UMEÅ UNIVERSITY

ACCEPTABLE RISK; FROM PRODUCT SAFETY TO AI ETHICS

Evaluating the EU AI act through the new legislative framework principles and a risk-oriented theoretical lens

Markus Brännström

Semester 9 HT 2024
Master Thesis, 30 hp
Law Programme, 270 hp
Supervisor: Lena Enqvist

Table of content

1 Introduction

1.1 Background

1.1.1 What is AI, and why does it matter?

Lately, the use of Artificial Intelligence (AI) in a public and professional context has exploded. New applications and levels of complexity are being introduced every day as AI is being imbedded in products, judicial decision making, and law enforcement. With technological advancements, concerns have been raised regarding potential risks that certain AI practices pose to the fundamental rights of individuals within the EU.

To address these concerns the European Commission, *in collaboration with stakeholder organisations, private companies, and academic experts*, introduced the EU Artificial Intelligence Act (AIA). The justification for regulation lies in the inherent risks of this technology, and the harm it can pose to fundamental rights, especially as the AI system's operations become *increasingly capable of autonomy* and are integrated into *decision-making processes* that can have a tangible effect on individuals.

The AIA, which seeks out to regulate the development and deployment of AI, defines AI-systems as "*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*".[1] A key characteristic of AI systems is their capability to infer. This capability to infer refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments. This capability also refers to models or algorithms that which ai systems derive from inputs or data.[2]

The purpose of the regulation is twofold: on one hand, to promote the uptake of human-centric and trustworthy AI, while ensuring a high level of protection of health, safety, fundamental rights, enshrined in the Charter of Fundamental Rights of the European Union; and on the other hand, to improve the functioning of the internal market and supporting innovation by laying

[1] Article 3(1) AIA.
[2] Recital 12 AIA.

down harmonised rules for placing on the market, and putting into service, AI systems in the Union.[3]

The AI act adopts a "risk-based approach", classifying AI systems based on levels of risk ranging from minimal, to high-risk, and *unacceptable*, in doing so, with *some exceptions*,[4] it forbids certain AI practices, while subjecting others to higher regulatory scrutiny.[5]

What is then a "risk-based approach?", the approach is hardly novel and is utilized in a coherent *product safety* legal framework: "The new legislative framework" (NLF) which the AIA is modelled after. The NLF is a framework of directives and regulations that harmonize technical standards regarding potentially harmful products, utilizing legislative harmonisation of essential requirements, voluntary compliance, and standardisation. And by adopting this regulatory framework, the AIA is in essence rendered a *harmonising product safety regulation*.

The choice of instrument and approach can be considered controversial. Framing the potential harms that AI poses to fundamental rights as "risks", with risk-based product safety regulation being the solution can be considered an inherently value-laden choice. And carry normative leanings.[6] Framing AI as a "product" can also be seen as a controversial distinction, considering that AI is a system delivered dynamically through multiple hands, and not a "one off" piece of machinery or a toy.[7]

The NLF is considered a success in harmonizing technical standards in potentially harmful products, it classifies the various contextual risks, *often expressed as hazards*, connected to specific sectors. An example can be taken from toy directive's that expresses risks as: "strangulation, asphyxiation, flammability".[8] The question can be raised is if a product safety approach is appropriate when the risks associated to AI are intangible, and has fundamental rights impacts rather than concrete harms?

[3] Article 1-2 AIA.
[4] See Article 5(1) h i-iii, and annex II.
[5] Article 5-6 AIA.
[6] Kaminski. Regulating the risks of AI. 2023. p. 1351.
[7] Edwards. Expert opinion, Regulating AI in Europe: four problems and four solutions. P.5
[8] Directive 2009/49/EC. Annex II.

The white paper on AI states that the use of AI can affect the values on which the EU is founded and lead to breaches of fundamental rights, including the rights to freedom of expression, freedom of assembly, human dignity, non-discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, as applicable in certain domains, protection of personal data and private life, or the right to an effective judicial remedy and a fair trial, as well as consumer protection. It states that, the risks might result from either flaws in the overall design of AI systems, or from the use of data without correcting possible bias, which can be a risk if a system is trained using mainly data from men leading to suboptimal results in relation to women.[9]

A risk-based approach was by no means the only legislative path forward, as other Union regulations that deal with possible impacts to fundamental rights, such as the GDPR, utilizes a "rights-based" approach. The GDPR working party issued a statement on the role of a risk-based approach in data protection legal frameworks where they stated that *"The so-called "risk-based approach" is not a new concept, […] The legal regime applicable to the processing of special categories of data can also be considered as the application of a risk-based approach: strengthened obligations result from processing which is considered risky for the persons concerned. It is important to note that – even with the adoption of a risk-based approach – there is no question of the rights of individuals being weakened in respect of their personal data. Those rights must be just as strong even if the processing in question is relatively 'low risk'. Rather, the scalability of legal obligations based on risk addresses compliance mechanisms. This means that a data controller whose processing is relatively low risk may not have to do as much to comply with its legal obligations as a data controller whose processing is high-risk."*[10] The statement expresses a concern that a risk-based approach could dilute the importance of protecting fundamental rights, in this case *privacy*, by considering a "low risk" fundamental rights infraction *as less important*.

Meanwhile, in the draft to the AIA, the choice of a risk-based legislative is motivated by following the model of the NLF, implemented through a comprehensive ex-ante conformity assessment through internal checks, combined with a strong ex-post enforcement, is presented as an effective and reasonable solution for high-risk AI systems.[11]

[9] White Paper on Artificial Intelligence: A European approach to excellence and trust. 2020. P. 11.
[10] Statement on the role of a risk-based approach in data protection legal frameworks p. 2.
[11] EU AI Act Draft 5.2.3.

The aim of this thesis will be to evaluate the AIA, through the principles of the new legislative framework and its main principles in order to evaluate:

- To *what extent* is the risk-based regulatory toolbox of the NLF an *appropriate model* for *protecting fundamental rights in the EU AI Act*?

1.2 Methodology and source material

In order to answer the proposed research question, I will examine the NLF and the legal documents that make up the framework. In order to contextualise the NLF into its current form, I will make a historical review of the *legacy directives* and resolutions that formed it to elucidate its strengths and weaknesses. In addition, putting the NLF into a historical context will clarify the decisions leading up to the AIA being modelled after the NLF.

After reviewing the historical background of the NLF I will evaluate it together with the AIA, *in light of the NLF-principles*, by utilizing a comparative methodology. The comparative method, albeit *mainly* used for comparing *different legal systems*, can be used for evaluating *different regulatory instruments within the same legal system*[12], the main goal being to attain a better knowledge of Jurisprudence.[13] The AI act, being modelled of the regulatory framework that makes up the NLF, makes it already a very similar legislation which rightfully might raise the question whether it is worthwhile comparing them at all? And perhaps a different methodological approach could yield more interesting results?

As stated earlier, the purpose of this evaluation of the AIA is *not to discern what is applicable law*, it is to broadly evaluate to what extent the chosen framework of legislation is appropriate. In order to conduct a theoretical discussion of the legislative methodology behind the regulation, I believe that it is crucial to examine the regulation in its context. The goal is to understand if that comparison, even if admittedly applied in an unconventional way, will adequately accentuate the underlying assumptions and structure of the regulation as a legal instrument. Therefore this methodology might produce a better conversation on legislative frameworks than what a purely *hierarchical examination of legal sources* could

[12] Zweigert, Kotz. 1998. p. 2.
[13] Nääv, Zamboni, Valguarnera p. 148.

yield. In its most basic application, comparative law is an intellectual activity with law as its object, and comparison as its process.[14] In a similar fashion, this thesis is an evaluation of a legislation by method of comparison.

In the process of comparison, despite the instruments sharing a similar regulatory structure, I believe that the differences will be drawn into the light and become visible. The similarities and differences themselves might shed some light on why the risk based regulatory approach was chosen. A "microcomparison"[15] of adjacent legislations. The objects of comparison will be the structure of the NLF legislation, with its principles, i.e., *its regulatory toolbox*, superimposed on the AIA.

The NLF regulatory toolbox, *which also permeates the AIA*, consists of, *legislative harmonisation*, *standardisation bodies*, *voluntary standards*, *presumed conformity*, and market surveillance. A commonality of all NLF legislative acts is also a *risk-based approach*[16]

The source *material* I have used in order to conduct this comparison is slightly disparate in dignity. While the *main bulk* of the presentation of the AIA and the NLF, including the historical review of "the new approach", is sourced from current or legacy official EU legislature as the AIA itself, the *Treaty of the functioning of the European Union* (TFEU) and its 1957 predecessor; *the treaty of Rome*, the relevant *regulations*, *directives*, council resolutions, and *decisions*, that makes up the NLF and its predecessor (i.e. "The new Approach") – some of the argumentative or theoretical parts of this thesis is sourced from *legal literature*.

The utilization of legal literature will serve as an *accentuation* of the previously mentioned legislative acts and could be placed in three categories: *Official publications*, *legal doctrine*, and *articles*.

The first category "official publications" is literature that either stems from, *or is adjacent to*, the underlying legislative process of the European Union, or in some capacity stems from official Union sources. This category contains the White paper on artificial Intelligence,

[14] Zweigert, Kotz. 1998. p. 2.
[15] Ibid. p. 5
[16] Not a NLF principle per se but, can be considered the foundational driver behind the justification of harmonisation. Without risk of harm there is no real need for harmonisation of technical standards.

publications from the *High-level expert group on artificial intelligence* appointed by the European Commission, and publications from the European Union Agency for Fundamental rights (FRA). Despite not being legal sources per se, the *relevance*, *reliability*, and *authority* of these sources must be considered to be sufficient, considering they are published by authors and groups that actively were participating and collaborating with the Commission in the drafting process

The second category "legal doctrine" contains legal literature of well-established authority, often used as course literature in legal studies, as Professor Catherine Barnard's *The Substantive Law of the EU: The four freedoms,* which is an authoritative source in the field of EU law. And *An introduction to comparative law* by Professor Konrad Zweigert and Dr. Hein Kötz. Widely considered to be the authoritative and foundational text in the field of comparative law. Also accompanied by Fillipo Valguarnera's section on comparative law in Maria Nääv's and Mauro Zomboni's publication on legal methodology.

The third category, "articles" is literature that is either used to problematise or discuss the information contained in legislative source material or the two previously mentioned categories of legal literature. This category of sources contains material that has a theoretical or argumentative nature, as the well cited publication from Professor Margot E. Kaminski; *Regulating the risks of AI,* which thoroughly examines the possible downsides on utilizing risk-based regulations in regard to AI. Professor Kaminski's article on the risks of AI regulation is cited throughout my thesis in order to balance and challenge my own argument, thus having a strong relevance to both the subject matter and the theoretical conversation on whether AI is an "acceptable risk" or not.

Professor Tobias Mahler's "Concept of risk" and "Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal" used both for giving substance to my theoretical framework and for his thorough examination of the risk-based approach of the AIA.

I cite the historian, Professor Grace Ballor, and her review of the CE marking, which directly overlaps with the emergence of harmonisation legislation finding relevance in her detailed investigation into the "old approach" of EU product safety legislation.

The articles are mainly used in the later, more analytical and argumentative part of my thesis, either to support or *challenge* my own argument, in order to portray more than a single perspective on the subject matter.

1.3 Theoretical framework

For this thesis, I intend to apply a "risk oriented" theoretical framework. In order to do so, I ought to first attempt to define what risk is as a legal concept. However, the concept of "legal risk", is at best a misnomer. There is no consensus on how legal risk should be defined as a central legal concept, the concept of "risk", *both inside and outside of legal practice*, is highly contextual. The word risk can denote anything from the possibility of harm to liability rules within contract law, or "who carries the risk" for a certain transaction.[17]

Consider the term "*risk of rain*". The word risk implies an outcome; "rain", which is a natural occurrence, and depending on which part of the world you reside in, *not* a very uncommon one. The term "risk of rain", when used by a meteorologist, means that by their scientific methods of prediction it is credible that within a more or less calculable margin of uncertainty – your family picnic will be *ruined by the elements*. This is what is referred to as a *quantitative* risk, while they certainly have a place in law, they are not the risks this thesis is concerned with.[18]

Legal risk tends to denote *qualitative* risks, where the risk considered is certain and credible *negative consequences* that is described or defined in either legislation or in contract clauses, this is a prescriptive approach to legal risk where the specific conditions need to qualify to the normative description[19].

Risks involve *negative consequences* of events. To know whether a legal norm has negative consequences, we need to apply the norm to a given set of facts and evaluate the result from the stakeholder's *subjective perspective* as either beneficial or detrimental for his assets or objectives. Risks imply *future* events, meaning that one needs to assume an ex-ante perspective, different from the ex-post perspective of a judge. Risks also entail *likely* future events, obliging us to address matters of *uncertainty*.[20]

[17] see Mahler, 2007, p. 28
[18] compare Mahler, 2022. p. 256
[19] See Mahler, 2007, p. 17.
[20] Ibid, p. 10.

In the *legal context*, Mahler argues, there is a necessary basic distinction between two conditions of uncertainty of an event regulated by a legal norm. Every legal norm consists of an antecedent and a consequent. If the consequent is "negative" then one needs to determine if the norm will be applicable This depends on two essential questions, Is the set of facts true; and second, whether the application of the norm to the set of facts then renders the consequence.[21]

Thus, legal risk, must be connected to some form of legal obligation or norm. As with a non-legal risk, it requires uncertainty. A legal risk that has been actualised is a legal fact. Criminal negligence, tort law, liability, product safety, or "force majeure"-clauses are non-exhaustive examples of legal responses to qualitative legal risks.

A qualitative or prescriptive approach to risk certainly is not the only relevant approach when defining legal risk, it does however align well with how Union harmonisation legislation seems to define it. The EU harmonisation legislation expressed in the NLF, and thus also the AIA, both described as "risk-based legislation" *risk* is consistently defined as "The probability of harm and the severity of that harm"[22]

The risks themselves, sometimes expressed as *hazards*, tend to be qualitative and range from concrete risks in the NLF legislative acts, as flammability, chemical, electrical or biological properties, hygiene, or radioactivity, to abstract qualifiers as the AIA's risk categories linked to possible human rights impacts.[23]

If legal risk can be anticipated, it can be mitigated with clauses or norms, typically based on some kind of risk assessment, liability clause, or regulation.

If risk can be appropriately mitigated to the point that contractual partners agree, or a certain threshold of risk is deemed tolerable in a normative sense, it implies the existence of an "acceptable" legal risk. *Product safety* is an example of such a practice where some risks or potential harm can be mitigated to an *acceptable threshold*. For this thesis I base my theory partly on the definition stated in the NLF where risk is defined as "the combination of the

[21] See Mahler, 2007, p. 11.
[22] See recital 6, and article 3(2) of the AIA.
[23] The Blue Guide, 2022. p. 48.

probability of an occurrence of harm and the severity of that harm", but I hypothesise that the foundational driver of harmonization legislation is and product safety is the interplay between risk/safety on one side, and the free movement of goods on the other. This seems to be corroborated in the justifications of the prototype legislative act of the NLF; the council decision for a new approach to technical harmonization and standards.[24] Acknowledging that; in this thesis, "risk" is not only treated as a descriptive term, but as a *prescriptive regulatory justification for harmonization legislation*.

Products may not ever be fully safe or without risk, but regardless of whether it is a choking hazard or a fundamental rights impact, it may be reduced if a correct response can mitigate it, such as the principles of the NLF. These principles set out technical standards and puts the responsibility of compliance largely on the manufacturers as a compromise between safety and commerce. This tells us that in the commissions choice to employ a risk-based approach on AI, it considers AI a commodity, one that carries risks *yet a commodity nevertheless*, the human impact risks of AI should, following that logic, be levied against the commercial interests of the developers that intend to put AI on the market. With the toolbox of the NLF it seems as if the Commission have decided that the risks concerning AI can be mitigated to something that approaches *acceptable*.

2 The approach of the new legislative framework

2.1 The "old vertical approach" to product safety

In this chapter I will describe the structure that the directives in the new legislative framework follow, to do this I need to briefly address the new legislative framework's legacy frameworks and some significant historical legislative events regarding product harmonisation in Europe.

The old approach to product safety and technical harmonization in the European Community was based on vertical detailed and sector-specific technical rules. [25] These technical rules were aimed at eliminating trade barriers by seeing to those products met unified safety standards in the member states. This approach was slow, inefficient, and inflexible.[26]

[24] Council resolution 85/C 136/01, p. 3 ff.
[25] Barnard, 2022. p. 588.
[26] Ballor, 2021. p.3

In the 1957 Treaty of Rome, a legal basis for product safety and standardization was found in Article 30 EEC, which set out to eliminate customs duties and quantitative restrictions between member states in regard to the import and export of goods, as well as of all other measures with equivalent effect.[27] Certain exceptions to the prohibition on quantitative restrictions were permissible if justified on grounds of (among other things); public morality, public order, public safety, the protection of human or animal life or health. As long as the restrictions didn't constitute arbitrary discrimination, or a disguised restriction on trade between Member States.[28] Article 100 of the EEC gave the Council, by means of a unanimous vote on proposals of the Commission, the competence to issue directives for the approximation of laws across the Member States as have a direct incidence on the establishment or functioning of the Common Market, but the EEC lacked effective means utilize voluntary standards as a means of technical harmonization.[29]

In 1960, some non-EEC affiliated European countries, eager to trade but uninterested in EEC's uniform tariff [30] established the European Free Trade Association (EFTA) and in order to bridge the gap between EEC and EFTA, the Comité Européen de Coordination des Norms (CEN) was established in to promote standardization across Europe.[31]

In the European Court of Justice (ECJ) landmark ruling "Cassis de Dijon" [32] (1979) the principle of mutual recognition was established which was a momentous milestone in trade and harmonisation of standards.

By stating that Member States could only block or restrict the marketing of products from other Member States based on failure to conform with essential requirements. Consequently, the ECJ

[27] Articles 30, 34, and 3, EEC
[28] Article 36 EEC
[29] Ballor, 2022. p. 5.
[30] Austria, Denmark, Norway, Portugal, Sweden, Switzerland, and the United Kingdom
[31] Ballor, 2022. p. 6.
[32] Case C-120/78 Rewe Zentrale v. Bundesmonopolverwaltung für Branntwein ("Cassis de Dijon") [1979] ECR 649, in which a retailer sought to import and sell the French blackcurrant liqueur Cassis de Dijon on the German market but was denied due to Germany's 25% minimum alcohol requirement for certain spirits(!), Cassis de Dijon that contained 15-25% did not have a high enough alcohol content to be sold legally in Germany. The retailer appealed to the ECJ, arguing that the restriction violated Article 30 of the EEC Treaty. The ECJ ruled that products lawfully produced and marketed in one Member State must be allowed in others, which established "the principle of mutual recognition", ensuring free circulation of lawfully produced goods unless restrictions are justified on specific grounds.

laid a foundation for future harmonisation legislation. Since only essential requirements could justify restricting the marketing of a product, harmonisation texts began to focus solely on those essential requirements.[33]

Despite establishing a *legal principle*, the ECJ did not establish any *practical means* in order to create trust in products that national authorities in the member states had not verified independently. By the end of the 1970s, technical barriers persisted, but a new strategy to realise market integration was starting to take shape.[34]

2.2 The new approach

In the 1980s a new regulatory framework of product harmonization emerged, dubbed as the "new approach".[35] The new-approach directives utilised a *horizontal* structure and applied throughout industries instead of being directed to individual sectors within that industry. The directives set out general principles rather than detailed rules and relied on private bodies to set voluntary standards.[36]

The four 1985 new approach principles that were outlined in the resolution were stated such that: Legislative harmonization would be limited to the adoption of the *essential safety requirements* with which products put on the market would have to conform and which should therefore enjoy free movement throughout the Community. [37] The task of drawing up the technical specifications needed for the production and placing on the market of products conforming to the essential requirements established by the directives would be entrusted to competent organizations in the standardization area. The technical specifications would be non-mandatory and would maintain a status as *voluntary standards*. National authorities would however be obliged to recognize that products manufactured in conformity with harmonized standards would be presumed to conform to the "essential requirements" established by the Directive. Producers would have the choice of not manufacturing in conformity with the standards but would then have an obligation to prove that their products still conformed to the essential requirements.

[33] The Blue Guide, 2022. p.
[34] Ballor, 2021. p. 9.
[35] Council resolution 85/C 136/01.
[36] Barnard, 2022. p. 588.
[37] Council resolution 85/C 136/01, Annex II.

The principles and the structure of the "new approach" resolution and directives were based on the experience gained from an earlier "prototype" of harmonization legislation: the Low Voltage Directive, put in use 1973.[38] The approach was meant to ensure free movement of goods, on the condition *that manufacturers guaranteed that their products were safe*. The Legislation set out the levels of protection that had to be achieved and did not evaluate the choice of technical solutions used to achieve those levels.[39]

The new Approach improved efficiency in legislating harmonised standards for products, yet lacked strong enforcement mechanisms, lacked a unified certification, and failed to address the differences between national systems for conformity assessments.[40]

In 1992 the Council released the council directive 92/59/EEC on general product safety. The directive was made to work alongside the directives of the new approach product safety legislations, with an explicit "lex specialis clause" making it only applicable to consumer product not regulated by other union legislation with the same objectives. The directive was poorly received and co legislators considered it incomplete and some of its provisions indistinct. The directive was reworked into Directive 2001/95/EC, also known as the General Product Safety Directive (GPSD) in 2001.

Following the New Approach council resolution, a 1987 directive regarding simple pressure vessels implemented the "EC Mark.", to be used to indicate a product's compliance with the essential requirements. It did not manage to provide concrete procedures for assessing conformity [41], yet laid the groundwork for the "CE marking" currently in use.[42]

Missing from the framework was coherent ways of affixing the CE mark, and means to survey presumed conformity ex-post, with the Requirements for accreditation and market Surveillance (RAMS) Regulation 765/2008, general principles on the accreditation of conformity assessment bodies, on CE marking, and on market surveillance to the New approach toolbox, completed by decision 768/2008/EV on a common framework for the marketing of products, which

[38] Barnard, 2022. p. 589; Council directive (73/23/EEC).
[39] Commission, "Package on Internal Market for goods", MEMO/07/54.
[40] Ballor, 2021. p. 11.
[41] Ballor p. 12.
[42] Implemented in council directive 93/68/EEC of 22 July 1993.

provides common principles and reference provisions for the purposes of legislation based on the new approach principles.

2.3 "The new legislative framework"

The NLF is the contemporary evolution of the "new approach" framework.[43] And consists mainly of:

- Regulation (EC) 765/2008, which contains the requirements for accreditation and the market surveillance of products.

- Decision 768/2008 on a common framework for the marketing of products, and

- Regulation (EU) 2019/1020 on market surveillance and compliance of products.

Alongside the acts that sets up the structure of the new legislative framework, it is also a set of decisions, directives and regulations[44], harmonising specific sectors such as, the Machine directive[45], the toy directive[46], and the IVD Regulation.[47] Three vastly different sectors, but with risk based NLF structured harmonising legislation using the "boilerplate" structure of the new approach regulatory principles, with the clarification brought by the new legislative approach regarding accreditation CE-markings and market surveillance.

Barnard incorporates all the guiding legislations, including those of the NLF and pinpoints *five* principles of the original new approach, adding *market surveillance* as a fifth. It could however be argued that there are *six*. The NLF is only partly the regulations and directives that gives it a structural framework, but the directives and regulations based on the framework all have a commonality: *Risk.*

[43] The "New Approach" and the "NLF" names for the frameworks are sometimes used interchangeably. Barnard, as an example, refers to the framework as its legacy framework; "the new approach", while adding the RAMS, the market surveillance, and accreditation directives *as principles* in its regulatory toolbox. See Barnard. P 589.
[44] Not part of the framework, but a complement to it, is The General Product Safety Regulation (GPSR), repealing its weaker predecessor the GPSD, establishes a baseline of product safety that applies broadly to non-harmonised sectors. It can be considered as a safety net that catches possible risks in products that has not been foreseen by experts in various sectors.
[45] Directive 2006/42/EC.
[46] Directive 2009/48/EC.
[47] Regulation (EU) 2017/746.

None of the operational sectorised directives and regulations that exist within the framework would have any justification for harmonization if it were not for a credible and potentially negative outcome, mitigated by the contents of the directive. Taking that into account the principles would then be as follows: Risks, legislative harmonisation of essential requirements, European standardization bodies, voluntary harmonised standards, and presumption of conformity

2.3.1 Risks – a justification for legislation

In the original 1985 council resolution on a new approach, it is stated that: Among the traditional principles justifying a Directive, the following aspects should be emphasized, member States have the responsibility of ensuring safety on their territory (in the home, at the workplace, etc.) of persons, domestic animals and goods, or the respect of other essential protection requirements in the general interest such as health, consumer or environmental protection. National provisions ensuring such protection must be harmonized in order to ensure the free movement of goods, without lowering existing and justified levels of protection in the Member States[48] The importance of a guarantee of safety is twofold, It is given that all consumers, including the most vulnerable, such as children, older persons or persons with disabilities, have the right to safe products.[49]

Risk based legislation protects the actual physical well-being of its citizens, but harmonisation of safety standards is a means of ensuring the free movement of goods within the union. A certification of safety is in the direct interest of the producers themselves since it can ensure consumers- of any member state – that their products are up to standards – *literally*.

It is a balancing act of market incentive and the fundamental "right" to health. The actual standards and responses to risks are contextual due to the varied nature of the harmonised sector yet justified by a need for *specificity* that cannot be covered by the "baseline" product safety net, granted by the GPSR. In the Machine directive, article 9 outlines specific measures to deal with potentially hazardous machinery, while Annex I, details specific hazards related to the use of machinery. A "Risk based" NLF legislation always deal with the combination of the

[48] Council resolution 85/C 136/01.
[49] See Recital 5 of the GPSR.

probability of an occurrence of harm and the severity of that harm[50]. minus appropriate mitigative action[51], in this case "appropriate mitigative action" is the material "contextual" content of the respective directive in addition to the NLF principles described below.

2.3.2 Legislative harmonization of essential requirements

Legislative harmonization is limited to the adoption of the essential safety standards with which products put on the market must conform. Once goods conform to these standards and a CE marking is placed on the goods, they must enjoy free movement throughout the union without the need for verification.[52] General principles governing the CE marking are set out in Regulation (EC) No 765/2008) and are defined as markings on a product by which the manufacturer indicates that the product is in conformity with the applicable requirements set out in Union harmonisation legislation providing for its affixing.[53]

A fundamental feature of Union harmonisation legislation is to limit legislative harmonisation to the essential requirements that are of public interest. Typically, such requirements target the protection of health and safety of consumer or worker, but can also cover protection of property, scarce resources or the environment. Essential requirements are designed to ensure protection from certain *hazards* associated with a product, or refer to provisions regarding materials, design, construction, manufacturing process, instructions drawn up by the manufacturer, or lay down principal protection objectives.[54]

Essential requirements *must be applied as a function of the hazard inherent to a given product*. Manufacturers have to carry out a risk analysis to first identify possible risks that the product may pose and determine the essential requirements relevant for the product. Essential requirements should define the results the manufacturer should attain, while leaving the specifics of the technical solutions up to the manufacturer. Precise technical solutions can be provided by harmonisation standards. This flexibility both allows manufacturers to choose the way to meet the requirements, and it allows for manufacturers to adapt to technological progress.[55]

[50] Directive 2009/48/EC, Annex I, 1. General principles.
[51] See chapter 1.3 above. Theory.
[52] Barnard, 2022. p. 589.
[53] Article 2(20) Regulation (EC) No 765/2008.
[54] The Blue Guide, 2022. p. 48.
[55] Ibid.

2.3.3 European standardisation organisations

The task of drawing up the technical specifications which satisfy the essential requirements is entrusted to specialist European standardisation organizations (ESOs), Comité Européen de Normalisation (CEN), Comité Européen de Normalisation Électrotechnique.[56] (CENELEC), and European Telecommunications Standards Institute (ETSI), acting by qualified majority voting, on a mandate from the commission.[57]

The mandate requires the ESOs to draft harmonised standards following a *standardisation request* issued by the Commission.[58] A harmonised must be clear which requirements are aimed to be covered to ensure that a manufacturer, public authority, or notified body know which essential requirements must be in place for a "presumption of conformity" to apply.[59]

Harmonised standard can after its adoption be challenged by the ESOs, also the reference to the harmonised standard can be removed from the OJEU by the Commission if the relevant edition of a harmonised standard is not reviewed or updated by the ESOs. Since essential requirements is based on the assumption that the harmonised standards reflect technological "state of the art" the ESO review the harmonised standards regularly in accordance with the relevant standardisation request. If a harmonised standard is not relevant anymore, it can no longer provide a presumption of conformity. Harmonised standards can however when necessary be revised by the ESOs, either by their own initiative or initiated by the commission directly via a standardisation request, or indirectly through a decision or formal objection.[60]

2.3.4 Voluntary harmonised standards

Harmonised standards mean a non-binding technical specification adopted by a standardisation body on the basis of a remit issued by the Commission in accordance with the procedures laid down in Directive 98/34/EC of the European Parliament and of the Council of 22 June 1998 laying down a procedure for the provision of information in the field of technical standards and regulations and of rules on Information Society services. [61]

[56] See chapter 2.1 above. The "old approach".
[57] Barnard, 2022 p. 589.
[58] The Blue Guide, 2022. p. 49.
[59] Ibid. p. 50.
[60] The Blue Guide, 2022. p. 54.
[61] Directive 2009/48/EC, Article 2, l.

These harmonised technical standards are *voluntary*, and the producer has the choice of either manufacturing in accordance with the standards set by the voluntary bodies or manufacturing according to other standards and then proving that they conform with the essential requirements of the directive through a system of tests and certificates.

Manufacturers can choose whether or not to apply such harmonised standard, if they choose not to apply a harmonised standard, they must demonstrate that the products conform to the essential requirements despite using other means of their own choice.[62]

2.3.5 Presumption of conformity

Products manufactured in conformity with harmonized standards are presumed to conform to the essential requirements established by the relevant NLF legislation for that product category.

European standards, including harmonised standards, may be based fully or partially on international ISO or IEC standards. However, the presumption of conformity is possible only when applying the European version published by reference in the Official Journal, because of possible technical modifications introduced in it to ensure adequate compliance with the legal requirements. Additionally, ISO and IEC versions do not contain information about which provision of the standard is relevant for which essential requirement, as this information is only included in their European version. If the manufacturer applies only a part of a harmonised standard or the harmonised standard does not cover entirely all relevant essential requirements, the presumption of conformity exists only to the extent the harmonised standard corresponds to the essential requirements. [63]

2.3.6 Market surveillance

Market surveillance is the instrument that competent national authorities utilise to monitor products *already on the market* in order to ensure that they comply with harmonisation legislation. Its legal basis is found in Regulation (EC) 765/2008, which contains the requirements for accreditation and the market surveillance of products (the RAMS-regulation), and Regulation (EU) 2019/1020 on market surveillance and compliance of products. Member

[62] The Blue Guide, 2022. p. 52.
[63] The Blue Guide, 2022. p. 51.

states must take all appropriate enforcement measures, including market surveillance, to ensure that non-conforming products are withdrawn from the market.[64]

A core function of market surveillance is to provide a high level of protection of public interests such as health and safety in general, health and safety in the workplace, the protection of consumers, of the environment and of public security and any other public interest protected by the EU legislation.[65]

In order to provide a high level of protection. market surveillance authorities assess the most appropriate action to be taken on a case-by-case basis and with a view to the *principle of the proportionality*. Actions taken are based on the *level of the risk,* if the economic operator is identifiable, the urgency, if previous measures have been taken against a specific product, etc.[66]

3 The EU AI Act

3.1 An overview of the general structure and aim of the AIA

The EU AI imposes harmonised rules for the use of AI in the Union, in order to "improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence"[67]. Additionally, the regulation aims to ensure a "high level of protection" of health, safety, and fundamental rights, against the harmful effects of AI systems in the Union, while boosting innovation and employment and making the Union a leader in the uptake of trustworthy AI.[68]

Due to the major impact that AI can have on society[69], it is vital for AI and its regulatory framework to be developed in accordance with Union values as enshrined in Article 2 of the Treaty on European Union (TEU), the fundamental rights and freedoms enshrined in the Treaties and, pursuant to Article 6 TEU, the Charter.[70]

[64] Barnard 589, recital (9) of Regulation (2019/1020).
[65] The Blue Guide, 2022. p. 104.
[66] Ibid. 46.
[67] Article 1(2) a and d, AIA.
[68] Recital 2 of the AIA.
[69] Recital 6 of the AIA.
[70] Studies on the possible fundamental rights impact of AI and other algorithmic or automated data processing has for long projected potential threats to several fundamental rights, as (but not limited to), fair trial and access to

Furthermore, beyond mitigating possible impact on fundamental rights, and protecting innovation and European competitiveness on AI on the global market, the regulation also aims to protect public interests, such as critical infrastructure and environmental protection.

The general structure of the AIA encompasses, as to be expected, every principle in the NLF toolbox which shall be demonstrated through following subchapters, with an emphasis however on the *requirements* for high-risk AI.

3.1.1 Risk classification

Being based on the NLF, the AI act utilises a risk-based approach[71] and classifies systems or practices according to the risk it poses. "Risk" is in the AIA understood as the "combination of the probability of an occurrence of harm and the severity of that harm"[72]. "Harm" then, being an impact on previously mentioned fundamental rights and freedoms. The classifications are hierarchical in dignity and range from unacceptable risk, high risk, systemic risk, limited risk, and minimal risk. [73]

The unacceptable risk category is expressed in Article 5 of the act and outlines *forbidden AI practices.* These include subliminal or manipulative techniques used to affect people's behaviour, especially if it leads to them making decisions that may cause harm. Systems that exploit people's vulnerabilities due to age, disability, or social circumstances to influence their behaviour. Social scoring based on behaviour or characteristics, leading to negative treatment of individuals or groups. Data collection for facial recognition from the internet or surveillance cameras is prohibited. Also prohibited is biometric remote identification in real time in public places for law enforcement.[74] Article 5 covers AI systems posing an unacceptable risk to fundamental risks, as the respect for human dignity, democracy, or the protection of fundamental rights.[75]

justice, freedom of expression, prohibition of discrimination, the right to free elections and data protection. (European union agency for fundamental rights Getting the future right, p. 10 - Artificial intelligence and fundamental rights, council of Europe study DGI (2017)12, p. 10-32,).

[71] Recital 26 of the AIA.

[72] Article 3(2) AIA.

[73] Article 5-6, AIA.

[74]The regulation allows for certain exceptions expressed in article 5:1 h iii, these exceptions include searching for victims of crimes such as human trafficking or to prevent an imminent threat to people's safety or terrorist attacks. The use of such systems requires prior approval from a judicial or administrative authority and may only be used if it is strictly necessary and proportionate.

[75] Recital 15, AIA.

The majority of the regulation concerns the "High risk" classification[76] an AI system is considered high risk when its use case is in any of the following areas:[77] Biometrics, Critical infrastructure, Education and vocational training, Employment, workers management and access to self-employment, Access to and enjoyment of essential private services and essential public services and benefits, Law enforcement, Migration, asylum and border control management, and Administration of justice and democratic processes.

The classification of an AI system as high-risk is based on the intended purpose of the AI system, in line with existing product safety legislation. The classification as high-risk does not only depend on the function performed by the AI system, but also on the specific purpose for which that system is used. [78]

Furthermore, a system is considered high risk if its intended to be used as a safety component of a product, or that the AI system in itself, is part of union a product, covered by the Union harmonisation legislation listed in Annex I, *and* is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the NLF harmonisation legislation listed in Annex I.

Limited risk AI systems are subject to lighter transparency obligations where developers and deployers must ensure that end-users are aware that they are interacting with AI. Systems related to products that are not high-risk in accordance with article 6 and are therefore not required to comply with the requirements set out for high-risk AI systems are safe when placed on the market or put into service. These safer systems are regulated by the GPSR that applies as a "safety net".[79]

The provisions and requirements of the AIA applies to certain operators, defined as providers, product manufacturers, deployers, authorised representatives, importers or distributors.[80]

[76] Article 6, AIA.
[77] See annex III. AIA.
[78] EU AI Act Draft 5.2.5.
[79] Recital 166, AIA.
[80] Article 3(8) AIA.

Although the majority of the obligations fall on the *providers* of high-risk AI systems, the AIA outlies several responsibilities and restrictions for the deployer, i.e. the *user* of the system.[81]

3.1.2 Requirements for High-risk AI systems

High-Risk AI products and de must comply with the requirements listed in chapter III, section 2, articles 9-15, while taking the product's *intended use* as well as the generally acknowledged *state of the art* on AI and AI-related technologies. AI systems which are components of products that is regulated under NLF harmonisation legislation, obliges providers to be responsible for ensuring that the product is compliant with all applicable requirements of all relevant product harmonisation legislation. [82] As with the NLF, CE markings apply, indicating that a provider's AI-system is in conformity with the requirements.

The requirements are risk management systems, data governance, technical documentation, record keeping, transparency and information to deployers, human oversight, and accuracy, robustness and cybersecurity. These requirements are not *in name* the same as the "essential requirements" that NLF legislation focuses on, but for *high-risk* AI systems and components they are for the providers *functionally* precisely that. Thus, the requirements and their *functions* shall be covered in following *subsections*.

3.1.2.1 Risk management systems

Article 9 (1) of the AIA states that for high-risk AI systems, a risk management system must be established, implemented, documented and maintained.

The risk-management system should consist of a *continuous, iterative process*, active throughout the entire lifecycle of a high-risk AI system, aimed at *identifying and mitigating the relevant risks of AI systems on health, safety and fundamental rights. Any known or foreseeable circumstances related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse should be included in the*

[81] A point of clarification since some confusion may arise with the terminology of the AIA: When a deployer is referred to as the *user* of a system, it should not be confused with the "end-user" or *consumer* of the system or product. The relationship between the provider and the deployer (as a user of a system) can be compared to the relationship between a producer and a retailer, or a service provider that uses a technical infrastructure that they themselves may not have produced. This being said, it's of course entirely possible to simultaneously be the provider *and* the deployer of an AI-system.
[82] Article 8, AIA.

instructions for use that are provided by the provider to ensure that the deployer is aware and takes them into account when using the high-risk AI system.[83]

This process of identifying and mitigating relevant risks is meant to lead to the adoption of appropriate and targeted risk management measures designed to address the identified risks.[84]

This process is both separate and connected to the deployer's obligation to perform fundamental right impact assessments found in article

In summary, the function of the risk management system requirement is to ensure that *the providers identify and mitigate risks throughout the lifecycle of their AI systems*.

3.1.2.2 Data governance

It is stated in article 10 (1) that if a high-risk AI system uses techniques involving the training of AI models with data, these systems shall be developed on the basis of training, validation and testing data sets that meet the certain quality criteria listed in article 10 (2) a-h.

The training, validation, and testing of data sets shall be subject to *data governance* and management practices *appropriate for the intended purpose of the high-risk AI system*. Those practices shall concern in particular: Relevant design choices; data collection processes and the origin of data, and in the case of personal data, the original purpose of the data collection.

Providers must implement relevant data-preparation processing operations, such as annotation, labelling, cleaning, updating, enrichment and aggregation, and the formulation of assumptions, in particular with respect to the information that the data are supposed to measure and represent.

Providers must assess the availability, quantity and suitability of the data sets that are needed and examine possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights or lead to discrimination prohibited under Union law, especially if data outputs influence inputs for future operations. If such biases are identified then appropriate measures to detect, prevent and mitigate those biases should be implemented.

[83] Recital 65, AIA.
[84] Article 9, paragraph 2 (d), AIA.

Providers must identification relevant data gaps or shortcomings that prevent compliance with the AIA, and how those gaps and shortcomings can be addressed

This requirement exists in large part in order to ensure that high-risk AI systems, *especially those using techniques involving the training of models*, performs as intended and does not become a *source of discrimination* prohibited by Union law.[85]

In summary, the function of the data governance requirement is to ensure that providers use high quality data that is relevant, representative, and absent of *bias*.

3.1.2.3 Technical documentation

Before a high-risk AI system is placed on the market, a technical documentation must be drawn up and kept up-to date. The technical documentation should in part demonstrate that the high-risk AI system complies with the requirements set out in article 11 of the AIA, and in part it should provide national competent authorities and notified bodies with the necessary information to assess the compliance of the AI system with those requirements.

The minimal requirements for technical documentation are listed in Annex IV, they consist of: A general description of the AI system which, among other things, states the system's *intended use*, a detailed description of the system's design and development, data and training information.

If applicable it should contain validation, testing and performance metrics, assessments and measures pertaining to human oversight and risk control,

Documentation of harmonised standards that has been applied or what other technical solutions has been used as an alternative, an EU declaration of conformity, a detailed description of the system in place to evaluate the AI system performance in the post-market phase including a post-market monitoring plan.

The EU declaration of conformity is detailed in article 47 of the AIA, providers of high-risk AI systems are obliged to draw up such a document, and it should state that the system meets the

[85] Recital 67, AIA.

requirements set out in section 2, and should contain all the information that is referred to in Annex V which pertains to identification of the system, a statement of compliance with Union law, references to relevant harmonised standards or other declarations of conformity, if applicable the name and identification of any involved notified body and their conformity assessment procedures performed, and information the name and function on who signed the declaration.

The post-market monitoring system shall actively and systematically collect, document and analyse relevant data which may be provided by deployers, or which may be collected through other sources on the performance of high-risk AI systems throughout their lifetime, and which allow the provider to evaluate the continuous compliance of AI systems with the requirements set out in Chapter III, Section 2.

The monitoring is key to ensuring that the possible risks emerging from AI systems which are able to continue to learn after being placed on the market can be more efficiently and timely addressed. Providers are required to have a system in place to report to the relevant authorities any serious incidents resulting from the use of their AI systems, meaning; "*incident or malfunctioning leading to death or serious damage to health, serious and irreversible disruption of the management and operation of critical infrastructure, infringements of obligations under Union law intended to protect fundamental rights or serious damage to property or the environment.*"[86]

In summary, the function of the technical documentation requirement is to ensure that the providers of high-risk AI-systems have documentation that demonstrate their *compliance with all union obligations* to the national competent authorities and notified bodies.

3.1.2.4 Record keeping

Article 12 of the AIA obliges providers of high-risk AI systems to implement automatic recording logs over the lifetime of the system in order to ensure a level of traceability of the functioning of the system appropriate to its intended purpose.

[86] Recital 155, AIA.

Logging capabilities shall enable the recording of events relevant for; identifying situations that may result in the system presenting a risk to fundamental rights, including the right to non-discrimination.[87]

If the system is designed as a remote biometric identification system described in Annex III point 1 (a), the logging capabilities should at minimum provide a recording of the period of each use of the system, the reference database against which input data has been checked by the system, the input data for which the search has led to a match, and the identification of the natural persons involved in the verification of the results.

These requirements are necessary to effectively mitigate the risks for health, safety and fundamental rights. Since no other less trade restrictive measures are reasonably available, those requirements are not considered as unjustified restrictions to trade.[88]

In summary, the function of the record keeping requirement is to ensure that the high-risk AI-systems have *traceability* and can be *investigated through automatic system logs*.

3.1.2.5 Transparency and information to deployers

Article 13 of the AIA states that high risk AI systems must be constructed in such a way that it provides a sufficient transparency for the deployers to interpret and evaluate a system's output, such as predictions, recommendations, decisions, or other content generated by the system.

The provider of the system must accompany the system with instructions for use that include concise, complete, correct and clear information that is relevant, accessible and comprehensible.

The instructions should minimally consist of the provider's identity and contact details, and the characteristics of the system including its *capabilities and limitations*, any changes to the system and its performance from the moment of the initial conformity assessment, human oversight measures, computational and hardware resources needed to operate the system, the system's expected lifetime and necessary maintenance measures, and if relevant a description on how to properly process logs in accordance to the record keeping.

[87] See below in chapter 3.1.6, on market surveillance.
[88] Recital 66, AIA.

Regarding the information regarding the characteristics of the system, these must in turn state the system's intended purpose, levels of accuracy, robustness, and cybersecurity, and potential risks to health, safety, and fundamental rights, stemming from *reasonably foreseeable misuse* of the system.[89]

If applicable, the information should also state the technical capabilities, its performance regarding specific groups or persons on which the system to be used, specifications for input data in terms of the training, validation and testing data sets used.[90]

These instructions must be made available in a language which can be easily understood by the target users, as determined by the member state the deployer is operating from.[91]

In summary, the function of the transparency and information requirement is to ensure that providers design their high-risk AI-system's operations and output to be clear and understandable to the user, including the system's *capabilities and limitations*.

3.1.2.6 Human oversight

Article 14 of the AIA states the requirement that high-risk AI systems must be constructed with an appropriate interface that allows a natural person, a *human*, to oversee its operations during its lifecycle. This human oversight aims to prevent possible risks to health, safety, and fundamental rights, that may stem from using the system, both in accordance with its intended use, *but also under conditions of reasonably foreseeable misuse*.

The oversight measures should be proportionate with the *risks* of the system and its *autonomy and context*. The measures can be built in by the provider and/or identified by the provider and implemented by the deployer, before the system is put on the market.

For high-risk AI systems that are developed for remote biometric identification systems, the measures must be implemented in such a way that no action or decision is taken by the deployer

[89] See below on accuracy, robustness, and cybersecurity, subchapter 3.1.2.7.
[90] See above on data governance, subchapter 3.1.2.2.
[91] Recital 72, AIA.

on the basis of the identification unless it has been *separately verified and confirmed by at least two natural persons with the necessary competence, training, and authority*, with an exception to systems used for law enforcement, migration, border control, or asylum where Union or national law considers the application of that requirement to be disproportionate.

In summary, the function of the human oversight requirement is to ensure that the high-risk system can be overseen by natural persons that can intervene in its operations when it is necessary, especially in proportion to the system's autonomy.

3.1.2.7 Accuracy, robustness, and cybersecurity.

Article 15 of the AIA states the requirement that high-risk AI systems must be made in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle.

The term accuracy is vague, and a clear definition is not found within the AIA or its recitals, but there are comparisons to the need for accuracy within various sector, as how Union legislation on legal metrology aims to ensure the accuracy of measurements and to help the transparency and fairness of commercial transactions. [92] Or in the health sector where the stakes for life and health are particularly high, sophisticated diagnostics systems should be reliable and accurate.[93] To address the technical aspects of how to measure the appropriate levels of accuracy in a system, the Commission cooperates with relevant stakeholders and organisations to encourage the development of measurement methodologies.

The robustness of a system refers to its resilience towards harmful or otherwise undesirable behaviour that can result from its usage, the system's internal limitations, or the environment in which the system operates.[94] Cybersecurity, similar to robustness, refers to the system's resilience, but from external behaviour as malicious third parties attempting the systems vulnerabilities.[95]

[92] Directives 2014/31/EU (35) and 2014/32/EU (36).
[93] Recital 47, AIA.
[94] Recital 75, AIA.
[95] Recital 76, AIA.

In summary, the function of the accuracy, robustness, and cybersecurity requirement is to guarantee the system's performance under expected conditions and resilience against attacks.

3.1.3 Standardisation bodies

The AIA, being modelled after the NLF, does not deviate from the established process where harmonised standards are drafted by ESOs such as CEN, CENELEC, or ETSI, based on standardisation requests issued by the commission.[96]

What can be added in regard to the standardisation process that is specific in the context of the AIA is that participants in the standardisation process shall seek to *promote investment and innovation in AI* as well as the competitiveness and growth of the Union market.

The stated goal is to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI that are consistent with Union values and fundamental rights and interests. [97]

The process should represent a wide range of interests, as all relevant stakeholders in the development of standards, in particular small and medium-sized enterprises, consumer organisations, and environmental and social stakeholders.[98]

3.1.4 Voluntary harmonised standards.

As with other NLF legislations, the ESOs draft voluntary harmonised standards based on standardisation requests issued by the Commission.

In the context of the AIA, the Commission issued a formal standardisation request to CEN and CENELEC in May 2023.[99] The request outlines the specific areas in which the Commission has deemed urgent for standardisation. Among these are harmonised standards pertaining to the requirements for high-risk AI, as defined in articles 9-15 of the AIA.[100]

[96] See above on European standardisation bodies, 2.3.3.
[97] Article 40 (3) AIA.
[98] Recital 121, AIA.
[99] Decision (C(2023)3215).
[100] See p. 4-5 in Decision (C(2023)3215).

Within the interim period of the AI act being implemented and the harmonised standards being drafted and put into use, the Commission, *together with stakeholder organisations*, are in the process of developing voluntary codes of conduct for high-risk AI systems, and voluntary codes of practice for general purpose AI-systems, especially those that are deemed to pose systemic risk.

These voluntary codes and practices are not legally binding in the sense that they can provide a presumption of conformity in the same manner as harmonised standards, but act as a *placeholder* for these, beneficial to providers or organisations that want to demonstrate compliance while they prepare for the full implementation of the AIA, its technical essential requirements, and the voluntary harmonised standards and the presumption of conformity they provide.[101]

3.1.5 Presumption of conformity

Once a standardisation body, as the relevant ESOs, have drafted a harmonised standard and published a reference in the OJEU, High risk or general-purpose AI systems that are compliant to these standards shall be presumed to be in conformity with the requirements set out in the requirements for high-risk AI systems stated in chapters 9-15 of the AIA.[102]

For each high-risk AI system that the provider develops, an EU declaration of conformity must be drawn up and kept it available to national competent authorities for 10 years after the system is put on the market or put into service. The declaration of conformity must demonstrate that high risk AI meets the requirements in articles 9-15 of the AIA, and if applicable, compliance to any other requirement in relevant EU harmonisation legislation. By drawing up this declaration, the provider assumes responsibility for the system's compliance[103]

High-risk AI systems should bear the CE marking in order to display its conformity with the requirements of the AIA. If embedded into a product a physical CE marking should be affixed, if the high-risk AI system is only provided digitally a digital CE marking should be used. Regardless of the physical or digital nature of the CE marking, member States should not create

[101] See articles 56, 95, and 96, and recital 117 of the AIA. More information on the "AI-Pact" and its stakeholder organisations, as the CEDPO, can be found on the Commissions official home page for its digital strategy for Europe: https://digital-strategy.ec.europa.eu/en/policies/ai-pact.
[102] See article 40(1).
[103] Article 47, AIA.

unjustified obstacles to the placing on the market of high-risk AI systems that comply with the requirements of the AIA and is affixed with a CE marking.[104]

3.1.6 Market surveillance

The previously mentioned Regulation (EU) 2019/1020 on market surveillance and compliance of products is applicable to AI systems covered by the AIA. Any reference to an *economic operator* under the market surveillance regulation should be understood as the term operator as it is defined in the AIA, a provider, deployer, importer, distributor, product manufacturer, or representative, and any reference to a product should be understood as any type of AI-system that falls withing the scope of the AIA. [105]

The market surveillance authorities report any information identified that might be of potential interest for the application of Union law on competition rules or the use of prohibited practices to the Commission and if applicable, national competition authorities. For high-risk AI systems related to products covered by NLF Union harmonisation legislation the market surveillance authority for the purposes of that Regulation shall be the authority responsible for market surveillance.

Although mostly the high-risk AI systems are subject to specific requirements and obligations under the AIA, market surveillance authorities may take measures in relation to all AI systems when they present a *risk* in accordance with the AIA.[106]

4 Evaluating the AIA through the principles of the NLF

4.1 Risks

In the model of the NLF risks are the justification for legislation, and the reasoning is twofold, on one side to protect public interests as health, safety, property, or the environment from certain *qualitative risks*, or *hazards*, in order to ensure a "high level of protection", and on the other side to allow the free movement of good within the internal market by providing a voluntary framework of compliance in trading with goods that has certain risk factors connected to them.

[104] Recital 129, AIA.
[105] Articles, 74, 2(1), and 3(8), AIA.
[106] Recital 156, AIA.

Realising the risks that are inherent to AI, the AIA uses the NLF risk-based model and expounds on it, in some parts taking a precautionary approach and denying market access whenever risks stemming from their use are deemed too high for risk-mitigating interventions.[107] For the most part, the provisions of the AIA target those AI systems and practices that it deems to be "high-risk", while limited risk AI systems are subject to lighter transparency obligations and in part regulated by the GPSR.[108]

Worth mentioning is the requirement of a risk management system in article 9 of the AIA, which consist of a continuous and iterative process that is planned and run throughout the entire lifecycle of a high-risk AI system. This requirement is an *intrinsic to the system as a product*, which places obligations on the provider *separate from the mandatory obligation of the deployer* to assess the system's potential impacts to fundamental rights prior to being put on the market or put in service. This practice deviates from the standard procedure of the NLF risk management where a risk is addressed and mitigated as a "one off event", so to say, which most likely is related to the difference in the nature of the risks that is addressed.

The AIA does not deviate from the NLF in that the risks that it addresses are still very much qualitative, *the difference* is in how previous product safety legislation has focused on specific and concrete hazards instead of complex and intangible human rights impacts that might complicate risk assessments using the same system.

4.2 Legislative harmonisation of essential requirements

In the NLF, legislative harmonisation is limited to the essential requirements that are of public interest and must be applied as a function of the *hazard* inherent to a given product. Essential requirements should define the results the manufacturer should attain, while leaving the specifics of the technical solutions up to the manufacturer.[109]

[107] Almada, Petit, 2023. p. 8.
[108] Recital 166, AIA.
[109] The Blue Guide, 2022. p. 48.

In the AIA, high-risk AI systems must conform with the requirements listed in chapter III, section 2, articles 9-15, while taking the product's *intended use* as well as the generally acknowledged state of the art on AI and AI-related technologies

These requirements are not in name the same as the "essential requirements" that NLF legislation focuses on, but for the providers of high risk they are functionally precisely that: legislative requirements that outline the minimum requirements of high-risk AI systems, without being specific to which technical solutions the provider of the AI must utilise in order to comply with the requirements.

The essential requirements are a function or response to specific and identified hazards inherent to a product, and *in this sense* the requirements in article 9-15 of the AIA is no different. An example from the NLF could be the requirement of protective casings fixed to dangerous machinery in order to address and mitigate the possibility of harm.[110] The obvious difference is the *nature* of the risks.

Examples from the AIA show functions to address *considerably more complex risks*, as the data governance requirement that is meant to ensure the use of high-quality data that is relevant, representative, and absent of *bias*, mitigating the risk of *discriminatory outcomes*.

The transparency and information requirement ensures that the high-risk AI-system's operations and output is clear and understandable to the user, including the system's *capabilities and limitations,* addressing and mitigating the risk of *over-reliance* on the system's output by the deployer or end-user.

Human oversight addresses the risk of AI systems operating without the control of natural persons that can intervene in its operations when necessary, which is especially important the more autonomous the system is, and depending on the system's context, as healthcare or judicial decision making.

While the legislative checks to address these complex issues are in place and mandatory human right impact assessments and compliance with the requirements is required, human rights

[110] Directive 2009/48/EC, Annex I, 3.4.7.

bodies as the European Union Agency for Fundamental Rights (FRA) warns that their experience with mandatory impact assessments, as required by the GDPR, focused mainly on *technical aspects,* and rarely addressed potential impacts on fundamental rights. They add that their research indicate that fundamental rights impact assessments are not carried out when an AI system does not, *or appears not to*, affect fundamental rights negatively.[111]

4.3 Standardisation organisations

The ESOs, as the Comité Européen de Normalisation (CEN), Comité Européen de Normalisation Électrotechnique. [112] (CENELEC), and European Telecommunications Standards Institute (ETSI), acting by qualified majority voting, on a mandate from the commission. [113]

Participants in the standardisation process in regard to the AIA shall seek to promote investment and innovation in AI as well as the competitiveness and growth of the Union market. The stated goal is to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of AI that are consistent with Union values and fundamental rights and interests.

While on one hand cooperating with private sector and stakeholder organisations can greatly benefit the standardisation process with the competence and experience, both technical and commercial, of the market, a counter argument can be made that relying on the cooperation between standardisation bodies, private companies or entities, and stakeholder organisations allows private actors to shape which legal obligations they themselves ought to comply with regarding their AI systems, allowing the risk that companies may define harms to fundamental rights in a way that favour their own interests.[114]

Other perspectives highlight the need for all actors in all nations to be represented in this dialogue. Since there is not a *single* ethical theory that holds *all the time,* similarly, a single set of utilitarian ethical principles with AI would perhaps not be appropriate to address the high

[111] FRA, 2021. p. 6.
[112] See chapter 2.1 above. The "old approach".
[113] Barnard, 2022 p. 589.
[114] Kaminski, 2023. p. 1402-1403.

complexity of our societies. From this perspective, it is desirable for all actors in all nations to be represented in this dialogue. [115]

4.4 Voluntary harmonised standards

The Commission has issued formal standardisation requests to CEN and CENELEC in order to instruct them to start the standardisation process. Among those is the previously mentioned request declaring the urgency for harmonised standards pertaining to the requirements for high-risk AI in articles 9-15 of the AIA.

While these are ongoing, the Commission and stakeholder organisations, together with SMEs and other private companies and entities are developing and adhering to voluntary codes of conduct and voluntary practices regarding the ethical use of AI.

The voluntary codes of conduct fills the gap during the interim period where the harmonised standards are being developed and considering that the actors involved are similar the finished standards, once published in the OJEU, should not bring any surprises to the providers in regard to what they need to accomplish in order to conform and comply with the standards set out by the ESOs. Which means that the providers that already now are demonstrating compliance will not need to change their operation in any substantial way, the conduct and practices they are voluntarily conforming to should then also *in theory* already grant a high level of protection for public interests, in accordance with the NLF model.

However, the risks that these harmonised standards are addressing are not the concrete type of hazards that the NLF would typically develop technical solutions to, and while they do provide means for providers of ai systems to demonstrate conformity, are the standards themselves sufficient?

Using the instruments for protecting individuals from product risks rely on a formalized logic of evaluation different from proportionality that typically guides fundamental rights assessments. Typical product safety instruments, as voluntary standards, aim to keep risk at an

[115] Vinuesa et al., 2020. p. 7-8.

"acceptable" level rather than minimize it, thus attempts to address fundamental rights concerns through a product safety lens can potentially overlook important concerns.[116]

4.5 Presumption of conformity

Products manufactured in conformity with harmonized standards are presumed to conform to the essential requirements established by the relevant NLF legislation for that product category. The means to attain this presumed conformity can in some cases be granted by self-assessment, but in many cases the assessment is made by notified assessment bodies.

The AIA follows this model and obliges the provider for each high-risk AI system they develop to draw up an EU declaration of conformity to in part, demonstrate that all requirements in chapter 9-15 of the AIA, or other relevant harmonisation, is met, and in part to make the provider assume responsibility for the system's compliance. Compliance is also demonstrated by affixing the CE marking to the system, digitally or physically when possible.

When these conditions are met, the AI system enjoys the same free movement within the internal market as products that are presumed to conform with harmonised standards under the NLF, until proven to not conform by market surveillance authorities.

4.6 Market surveillance

The main function of market surveillance is to provide a high level of protection of public interests.[117] In order to provide a *high level of protection*. market surveillance authorities assess the most appropriate action to be taken with a view to the *principle of the proportionality*. Actions taken are based on the *level of the risk*.[118] Market surveillance authorities may take measures in relation to *any risk category of AI system* when they present a *risk* in accordance with the AIA.[119]

Contrary to the concrete risks of the NLF, the risks the market surveillance authorities must contend with are intangible and contextual such as bias, discrimination, or over-reliance on AI-

[116] Almada, petit, 2023. p. 18.
[117] The Blue Guide, 2022. p. 104.
[118] Ibid. p. 46.
[119] Recital 156, AIA.

outputs. The harms of AI are not visible through physical inspection and puts higher expectations on the market surveillance technical competence.

The risks are not however purely technical in nature, and stem from complex and judicial definitions, whose effects most likely only can be ascertained over time, if discovered at all. The competence overlaps from market surveillance authorities monitoring digital products as non-AI software, smart electronics, or medical equipment, might however set a solid foundation for monitoring the post market AI products and their potential human rights impacts. Not because of the technical aspects of these products, but because of the how these products relate to health, (cyber)security, personal data protection, and data processing.

Portraying the NLF risks as *purely concrete hazards* are as time passes becoming an anachronistic statement, but it is however true that the NLF legislation was conceived and implemented to address precisely that type of product safety, and the burning question is how the framework and its surveillance mechanisms adapts to the rapid technological advance the world has seen the last few years in regard to AI.

5 Conclusion

This thesis set out to evaluate to *what extent* the risk-based regulatory toolbox, i.e. *principles of the NLF*, is an *appropriate model* for *protecting fundamental rights in the AIA.*[120]

I will first answer directly, then revisiting my findings to clarify what I mean by my, perhaps *contradictive conclusion*. The answer to the question must be that the classical NLF model *is not* entirely appropriate for taking on such a *complex task* as protecting fundamental rights in the context of the AIA.
However, the question I posed and based this thesis on was in hindsight *naïve* and took the regulation at *face value*. The AIA *is not* a pure NLF "risk-based" legislation, as it *explicitly claims to be*.

With the accommodations that have been made in the AIA to *forbid certain practices*, and the *rigorous checks, obligations, and requirements that providers must meet* in order to put their AI system on the market or into service, with the *cooperation with all manner of actors as*

[120] See above chapter 1.1.1 "What is AI and why does it matter" *in fine*.

stakeholders and experts, the AIA is arguably *more appropriate* than a rights-based top-down legislation in the style of GDPR would have been.

To reach this conclusion I utilised a comparative methodology, evaluating the AIA through the lens of the NLF. Besides the relevant legislative material I have used various materials and literature in order to both contextualise the regulation, and to deepen the conversation regarding the use of risk based regulations, materials such as official publications from the Commission or adjacent bodies, expert groups involved in the drafting of the regulation, and independent scholars who has through articles shared their qualified perspectives regarding the appropriateness of using the NLF product safety framework for matters pertaining to fundamental rights.[121]

I have framed my research into this topic within a risk oriented theoretical framework where I have specified what types of legal risks that are relevant to define in order to analyse the question and the materials, and I have predicted that this view on risk to function as a prescriptive regulatory justification for harmonization legislation, the *foundational driver* that allows for the balancing act of addressing *both;* the safety of public interests by mitigating potential harm to withing an acceptable threshold, *and*; allowing manufacturers or *providers* to put their products on the market under the presumption of conformity.[122]

In order to conduct a comparison, I have presented the NLF, first in a *historical context* and then as the model *functions today* with the additions of regulations and directives that has added to the regulatory toolbox over time. Here many of the important concepts that later will be important for understanding the AIA as a *product safety* regulation has been introduced. The chapter structure followed the principles of the framework, which was "echoed" in subsequent chapters in order to clearly distinguish the potential change in implementation of the NLF principles in the AIA [123]

The object of comparison and the main subject of this thesis has been the AIA. As previously mentioned the subchapter structure followed the principles of the NLF in order to identify changes in implementation, and as expected both similarities and differences were found. In

[121] See above, Chapter 1.2, "Methodology and source material".
[122] Ibid. Chapter 1.3, "Theoretical framework".
[123] Ibid. Chapter 2, "The approach of the new legislative framework"

order to reduce redundancy, the chapter simply presented the AIA in terms of a product under the NLF principles, and the result of the comparison was discussed in a more thorough manner in the following chapter.[124]

My comparison has shown that the implementation of the NLF principles in the AIA has in some cases more or less been inserted unchanged, notably regarding standardisation bodies, the presumption of conformity, and post market surveillance. AI, since it is considered a potentially harmful product will inevitably be under similar limitations and freedoms as other goods where the possibility for harm is inherent in the product, and where that harm is mitigated by essential requirements and harmonised standards. And even if the challenges of post market surveillance are significantly more complex as it pertains to potential fundamental rights impact rather than concrete hazards, the underlying mechanism, and competence of the market surveillance authorities remains largely similar.

Significant differences emerged however in how the NLF toolbox was adapted to the AIA, specifically in the manner the product requirements are presented and enforced; how the actual standards are drafted in cooperation with market actors and stakeholder organisations, and of course; *the nature of the risks themselves*.

The risks covered in the AIA is in contrast to the traditional NLF based product safety legislation systemic and diffuse. Even though the definition that is stated in every NLF regulation or directive entirely mirror the definition found in the AIA, namely that risk is "The probability of harm and, the severity of that harm". The application of this definition to quantify the severity of a societal harm as misinformation, discrimination, or faulty judicial decisions made by low quality datasets, is far more complex than assessing the severity of a physical injury or setting tolerance levels for radioactivity in physical products.

But even the nature of the risks themselves is only part of the problem, the difficulty of determining the occurrence and the severity of the harm certainly brings an added layer of complexity to AI risk management and impact assessments, but the sheer ubiquitousness of AI, with all its complexity makes it such a common commodity that it can neither be ignored nor forbidden. The framing of the requirements in articles 9-15 of the AIA casts a wide net in regard

[124] Ibid, Chapter 3, "The EU AI Act".

to risk mitigation which, *after the relevant checks are in place to achieve a presumption of conformity*. The requirements for high-risk AI systems, drafted by a long cooperation between ESOs, stakeholder organisations and academic experts, mimic the NLF essential "minimal" requirements a product must conform to, but the requirements are not sectorial or based on specific product categories. They are general and covers all perceived risks that the participants in the process predicted or projected, setting a high level of protection against a wide range of risks. But the AIA also uses a precautionary approach and forbids certain practices regardless of potential risk mitigation response. This is one of the accommodations that have been made that shows that the AIA is not entirely a risk-based legislation, it is in practice a hybrid regulation between the bottom-up market incentivising NLF approach, and a top down precautionary and restrictive approach.

The AIA presents a compromise that allows for some low-risk AI practices to be regulated through the GPSR as any other product, while imposing strict requirements for any practice that are deemed to potentially carry certain risks that falls withing the high-risk category. As such it becomes more appropriate that a completely restrictive approach would be, simply because it allows the market a *certain amount* of voluntary compliance while reserving the authority to intervene if necessary.

But why is this desirable? Because just as in the ancient fable, AI has already sprung out of Pandoras box and *can not be put back in again*. Restricting AI too much would restrict trade, innovation, and technological leadership, effectively giving *walkover* of regulatory control to other global actors whose values might not align with the values of the EU, which the AIA is based on. Ignoring AI regulation would inevitably realise all projected risks and harms the AIA is set out to mitigate.

The AIA is not a perfect instrument, but it is a *realistic one.* It does not assume that a risk-based approach alone can protect fundamental rights. By applying *key principles* of the NLF framework to *high-risk AI-systems*, while *maintaining and incorporating stakeholder input*, and *precautionary safeguards* against unacceptable practices, the AIA represents a pragmatic compromise. The AIA realises that AI is not just a market commodity or decision-making tool, but a technology that *for better or worse is here to stay*, and despite its benefits, carry significant risks that must be managed in a way that it both safeguard the fundamental rights of individuals, and allow for innovation and trade within a framework that respect core EU values.

Bibliography

Legislative material

Regulations

Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices.

Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence.

Regulation (EC) 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products.

Regulation (Eu) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation.

Directives

Council Directive 93/68/EEC of 22 July 1993 amending various directives in relation to CE marking.

Directive 2001/95/EC of the European Parliament and of the Council of 3 December 2001 on general product safety (General Product Safety Directive, GPSD).

Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery.

Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys.

Directive 2009/49/EC of the European Parliament and of the Council of 18 June 2009 amending Council Directives 78/660/EEC and 83/349/EEC as regards certain disclosure requirements for medium-sized companies and the obligation to draw up consolidated accounts.

Directive 2014/31/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of non-automatic weighing instruments

Directive 2014/32/EU Of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of measuring instruments (recast)

Decicions

Decision (C(2023)3215) – Standardisation request M/593, Commission Implementing Decision of 22.5.2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence

Decision No. 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products.

Council Resolutions & Statements

Council Resolution 85/C 136/01 on a new approach to technical harmonization and standards.

Working Group Statement on the role of a risk-based approach in data protection legal frameworks

Proposals and drafts

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.

Case law

Case (120/78) Rewe-Zentrale AG mot Bundesmonopolverwalttung fur Branntwein (Cassis de Dijon). EU:C:1979:42.

Literature

AI HLEG. (2019, June 29). Policy and investment recommendations for trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence.

Almada, M., & Petit, N. (2023). The EU AI Act: A medley of product safety and fundamental rights? (RSC Working Paper 2023/59). Robert Schuman Centre for Advanced Studies, European University Institute

Ballor, G. (2021). The origins of CE marking: Standards, business, and the European market in the 1980s–1990s. Harvard Business School Working Paper, No. 21-142.

Barnard, C. (2022). The substantive law of the EU – The four freedoms (7th ed.). Oxford University Press.

Edwards, L. (2022). Regulating AI in Europe: Four problems and four solutions. Ada Lovelace Institute.

European Union Agency for Fundamental Rights. (2021). Getting the future right: Artificial intelligence and fundamental rights.

Kaminski, M. E. (2023). Regulating the risks of AI. Boston University Law Review, 103, 1347.

Mahler, T. (2007). Defining legal risk. Norwegian Research Center for Computers and Law (NRCCL), University of Oslo.

Mahler, T. (2022). Between risk management and proportionality: The risk-based approach in the E.U.'s Artificial Intelligence Act proposal. Nordic Yearbook of Law and Informatics 2020–2021, 247–270.

Nääv, M., & Zamboni, M. (Eds.). (2018). Juridisk metodlära (2nd ed.). Valguarnera.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11, Article 233. https://doi.org/10.1038/s41467-019-14108-y

White Paper on Artificial Intelligence – A European Approach to Excellence and Trust. (2020, February 19). European Commission. http://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Zweigert, K., & Kötz, H. (1998). Introduction to comparative law (3rd rev. ed., T. Weir, Trans.). Clarendon Press.