Insight* (1st ed., Vol. 11, pp. 32-35). SAGE Publications. https://journals.sagepub.com/doi/full/10.1177/2041905820911746

Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. In *Computer Law & Security View* (Vol. 45). Elsevier. https://www.sciencedirect.com/science/article/pii/S0267364922000097

Townsend, B. (2021, September 30). Decoding the Proposed European Union Artificial Intelligence Act | ASIL. *American Society of International Law*. https://www.asil.org/insights/volume/25/issue/20

Varošanec, I. (2022). On the path to the future: mapping the notion of transparency in the EU regulatory framework for AI. In *International Review of Law, Computers & Technology* (2nd ed., Vol. 36, pp. 95-117). Routledge. https://www.tandfonline.com/doi/full/10.1080/13600869.2022.2060471

The Verge. (2017, September 4). *Putin says the nation that leads in AI 'will be the ruler of the world'*. The Verge. Retrieved March 14, 2023, from https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world

von Eschenbach, W. J. (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. In *Philosophy & Technology* (Vol. 34, pp. 1607–1622). https://link.springer.com/article/10.1007/s13347-021-00477-0

Data Appendix

Appendix A, Analysed and Consulted Literature

Author(s)	Title	Year of Publication
Aizenberg & van den Hoven	Designing for human rights in AI	2020
Artzt & Dung	Artificial Intelligence and Data Protection: How To Reconcile Both Areas from the European Law Perspective	2023
Belenguer	AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry	2022
Brkan, Claes, Rauchnegger	European fundamental rights and digitalization	2021
Dafoe	AI Governance: A Research Agenda	2018
Data Protection Working Party	Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01).	2018
Ebers, Hoch, Rosenkranz, Ruschemeier, Steinrötter	The European Commission's Proposal for an Artificial Intelligence Act -- A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)	2021
European Commission	Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence	2021

	(Artificial Intelligence Act) and amending certain union legislative acts.	
European Parliament	Report on a comprehensive European industrial policy on artificial intelligence and robotics	2019
European Union	Charter of Fundamental Rights of the European Union	2012
Gerrards	The fundamental rights challenges of algorithms	2019
Greenstein	Preserving the rule of law in the era of artificial intelligence (AI)	2021
Harrisson	Researchers Reveal That Paper About Academic Cheating Was Generated Using ChatGPT (Newspaper Article)	2023
Hickman & Petrin	Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective.	2021
HLEG	Ethics Guidelines for Trustworthy AI	2019
Kesa & Kerikmäe	Artificial Intelligence and the GDPR: Inevitable Nemeses?	2020
Lorè, Basile, Appice, Gemmis, Malerba, Semeraro	An AI framework to support decisions on GDPR compliance.	2023
McFadden, Taylor, Osborn	Harmonising Artificial Intelligence: The role of standards in the EU AI Regulation	2021

Madiega	Parliament's negotiating position on the artificial intelligence act	2023
Oltermann	European politicians duped into deepfake video calls with mayor of Kyiv	2022
Procter, Tolmie, Rouncefield	Holding AI to Account: Challenges for the Delivery of Trustworthy AI in Healthcare	2023
Raposo	Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence	2022
Ruschemeier	AI as a challenge for legal regulation - the scope of application of the artificial intelligence act proposal	2023
Ryan	In AI We Trust: Ethics, Artificial Intelligence, and Reliability	2020
Sartor	Artificial intelligence and human rights: Between law and ethics	2020
Schaake	The European Commission’s Artificial Intelligence Act	2021
Schippers	Artificial Intelligence and Democratic Politics	202
Stahl, Rodrigues, Santiago, Macnish	A European Agency for Artificial Intelligence: protecting fundamental rights and ethical values	2022
Townsend	Decoding the Proposed European Union Artificial Intelligence Act	2021
Varošanec	On the path to the future: mapping the notion of	2022

	transparency in the EU regulatory framework for AI	
The Verge	Putin says the nation that leads in AI ‘will be ruler of the world’ (Newspaper article)	2017
Von Eschenbach	Transparency and the Black Box Problem: Why We De Not Trust AI	2021

Appendix B, Coding Scheme (Including Examples)

Category	**Codes**	**Examples**
AI Challenges	AI Challenges (1) Algorithmic Transparency (Black-Box Technology) (2) Discriminatory and Biassed AI (3) Human Dignity (4) Freedom (5) Equality (6) Solidarity	(1): *"Given the accountability duties under the GDPR, the user of AI will not only need to ensure the machine's decision-making process is fair but also to demonstrate this is the case. This is likely to be challenging where that decision is taken in a "black box", though the use of counterfactuals and other measures may help." (Artzt & Dung, 2023, p.42).* (2): *"Third, algorithms may have discriminatory effects. Surely the right to non-discrimination does not require that everyone is treated the same. Indeed, in many cases this would create unfair situations, because all human beings are different. In fact, it would be best to treat everyone in accordance with their own, uniquely personal characteristics, merits, needs and behaviour" (Gerards, 2019, p.207).* (3): *"Therefore, individuals' ability to enforce their autonomy significantly relies on their awareness of being subjected to algorithmic profiling and their ability to contest the rationale behind algorithmic decisions"* *(Aizenberg & van den Hoven, 2020, p.7)* (4): *"Second, there may be an impact on rights such as the freedom of expression and access to*

Category	**Codes**	**Examples**
		information. The exercise of these rights is greatly facilitated by the availability of search engines, social media and internet forums. At the same time, our access to the wealth of available information is strongly determined by algorithmic analyses of our viewing, reading and clicking behaviour." (Gerards, 2019, p.206). (5): *"AI bias, for example, can be introduced to algorithms as a reflection of conscious or unconscious prejudices on the part of the developers, or they can creep in through undetected errors. In any case, the results of a biased algorithm will be skewed, potentially in a way that is offensive to people who are affected. Bias in an algorithm may come from input data when details about the dataset are unrecognized."* (Artzt & Dung, 2023, p.42). (6) "It recognizes that individuals' ability to exercise their rights, and therefore uphold their dignity, can be compromised in circumstances such as "maternity, illness, industrial accidents, dependency or old age, and in the case of loss of employment*""(Aizenberg & van den Hoven, 2020, p.8).*
Assessments (GDPR, HLEG Framework)	(1) Transparency Provisions (2) Non-Discrimination/Bias	*(1):* *GDPR:* *"Article 12 of Chapter III of the GDPR,*

Category	**Codes**	**Examples**
	Provisions	*concerned with the rights of the data subject, establishes the principle of transparent information and communication" (Kesa & Kerikmäe, 2020, p.76)* *HLEG:* *Human agency. Users should be able to make informed autonomous decisions regarding AI systems. They should be given the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree and, where possible, be enabled to reasonably self-assess or challenge the system. (HLEG, 2019, p.16)* *(2)* *GDPR:* *"The requirements are as follows: - Fairness, which includes preventing individuals from being discriminated;" (Artzt & Dung, 2023, p.49).* *HLEG:* When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. This needs to be addressed prior to training with any given data set. *(HLEG, 2019, p.17)*
GDPR Assessment	(1) GDPR AI-Provisions (2) GDPR Conflicts	(1): *"where automated decision making takes place, there is a "right of explanation""*

Category	**Codes**	**Examples**
		(Artzt & Dung, 2023, p.43) (2): *"The obligation boils down to providing "meaningful information about the logic involved". This can be challenged if the algorithm is opaque"* *(Artzt & Dung, 2023, p.43)*
Trustworthy AI	(1) HLEG Requirements (2) TAI Conflicts	*(1):"1.1 Human agency and oversight AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight. " (HLEG, 2019, p.15)* *(2):* *"training, validation and testing datasets must be relevant, representative, error-free and complete'. Experts point out that the idea of a completely error-free dataset is utopian." (Raposo, 2022, p.108)*
AIA	(1) Risk Categorisation (2) Innovation	

Category	Codes	Examples
	(3) High Risk Obligations (4) Transparency Obligation (5) Governance (6) Codes of Conduct	*(1)* *"The Act addresses three categories of risk: 1. Prohibited Systems. Prohibited systems include AI systems that manipulate human behavior and/or exploit persons' vulnerabilities; social scoring systems; and, save for certain exceptions, "real-time" and "remote" biometric identification (or facial recognition) systems.* *2. High-Risk Systems. While not clearly defined, a "high-risk" system is understood to be one that poses significant risk to health, safety, and fundamental rights. Although the AI Act applies generally to all AI systems, certain provisions contained within the Act (and provided for in Title III) apply specifically to those considered high-risk. (...)"* *(Townsend, 2021, p.4)* *(2)* *"AI regulatory sandboxes establish a controlled environment to test innovative technologies for a limited time on the basis of a testing plan agreed with the competent authorities."* *(European Commission, 2021, p.15)* *(3)* *"The effect of Article 40 is that providers of high-risk AI systems may demonstrate compliance with the onerous set of requirements listed in Chapters 2 and 3 of the Regulation by complying with officially*

Category	**Codes**	**Examples**
		adopted "harmonised standards" that cover them." (McFadden et al., 2021, p.8) *(4)* *"The Act also introduces new legal obligations (such as monitoring, reporting, and transparency obligations) to manage those systems that, although not prohibited, are considered high risk." (Townsend, 2021, p.3)* *(5)* *"The Board will facilitate a smooth, effective and harmonised implementation of this regulation by contributing to the effective cooperation of the national supervisory authorities and the Commission and providing advice and expertise to the Commission. It will also collect and share best practices among the Member States" (European Commission, 2021, p.15).* *(6)"Moreover, those that design and deploy low- or minimal-risk systems are encouraged to adhere to voluntary codes of conduct." (Townsend, 2021, p.5)*
AIA Assessment	(1)AIA Shortcomings (1) Ambiguities (2) Incomplete Scopes (3) Problematic Definitions (4) Missing Mechanisms (5) Faulty Instruments	(1): *"First, the norms require manipulative intention on the part of the person or entity that develops, launches in the market or professionally uses these AI systems.*

Category	Codes	Examples
		However, it is not clear how intent is to be proven if that intention is not declared. *Surely few AI providers explicitly state that they use AI for the purpose of manipulating behaviours and emotions. "(Raposo, 2022, p.93).* (2): *"The prohibition does not cover all scoring systems. The norm requires the scoring AI system to operate 'over a certain period of time', thus excluding episodic scoring from the ban. "(Raposo, 2022, p.94).* (3): *"As a result, the AIA in conjunction with Annex I covers almost every computer programme, merging expert systems, machine learning and statistical approaches together in a definition of AI." (Ruschemeier, 2023, p.369).* (4): *Therefore, there is a need for clear provisions that avoid duplicate tests for such AI systems which would otherwise cause inconsistency in the legal framework with contradictory assessments, accompanied with additional expenditure for companies. Ebers (Ebers et al., 2021, p.596).* (5): *"However, self-assessment has been criticised for its unreliability, cloudiness and discretionary nature and thus the strengthening*

Category	Codes	Examples
		of ex ante obligations has been strongly advocated" (Varošanec, 2022, p.105).

Appendix C, Co-Occurrence Analysis (Atlas.ti)

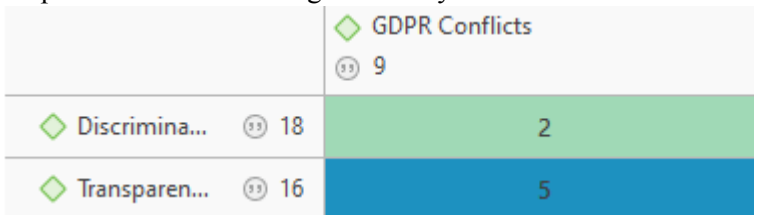
Graphic 1: Results showing the amount of Transparency and Non-Discrimination provisions reflected in identified GDPR AI Provisions.

	GDPR AI-Provisions 21
Non-Discri... 8	3
Transparen... 17	10

Graphic 2: Results showing the amount the frequency of AI challenge code uses.

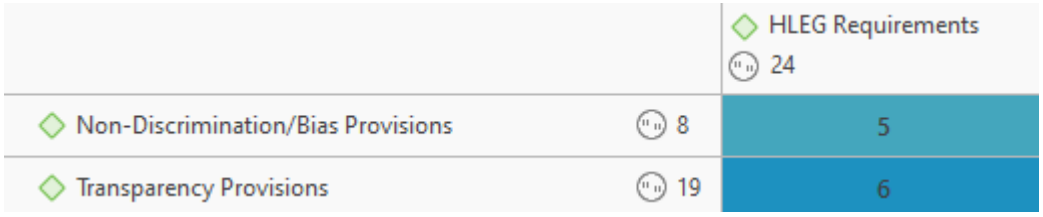
	AI Challenges 31
Dignity 31	31
Discrimina... 18	15
Equality 13	13
Freedom 12	12
Solidarity 4	4
Transparen... 16	11

Graphic 3: Results showing how many of the identified AI challenges for GDPR are attributed to the main challenges discussed in this thesis



Graphic 4: Results show how many times the Discriminatory & Biassed AI and the Algorithmic Transparency & Black Box technology code have been used together

Graphic 5: Results showing the amount of Transparency and Non-Discrimination provisions reflected in the HLEG requirements.



Graphic 6: Results showing the different codes of AIA shortcomings and the frequency of their use

	AIA Shortcomings 42
Ambiguities 9	8
Faulty Instr... 17	16
Incomplet... 6	6
Missing M... 10	10
Problemati... 7	7

Appendix D, Coding Overview (Graphic 1)

	AIA Documents 8 125	GDPR 5 55	Rights & Values A... 7 41	Trustworthy... 4 34	Summe
Dignity 31		6	25	1	32
Discriminatory & B... 18	1	5	12	1	19
Equality 14	1	4	9	1	15
Faulty Instruments 17	17				17
Freedom 12			12	1	13
GDPR AI-Provisions 21		21			21
GDPR Conflicts 9		9			9
Governance 3	3				3
High-risk obligatio... 15	15				15
HLEG Requirements 24				24	24
Incomplete Scopes 6	6				6
Innovation 2	2				2
Missing Mechanis... 10	10				10
Non-Discriminatio... 8		3		5	8
Problematic Defini... 7	7				7
Risk Categorisation 3	3				3
Solidarity 4		1	3		4
TAI Conflicts 20	3	6	10	3	22
Transparency (Blac... 16		8	8	1	17
Transparency Oblig... 4	4				4
Transparency Provi... 17		11		6	17

This graphic shows all codes that have been utilised, along with the frequency of their use including in which document group and how often these codes have been used in the specific group.