The EU Artificial Intelligence Act Proposal - Technical Requirements and Viable Solutions for High-Risk AI Systems

Philipp Bohlen, Artur Dox, David Drexlin, Dimitrije Kovacic, Nicolai Pietrzyk, Lennart Schulze
Baden-Wuerttemberg Cooperative State University (DHBW)
Stuttgart, Germany

***Abstract*— Artificial Intelligence (AI), despite its powerful capabilities, poses severe risks to its users when employed in productive context. In response, industry, science, and politics have issued non-binding recommendations for trustworthy AI. In April 2021, the European Commission published the first-ever proposal for a binding regulation of AI systems and their stakeholders with the so-called AI Act. To ease understanding of and compliance with the technical obligations set out therein for providers of AI systems, the following contributions are made: First, formal software requirements are extracted from the proposal in a legal requirements engineering process. Second, available software solutions that assist in fulfilling the requirements are systematically identified. Third, the extent of their support it evaluated through a technical review. In total, 95 requirements were established in eight categories, for which 36 software solutions were identified. The overall requirement fulfillment support score returned low, indicating the need for adapted solutions and manual adoption efforts for developers to achieve compliance with the current version of the act. Issues to address in a revision of the proposal are presented.**

***Index Terms*—Artificial Intelligence, Trustworthiness, European Union, Regulation, Requirements Engineering, Software Solutions, Technical Review.**

I. Introduction

With affordability of data storage and the level of computational power dramatically increasing, Artificial Intelligence has been on the rise for over a decade. Providing the ability to relieve human beings from arduous work, to create new-of-a-kind insights and values, and to solve challenges previously intractable, this technology has grown to great importance in science, economy, and society.

However, Artificial Intelligence has been attributed to pose severe risks from a security, privacy, legal and ethics standpoint. These appear when personal data is processed erroneously, seminal decisions are performed autonomously, a system's behavior cannot be understood by a human being, or outputs were flawed by bias. Especially in sensitive areas such as critical infrastructure, health, public services and administration, law and justice, employment, education, and product safety these threats are severe.

Therefore, technical bodies, think tanks, intergovernmental organizations, and corporate entities have independently issued their recommendation on the development and use of AI technology to achieve trustworthiness. Since 2016, several countries from Europe, America and Asia have been publishing their own Ethical Guidelines on Artificial Intelligence to regulate the technology. As of September 2019, a total of 84 guidelines was published, which could be utilized as guidance [1].

Now, within the scope of the European Commission's theme of A Europe fit for the digital age, the European Union leaps one step ahead by establishing the first-ever binding regulation for risky AI worldwide. The long-awaited proposal from April 2021 aims at defining the rights, obligations, and constraints of the various stakeholders of AI systems in its member states [2].

A. The Artificial Intelligence Act Proposal

The scope of the act is restricted to productive AI systems, explicitly excluding military applications and research. For AI systems on the market or in use by organizations, a risk-based category system is introduced, according to which a system can either pose a prohibitive amount of risk, high risk, limited risk, or no risk. Depending on the assigned category, different measures are foreseen, with the first category being entirely banned and the last not falling subject to any restriction.

In line with this scheme, the proposal, after setting the context and legal basis in the introduction, is divided into 12 titles. Prohibitive AI systems are concerned in title II. For high-risk AI systems, an extensive list of obligations for the technical implementation of these systems, their providers, users, and other parties involved is set out in title III. The remaining titles contain general provisions and definitions of key terms used throughout the act (I), transparency obligations (IV), measures to support AI innovation (V), governance and enforcement mechanisms (VI-X) as well as final provisions (XI) and remarks (XII). Additionally, the annexes to the proposal provide further details referred to throughout the legislation.

This research specifically concerns the requirements set out for high-risk AI systems described in Title III, Chapter 2. This class of AI systems may be considered the dominant objective of the regulation, as they pose severe risks to fundamental rights, the safety of natural persons and the protection of their personal data, while equally offering large benefits [3].

In this regard, the proposal comprehensively addresses the important aspects and relevant stakeholders in the development and use of safe AI. To effectively enforce their compliance with these obligations, the current version of the act imposes a cap of 30 million in penalties, or 6% of yearly turnover for commercial AI providers. Fines of this magnitude have already proven to give strong emphasis to new legal requirements at the time the GDPR was introduced [2].

To comply, a high-risk AI system claimed to adhere by the regulation's obligation must be assessed by an independent authority and registered in an EU-wide database before release into production as well as after every major revision.

Due to the legislative focus of a regulation, however, the proposal is limited in technical details, not clearly laying out the technical requirements nor proposing corresponding solutions for providers of AI systems. In addition, by the mere extend of the regulation, companies and the technical community may lack the resources and capacity to review the entire legal text to identify the implications it has for their AI systems.

To account for this, the proposal itself remarks that specific technical specifications or standards will be required to verify conformity in the future [2]. Their complex agreement process, nonetheless, leaves high-risk AI system providers with uncertainty about the current proposal at hand.

B. Research Objective

This paper aims to analyze the legal requirements for high-risk AI systems set out by the EU proposal for an Artificial Intelligence Act, to identify and to evaluate suitable software solutions to achieve compliance with those requirements.

The overarching objective is to contribute to the joint effort of elaborating a more detailed set of technical requirements along with corresponding software solutions that support providers of AI systems in adapting their systems to comply with the future regulation.

Consequently, to achieve the outlined objectives, the following research questions (RQ) will be answered in this paper:

1) What is the impact of legal requirements established in the EU Artificial Intelligence Act Proposal for the technical implementation of high-risk AI systems?
2) Which currently available software solutions are apt to support the satisfaction of the respective technical requirements in a high-risk AI system?
3) To what extent do these solutions support compliance with the technical requirements, which gaps remain, and what recommendations can be drawn for providers of high-risk AI systems with regard to their employment?

Finally, it is not intended to systematically evaluate the proposal and its quality nor to compare the results with other publications. Neither it is aimed to develop a sample AI system complying with the act. Instead, the findings shall serve as reference asset to the technical community.

II. Background

To position the AI Act Proposal's contribution in the vast field of AI governance, a specification of terms and summary of related publications will be provided first.

A. Terminology

Machine Learning (ML): In a broader sense, ML refers to a computer program that can learn to behave in a way that is not explicitly programmed by the author of the program [4]. In a narrower sense, ML can be defined as computational methods that detect patterns in data and use this information to make accurate predictions [5].

Explainability: In the context of AI, the purpose of eXplainable AI (XAI) is to explain the outputs from AI systems, rendering them more comprehensible to human beings and thus, rendering complex algorithms more transparent [6].

Transparency: An AI model is transparent if it is inherently understandable to human beings on its own [7]. It additionally refers to the need to describe and reproduce the procedures through which an AI system produces a decision [6], which is similar to the aim of explainability.

Interpretability: Interpretability is closely related to Explainability and is defined as the ability to provide explanations that are understandable to humans. In the ML community, interpretability is used more often than explainability [6]. These three terms are customarily used interchangeably [8] [9]. Since explainability, in an academic sense, is defined more concisely, henceforth the term explainability will be used to group the three.

Fairness: Fairness is one of the goals of XAI. An explainable ML model shows how the input leads to a certain results and thus, allows for an analysis of fairness of the given model [10][11]. Explainability can help to avoid an unfair usage of a ML model's output [6].

Trustworthiness: Trustworthiness is regarded as the main purpose of XAI [12][13]. It is considered a confidence measure of whether a ML model will act as expected on a given task. A model that behaves as expected is trustworthy. However, a trustworthy model does not necessarily imply that it can be explained on its own [7].

The EU sets trustworthiness as overarching objective for productively used high-risk AI systems [14]. Therefore, in line with related taxonomies [1], trustworthy AI, henceforth, is used to subsume the terms explainability, safe, robust, and fair AI.

AI System: Software that is developed with one or more of the following techniques: 1) ML and Deep Learning approaches including supervised, unsupervised and reinforcement learning; 2) logic- and knowledge-based approaches and (symbolic) reasoning including expert systems; 3) statistical approaches including Bayesian estimation, search and optimization methods [2].

High-Risk AI System: 1) An AI system that belongs to one of the following areas: biometric identification and categorisation of natural persons, management and operation of critical infrastructure, education and vocational training, employment, workers management and access to self-employment, access to and enjoyment of essential private services and public services and benefits, law enforcement, migration, asylum and border control management, administration of justice and democratic processes. Further details regarding these areas can be found in Annex 3 of the AI Act Proposal. 2) An AI system that is used as a safety component of a product or is itself a product and is required to go through a conformity assessment with the intent to be put on the market, as covered by the Union harmonisation legislation listed in Annex 3 of the AI Act Proposal [15].

Supplementary definitions are provided in title I of the AI Act Proposal and in literature [1].

B. Overview over technical recommendations on trustworthy AI

In the following, a brief overview is provided on existing standards or technical recommendations for trustworthy AI. In 2017, IEEE and the IEEE Standards Association (IEEE SA) published the second version of their seminal document "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems" [16] which provides insights and recommendations, both technical and legal, for the design, development and implementation of ethical autonomous and intelligent systems (A/IS). It was created with input from multiple committees from the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: Ethically Aligned Design. The document provides the following recommendations for implementation: 1) Well-being metrics: Contrary to standard economical metrics, well-being metrics include psychological, social, economic fairness and environmental factors and A/IS should be tested according to these metrics to measure their impact on human well-being. 2) Embedding Values into Autonomous and Intelligent Systems: Norms of the community in which a system is intended to be used in should be embedded in the system itself. 3) Methods to Guide Ethical Research and Design: Developers should use value-based design methods to create sustainable systems. 4) Affective Computing: A/IS that are used in the context of human society should not cause harm by misusing human emotional experience. In total, Ethically Aligned Design provides very high-level recommendations that are rather visionary than practical hands-on advice for developers [16].

The International Organization for Standardization (ISO) currently develops ISO/IEC JTC 1 /SC 42 [17], a standardization for AI. It is part of the standards development environment ISO/IEC JTC 1 on Information Technology [18] and its purpose is to provide guidance to committees from the International Electrotechnical Commission (IEC) and ISO that develop AI applications. The standard covers several aspects, ranging from functional safety and AI systems, bias in AI systems and AI-aided decision-making, and assessment of the robustness of neural networks to quality evaluation guidelines for AI systems. While these standards appear promising and are more detailed than Ethically Aligned Design, they are currently still under development.

C. Overview over regulatory recommendations on trustworthy AI

Overall, there do not exist many standards with technical recommendations for trustworthy AI and more work has been done on regulatory recommendations, which will be discussed in the following. In September 2020, the United Nations Educational, Scientific and Cultural Organization (UNESCO) published a first draft of the Recommendation on the Ethics of AI [19], aimed at providing values and principles on how AI systems should work for the good of humanity, individuals and the environment, and to prevent harm. It also provides policy recommendations, emphasising on gender equality and environment protection.

In February 2020, the Pontifical Academy for Life from the Roman Catholic Church, Microsoft, IBM, the Food and Agriculture Organization of the United Nations (FAO) and the Italian Ministry of Innovation jointly signed a document titled "Call for an AI Ethics" [20], in which they outline six principles to promote ethical AI. These principles include transparency, inclusion, responsibility, impartiality, reliability and security and privacy.

The International Telecommunication Union (ITU) build a digital platform called "AI for Good" with the aim to promote the United Nations Sustainable Development Goals (SDGs) and serve as the United Nations (UN) platform on AI [21]. UN Global Pulse is the initiative of the UN Secretary-General on Big Data and AI for sustainable development, humanity and peace, with the objective to support the development and implementation of Big Data and AI ideas for the public good [22].

The 2020-2021 World Economic Forum (WEF) Global Future Council on Artificial Intelligence for Humanity is currently working on identifying technical solutions to address issues of AI fairness to be able to consult policy makers and organizations [23].

The Organization for Economic Co-operation and Development (OECD) Principles promote that AI shall be innovative, trustworthy, and respecting human rights and democratic values. They were adopted by the OECD member countries in May 2019 [24].

Finally, the purpose of the Ad-hoc Committee on AI (CAHAI) from the Council of Europe (CoE) is to examine feasibility and potential aspects of a legal framework for the development and application of AI, with regard to the standards from the Council of Europe on human rights, democracy and rule of law [25].

The AI Act Proposal examined in this paper has its origins in the establishment of a High-Level Expert Group on AI (HLEG), which consisted of 52 experts in the field, with the aim to advise the European Commission on the implementation of their strategy on AI [2].

In conclusion, while there exist many initiatives on providing recommendations and regulation for AI in similar fields of concern, a unified, binding instrument has been missing. Few provide practical recommendations that can be directly applied by organizations to comply with proposed regulation. Therefore, reaffirming the seminality of the AI Act Proposal, there exists a clear need to accompany it with recommendations for technical solutions.

III. Methodology

Subsequently, the research methodology is outlined in a global perspective followed by the detailed specification of each step contained therein.

A. Overview

While an overall objective is pursued of delivering recommendations to high-risk AI system providers regarding the choice of technical solutions apt to satisfy the AI Act Proposal's regulatory obligations, a tripartite methodology is designed corresponding to the three research objectives. In this approach, first, technical requirements are derived in structured manner from the regulatory obligations contained in the AI Act Proposal. Second, software solutions are identified in the areas covered by the requirements. Third, the solutions' effectiveness with respect to satisfaction of the obligations is assessed. The outcome of each stage is designed to constitute an independent artefact, which shall prove useful to different types of stakeholders in possession of varying levels of capacity to act upon the implications of the Act independently: While the budget-scarce small-sized company may directly start from the delivered final recommendations on software solutions to adjust their AI system for regulatory compliance, the independent AI software architect may observe the technical requirements in a first approach and design their own solution in correspondence. Herein, the selected methods follow accepted academic literature and technical standards, amended for the specifics of the AI Act Proposal.

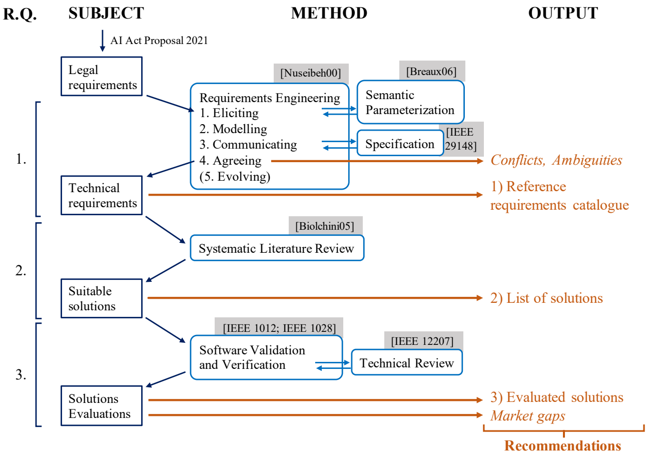

Fig. 1. Overview of the research design

To arrive at recommendations for the usage of specific technical software solutions that support the fulfillment of the requirements set out in the AI Act Proposal, a tripartite approach will be employed. First, requirements engineering [26] from the domain of software engineering allows to produce a set of technical requirements in a generalized five-step process. To render this method applicable to the first research objective its first step is substituted with a proposal by [27]. It sets out a formalised way to translate legal text, accounting for its special properties, into unambiguous demands. To formalize the requirements modeled based on them, a Software Requirements Specification according to the IEEE Standard 29148-2011 on Systems and Software Engineering [28] is produced, acknowledging its conciseness, formality, and acceptance in industry. In analysis for acceptance of the defined requirements, constituting the fourth step, overlaps, ambiguities and conflicts can be identified, accountable to shortcomes in the AI Act Proposal.

Second, to identify the most customary software solutions that principally may serve the fulfillment of the AI Act Proposal, a systematic literature review is conducted based on the technical requirements from before. The parameterized method by [29], , equally leveraging five steps, is targeted for the domain of software engineering. It is chosen due to its linear approach, allowing to control the relevance of the information extracted.

Finally, to evaluate the extent of the requirement support introduced by the capabilities of the software solution, the most scholarly established ones among the identified solutions are examined. The process of software verification, forming part of the conventional software life cycle [30], allows to verify the conformance of a software artefact with its technical requirements. It foresees the assessment of the remaining processes in the software life cycle, as governed in [31]. From different applicable approaches to realize that assessment, a limited technical review, as defined in [32], proved to possess the most favorable cost-effect ratio. There, the subject of evaluation is set to an AI system that employs the software solution of concern; the solution as an artefact itself is not verified.

From an aggregated requirement fulfillment support score computed per solution and a qualitative evaluation based thereon, final recommendations will be drawn regarding the usage of the solutions to comply with the AI Act Proposal. In this process, unaddressed requirements will be reported as potential gaps in the AI system software market.

B. Research Question 1

The approach by [26] synthesizes five steps of requirement engineering from previous literature in this field: eliciting demands, their modelling as requirements, their formal definition, and their discussion and approval, followed by their continual maintenance.

While requirements customarily are elicited from stakeholder demands subject to conflicts and negotiations, regulatory compliance is required by law. Thus, to extract the intrinsic requirements from legal text, intermediate approaches are proposed [27], [33], [34], [35], [36], [37], [38], [39]. From the two major approaches [27] and [38], both translating the legal text into intermediate representations before formally analyzing it, the former employs restricted language modelling and the latter a visual mapping in this step. Since the AI Act Proposal enforces plentiful obligations on a multitude of actors, it exceeds the capacity of visual mapping before becoming too convoluted, which is why [27] provides an appropriate approach, denoted as Semantic Parameterization.

As precondition, the articles of the EU AI Act Proposal with immediate relevance for the technical design of high-risk AI systems are identified. From these, each paragraph, amended with potential cross-references to other parts of the Act included therein, is transformed to restricted language statements in case of language ambiguity in the original text. Depending on the linguistic character of these statements, such as the choice of verbs or use of subordinate clauses, formal obligations, rights, and constraints are extracted, the last governing the applicability of the two previous. The resulting system is interconnected through references.

In the modelling step, obligations are transformed into technical requirements according to an $m : n$ scheme, that is one obligation may yield multiple requirements and one requirement may be derived from multiple obligations. Constraints, depending on their nature, result either as part of the content of the requirement or of its applicability description. Hierarchical relations between requirements, similarly, are captured in the applicability.

As third step, the requirements are formalized as field-value pairs according to [28]. In addition to the proposed information content, type, rationale, and difficulty, the fields origin, fit criterion, applicability, and category are introduced in response to the use of the Semantic Parameterization and subsequent methods. The *fit criterion* is defined as a negatable condition entailed by the state of an AI system that fulfills the requirement. The *requirement type*, as is customarily, classifies requirements into functional, and non-functional requirements for the software artefact, and process requirements regarding its interaction with different stakeholders in its life cycle. The *rationale* explains the underlying reasoning of a requirement. The three-class *difficulty* provides an indication of the realization effort to fulfill the requirement, derived from the complexity of its content.

Finally, as part of the agreement step, each requirement is cross-checked against the set of all other requirements. Conflicts, if not solvable, as well as insufficiently precise requirements are reported as conflicts and ambiguities, respectively, that are inherent to the AI Act Proposal. The maintenance of the requirements, as iterative fifth step, is subject to updates to the AI Act Proposal or from the involved stakeholders, and is thus inapplicable.

C. Research Question 2

The final requirements specification serves as input to the systematic literature review. For this purpose, the requirements are grouped based on their content and for each group, suitable frameworks in the domain of artificial intelligence that are apt to contribute to the fulfillment of one or more of the corresponding requirements are identified.

The proposal by [29] erects five steps for a systematic review in software engineering: 1) question formularization, 2) sources selection, 3) studies selection, 4) information extraction, and 5) results summarization. First, within question formularization, the review objective, review research questions, the review approach and the measurement of the outcomes are defined. Second, applicable sources, such as publishers or conferences, in which studies shall be searched are fixed. Third, the studies, that is the publications, are selected based on inclusion and exclusion criteria. Fourth, from these, the relevant information is extracted. Based on these findings, fifth, the results are summarized in section IV.

These five steps ought to be performed within three stages: The first and the second as well as part of the third constitute the review planning. Part of the third and the fourth form the review execution. The fifth step corresponds to the results analysis stage. Review planning and review execution in turn shall be followed by an evaluation of the results from that stage, respectively.

Important parameters of review planning and evaluation and review execution are captured in a review protocol, illustrated in table I.

The result of the review consists of a set of software solutions per category, joint with an academic search results metric used as heuristic for the relevance of the solution in the scientific AI community. These outcomes will be used further in the last part of the methodology.

D. Research Question 3

To assess the conformance of a software product with some specifications, standards, or requirements software verification and validation is eligible, defined as integrated constituent of the software life cycle in [30]. Therein, tests during the implementation stage of the artefact, such as qualification and acceptance tests, may be distinguished from holistic a-posteriori approaches, such as reviews and audits [40], which assess all other life cycle stages [31].

Since qualification tests shall be performed by developers [30], acceptance tests are targeted at the acquirer of a software product [30], and audits shall be performed by independent authorities [30] - such as *notified bodies* envisioned in the AI Act Proposal [2] - technical reviews are deemed appropriate to evaluate the capacity of the software solution with respect to requirements [30].

The employed method is based on the technical review process specified in [32], from which a five-step approach towards examining software artefacts is derived: 1) provisioning of input material for the review, 2) validation that the entry criterion for the review is satisfied, 3) the software examination itself, 4) validation that the exit criterion is satisfied, and 5) output production. The objective of the review is to quantify the aptness of selected software solutions to satisfy the requirements engineered from the AI Act Proposal when used in a high-risk AI system through a manual analysis of its functioning. Thus, the solutions themselves will not be assessed for compliance with the requirements, but for their ability to support their achievement in an integrated system. There, the software product that is subject of the review is a generalized high-risk AI system that employs the respective software solution of concern, which will be evaluated against the set of applicable requirements.

For each requirement group, the three identified software solutions with the highest score for the academic relevance metric will be selected for assessment. For each solution, the following process is performed:

1) Input provisioning: The review objective as defined above, the requirements specification from subsection III-C, this procedure guidance, and the software product are provided. The last is restricted to technical documentation and complementary literature and artefacts, not however access to a running instance of the solution, owing to resource constraints.

TABLE I
SYSTEMATIC LITERATURE REVIEW PROTOCOL (EXCERPT)

Step	Value
PLANNING	
1. Question formularization	
1.1. Question focus	Identify relevant software solutions and related artefacts that could support the satisfaction of at least one technical requirement in a high-risk AI system
1.2. Question quality and amplitude	
1.2.1. Problem	Within the research landscape of ethical AI and a plethora of recommendations for AI system requirements, it is difficult to understand which software solutions are effective in satisfying the first binding requirements defined in the AI Act Proposal
1.2.2. Research question	Which published software solutions, programs, frameworks, tools, packages, or libraries could support fulfillment of at least one technical requirement from either category of the AI Act Proposal?
1.2.4. Intervention	Evaluation of software introductions, publications, reviews, comparisons, and overviews
1.2.5. Control	Reviewers’ knowledge of related software solutions and their acceptance in the community
1.2.6. Effect	Set of software solutions and their relevance
1.2.7. Outcome measure	# of identified software solutions and # of search results for each framework from scholar.google.com
1.2.9. Application	Software developers of AI systems (AI Act Proposal)
2. Sources selection	
2.1. Criteria definition	Relevance for AI system developers or researchers AND ability to search through publications AND (conference OR journal OR publisher OR institution publications series)
2.3. Identification	
2.3.1. Search methods	Web search engine, sources web page search engine
2.3.2. Search string	(’AI’ OR ’Artificial Intelligence’ OR ’Machine Learning’) AND (’solution’ OR ’software’ OR ’framework’ OR ’approach’ OR ’program’ OR ’algorithm’ OR ’procedure’ OR ’library’ OR ’package’) AND [REQ. CAT. NAME INCL. VARIATIONS]
2.3.3. Sources list	IEEE, ACM, NIPS, ACM SIGMOD, JMLR, Arxiv.org, Springer, Researchgate, Github.com, Stackoverflow.com, Gartner.com, SAS Publishers, Rheinwerk Verlag, Proceedings of International Conference on Machine Intelligence and Data Science Applications, IBM J. Res. Dev.
3. Studies selection	
3.1. Studies definition	
3.1.1 Inclusion and exclusion criteria definition	Includes reference to relevant software solution published by author or company AND NOT includes references to beta versions or unpublished software
3.1.2 Studies types definition	Paper, proceedings, technical reports, webpages, GitHub repositories, forum hyperlink references
3.1.3 Procedures for studies selection	1) Use 2.3.2 to search sources 2) Include studies that meet 3.1.1 criteria 3) Analyze selected study and extract information on software solutions in format of 4.2 4) Retrieve no. of academic search results for the identified solution on scholar.google.com by searching for [Solution name] + [’AI’, if name does not contain explicit AI reference]
PLANNING EVALUATION	The protocol was iteratively executed with subset of sources and refined in response to recognized issues.
EXECUTION	
4. Information extraction	
4.2. Data extraction form	Solution name, category, description, publisher, academic publication, # scholarly search results

To allow for a thorough assessment nonetheless, complementary literature and artefacts can comprise of software development and architecture descriptions, maintenance manuals, release notes, source code repositories, marketing material, and user question and answer protocols, each retrieved from the original solution publisher or trusted sources.

2) Entry criterion validation: Technical documentation and complementary literature and artefacts, if necessary, are available in sufficient number, extent, and depth. Sufficiency is defined as the reviewer being able, in a preliminary assessment, to maintain that all applicable requirements can be assessed according to this procedure only from the provided information, or that a lack of information is objective evidence of failure to support the requirement.

3) Examination Procedure: Per requirement to assess against, the available technical documentation and complementary literature and artefacts are searched for relevant information. From evidence regarding the functionality and non-functional properties such as architecture, interoperability, operational or maintenance conditions, the reviewer establishes the extent to which a solution supports a high-risk AI system’s compliance with the requirement along four levels.

- 0 - No support. Integration of the solution does not contribute to satisfying the requirement.
- 1 - Limited support. Integration of the solution partially contributes to satisfying the requirement but considerable effort remains to fulfill it.
- 2 - Moderate support. Integration of the solution contributes to satisfying the requirement but some effort remains to fulfill it.
- 3 - Extensive support. Integration of the solution substantially contributes to satisfying the requirement, leaving no or minimal effort to fulfill it.

Here, effort refers to the delta between the contribution of the software solution and the target state of the fulfilled requirement, which is provided by a high-risk AI system that can fulfill the fit criterion of the requirement. This delta can be closed with manual development or administration activities or with further software solutions.

4) Exit criterion validation: All requirements pertaining to the category were assigned level 1-3 or conclusively assigned level 0.

5) Output production: The evaluations per requirement are stored. In addition, for each solution a level-weighted requirement fulfillment support score is computed as

$$\sum_{r=1}^{r_{max}} \frac{l_r}{3} \times \frac{1}{r_{max}} \tag{1}$$

where r is the integer requirement ID, r_{max} is the number of requirements to consider, usually the number of applicable requirements of the category, and l_r is the fulfillment support level assigned for the requirement with ID r. Thus, the score returns the portion of requirements whose satisfaction in a high-risk AI system is fully supported when employing the software solution, where 100% equates to all requirements being evaluated as level 3.

Concluding the methodology, a qualitative analysis of the quantitative results from the third step allows to achieve the overall research objective. Thereby, recommendations for an effective use of software solutions to comply with the AI Act Proposal's technical obligations are pronounced and gaps and ambiguity-induced uncertainties that should be considered are pointed out.

IV. Results

In line with the research objective, the results from execution of the research design will be portrayed in order of the research questions.

A. Requirement Engineering

After an initial analysis of the proposal, Articles 9 to 15, in *Title III*, *Chapter 2 - Requirements for High-Risk AI systems*, were classified as relevant as they contain immediate technical obligations for high-risk AI systems and specify the conditions they must satisfy. Hence, the clauses pertaining to this chapter, which were found to be linguistically unambiguous, will be used as basis for the Semantic Parameterization. Each article, thus, produces a set of obligations, requirements, and constraints, constituting the eliciting step of the requirement engineering process.

To demonstrate how the articles in the legal text were transformed into finished requirements, the engineering of one requirement is examined in table II and table III.

Table II depicts the erection of obligations, rights, and constraints from analysis of the original legal text. There, the verb indicating whether the sentence yields an obligation or a right is highlighted in **bold and underlined**, the details about an obligation are formatted **bold**, and details about a constraint, governing the applicability of the obligation, are formatted *italic*. In this case, the requirement arises from two paragraphs in article 9. 'Shall', in the legal sense, implies an obligation (O) for the high-risk AI system, for which reason both are transformed into such, respectively. While the content of art. 9 (5), only specifies the content of the obligation, art. 9 (2d) additionally conditions the scope of its corresponding obligation, normally translated into a constraint. However, as the content of the constraint is superfluous in light of the additional requirements arising from the remaining paragraphs, it was not employed as such to restrict O9.5.

Table III subsequently shows the result of analysis of the two obligations to arrive at a requirement. The *description* as content of the requirement introduces the obligation to test the system. The directly deducible, subjective motivation for a testing procedure, next to the articles requiring it regardless of consent to it, is presented in the *rationale*. Because the existence of technical test routines is a technical requirement compared to an organizational one, but one with no functionality for the user of the system, the *type* is set to non-functional. Since testing is obligatory in any software project, the additional workload is minimal, rendering the *difficulty* low. The *applicability*, besides applying the requirement to all types of high-risk AI systems defined in the AI Act Proposal, conditions the requirement on the existence of a risk management system in the system, which is defined in another requirement. Finally, the *fit criterion* specifies a scenario resulting from a system that implements the test routine with the specified purpose, which can be probed to assess conformance in a later stage.

TABLE II
Sample Requirement: Semantic Parameterization

Art. 9 (2) d: The risk management system [...] **shall comprise the following steps: [...] adoption of suitable risk management measures** *in accordance with the provisions of the following paragraphs*	Art. 9 (5): High-risk AI systems **shall be tested for the purposes of identifying the most appropriate risk management measures**. [...]
↓	↓
O9.5: The risk management system comprises of suitable risk management measures	O9.15: To identify the risk management measures, the high-risk AI system is tested

TABLE III
Sample Requirement: Requirement Specification

ID	9.14
Origin	O9.5, O9.15
Description	The high-risk AI system shall be tested with the purpose of identifying appropriate risk management measures.
Rationale	Art. 9 (2)(d), Art. 9 (5); Testing a high-risk AI system reveals the risks associated with its use that are hard to expect or predict.
Difficulty	low
Fit Criterion	The risk management measures adopted in the finalised risk management system were informed by the results of a technical testing procedure performed on the high-risk AI system.
Type	Non-Functional Requirement
Applicability	All (high-risk AI systems), given req. 9.1 is fulfilled
Category	Testing

Following this scheme, a total of 95 requirements was erected from the seven articles. The total number of obligations, rights, and constraints extracted from each article is shown in table IV.

TABLE IV
Number of obligations, rights, and constraints erected from relevant articles of the AI Act Proposal

Article	**# Obligations**	**# Rights**	**# Constraints**
9 - Risk management system	22	-	7
10 - Data and data governance	14	2	12
11 - Technical documentation	22	1	1
12 - Record-keeping	8	-	7
13 - Transparency and provision of information to users	3	-	8
14 - Human oversight	10	-	6
15 - Accuracy, robustness and cybersecurity	10	-	2

Based on the requirements' content, regarding the aspect of the AI system they address, and on their origin among the articles, each requirement was assigned to one of eight categories, which are shown in table V.

TABLE V
REQUIREMENT CATEGORIES

Category	Description	# Req.
Risk Management	predominantly process or functional requirements regarding the implementation of risk mitigation procedures	15
Testing	mainly non-functional requirements concerning testing routines and procedures in the high-risk AI system's life cycle	5
Dataset Properties	mostly non-functional requirements addressing the quality and content of training, validation and test sets put into the system	10
Technical Documentation	predominantly non-functional and process requirements regarding the scope of the information about the system included	23
Record Keeping	predominantly functional requirements regarding the logging of system behavior and access to these	11
Explainability	mostly process requirements on the transparency of operations of the system and the content of instructions of use	10
Human Oversight	requirements concerning interfaces and procedures for human beings to control the operation of the system	9
Accuracy, Robustness, Cybersecurity	process and non-functional requirements mitigating the proneness of the system to errors	12

The exhaustive requirement specification can be found in Appendix I.

B. Software Solution Identification

Based on the requirement categories, software solutions with the potential to support the fulfillment of the requirements were systematically searched. From a technical standpoint, the two categories *Risk Management* and *Technical Documentation* did not yield any requirements specific to AI systems compared to general IT systems. Since numerous reviews and market analyses are available in these domains, they were excluded from further research.

To demonstrate the review findings on the solution-level, in table VI, the resultant software solutions for the category *Accuracy, Robustness, Cybersecurity* are portrayed. While no AI-specific cybersecurity solution was identified, five were assessed to be relevant for the robustness- and accuracy-related requirements. Out of these, the three with the highest academic relevance score, Foolbox Native, IBM Adversarial Robustness Toolbox, and IBM CNN-Cert were selected for evaluation.

In total, 36 unique software solutions were identified. Among these, individual solutions were returned for several categories, such as Local Interpretable Model-Agnostic Explanation (LIME) (Explainability; Human Oversight), Neptune.ai (Record Keeping; Human Oversight), RuleX AI (Explainability; Human Oversight), and SHapley Additive exPlanations (SHAP) (Explainability; Human Oversight). In addition software suites contained individual tools that were relevant for different categories, such as Amazon Sage Maker (Testing; Dataset Properties; Record Keeping; Human Oversight), IBM Research Trustworthy AI 360 Toolkit (Explainability; Human Oversight; Accuracy, Robustness, Cybersecurity), IBM Watson (Testing; Dataset Properties), and Tensorflow (Dataset Properties; Human Oversight).

TABLE VI
SOFTWARE SOLUTIONS FOR SAMPLE CATEGORY *Accuracy, Robustness, Cybersecurity*

Name	Publisher	Original Publication	# Academic Search Results
CORTEX CERTIFAI	CognitiveScale	[41]	59
Foolbox Native	Rauber, J.	[42]	498
IBM Adversarial Robustness Toolbox	IBM	[43]	305
IBM CNN-Cert	IBM	[44]	85
IBM Research AI Fairness 360 Toolkit	IBM	[45]	48

The distribution over the categories, including multi-category solutions, is depicted in table VII.

TABLE VII
NUMBER OF SOFTWARE SOLUTIONS PER CATEGORY

Category	# Identified Software Solutions
Testing	3
Dataset Properties	11
Record Keeping	7
Explainability	5
Human Oversight	9
Accuracy, Robustness, Cybersecurity	5

C. Software Solution Evaluation

Finally, the three software solutions per category with the highest number of academic search results were evaluated for their aptness to satisfy the category requirements when employed in high-risk AI system. As process requirements necessitate organizational effort, they were excluded from the technical review.

Continuing the sample from subsection IV-B, table VIII shows the evaluations per requirement for the highest-relevance software solution in category *Accuracy, Robustness, Cybersecurity*: Foolbox Native. The given explanations show why different levels of fulfillment support were assigned, referencing the evidence that provided the underlying information. Out of the nine applicable requirements, using Foolbox Native would at least partially facilitate the fulfillment of seven. The level-weighted requirement fulfillment support score computes to 56%.

In table IX, the portion of applicable category requirements by evaluation level is provided for the three selected software solutions of each category along with their weighted aggregated requirement fulfillment support scores. From the 95 original requirements across eight categories, 37 across six categories were applicable. The overall rounded mean requirement fulfillment support score over all categories is 34%. On the category level, the decreasing rounded mean scores are 78% for Explainability, 46% for Accuracy, Robustness, Cybersecurity 42% for Testing, 37% for Dataset Properties, 26% for Human Oversight, and 23% for Record Keeping.

TABLE VIII
EVALUATION OF SAMPLE SOFTWARE SOLUTION *Foolbox Native* IN REQUIREMENT CATEGORY *Accuracy, Robustness, Cybersecurity*

Req. Id	**Level**	**Explanation**	**Evidence**
15.1	*N/A*	*Process requirement*	*N/A*
15.2	1	Foolbox provides attack models for adversarial training. There is a trade-off between robustness ('robust accuracy') and accuracy ('standard accuracy'). A consistent level of robustness through should lead to a consistent level of accuracy.	[46]
15.3	3	Foolbox provides a variety of adversarial attacks to benchmark the robustness of machine learning models.	[47]
15.4	2	Foolbox provides adversarial training, which helps mitigating adversarial attacks, but is not sufficient to achieve cybersecurity as a whole.	[48]
15.5	*N/A*	*Process requirement*	*N/A*
15.6	0	Foolbox provides adversarial training, but does not address technical redundancy or fault prevention.	[46]
15.7	0	Foolbox provides adversarial training, but does not address biased outputs through 'feedback loops'.	[46]
15.8	2	Adversarial training mitigates adversarial attacks, being a popular way of AI-System manipulation, but does not generally prevent unauthorized access by third parties.	[46]
15.9	*N/A*	*Process requirement*	*N/A*
15.10	2	Data poisoning is considered a specific strategy of adversarial attacks, which are addressed by the framework.	[47]
15.11	3	Adversarial examples are considered a specific strategy of adversarial attacks that are explicitly addressed by the framework.	[47]
15.12	2	Model flaw exploitation is considered a specific strategy of adversarial attacks, which are addressed by the framework.	[47]

TABLE IX
OVERVIEW OF EVALUATION OF SOFTWARE SOLUTIONS PER REQUIREMENT CATEGORY

Category	**Software Solution**	**Level 0**	**Level 1/2/3**	**Score**
Testing (4 req.)	Amazon Sage Maker	25%	50%/25%/0%	33%
	Watson OpenScale	0%	25%/75%/0%	58%
	Azure ML	25%	50%/25%/0%	33%
Dataset Properties (7 req.)	IBM SPSS Modeler	29%	29%/14%/29%	48%
	SAP Data Services	29%	43%/14%/14%	38%
	Informatica Data Quality	57%	14%/29%/0%	24%
Record Keeping (10 req.)	TensorBoard	50%	30%/20%/0%	23%
	Amazon CloudWatch	50%	0%/40%/10%	37%
	DataDog	80%	10%/10%/0%	10%
Explainability (1 req.)	SHapley Additive exPlanations (SHAP)	0%	0%/100%/0%	67%
	Local Interpretable Model-Agnostic Explanation (LIME)	0%	0%/100%/0%	67%
	IBM AIX360 Toolkit	0%	0%/0%/100%	100%
Human Oversight (6 req.)	SHapley Additive exPlanations (SHAP)	67%	17%/0%/17%	22%
	Local Interpretable Model-Agnostic Explanation (LIME)	67%	17%/17%/0%	17%
	MLflow	50%	0%/33%/17%	39%
Accuracy, Robustness, Cybersecurity (9 req.)	Foolbox Native	22%	11%/44%/22%	56%
	IBM Adversarial Robustness Toolbox	22%	11%/44%/22%	56%
	IBM CNN-Cert	33%	56%/11%/0%	26%

In the case of *Accuracy, Robustness, Cybersecurity*, it is recommended to employ either Foolbox Native or IBM Adversarial Robustness Toolbox in the high-risk AI system as their functionality is similar, each achieving a fulfillment support score of 55%. However, their requirement coverage is not complementary, rendering the use of both simultaneously superfluous. Part of the uncovered requirements are those that go beyond the AI-specifics robustness and explainability and instead include traditional security aspects. To fulfill these, it should be attempted to use conventional IT security practices and solutions, jointly with the novel AI-specific solutions.

Similarly, examining the assessed software solutions' individual explanations and level assignments per requirement demonstrates which solutions harmonize satisfactorily and which requirements remain entirely uncovered in each category. Thereby, recommendations on how to most effectively comply with the AI Act Proposal using established software solutions in high-risk AI systems are provided.

V. DISCUSSION

Overall, the results of the last step within the research process show that there are various technical solutions and frameworks which can be considered useful to comply with the proposed regulation on AI. Nevertheless, the individual evaluation scores indicate that few requirements and categories can be fully covered by the identified software solutions. In fact, 11 of the total 95 defined requirements were consistently evaluated as level zero, meaning their fulfilment cannot be supported by implementing the considered frameworks at all. This limitation of results can be attributed to the following factors which have become apparent in the course of the research:

During the analysis of the AI Act Proposal and the process of deriving technical requirements, various shortcomings in the level of detail have been identified. This impeded the derivation of clear technical implications for RQ1. An overview of vague or ambiguous terms and phrases has been composed and is provided in appendix IV. Without a clear definition of, for instance, what measures are considered in accordance with "recognised standards" (art. 12 (1)) or what level of transparency towards the user is "sufficient" (art. 13 (1)), the fulfilment of requirements containing such ambiguities can only be evaluated on a high level. In some cases, this has led to the respective requirement being evaluated as level zero.

The analyzed chapter is divided into 15 articles. Among these articles several overlaps and dependencies have been identified. While explainability as defined in subsection II-A can be considered a key aspect of trustworthy AI systems [6][7], it is not explicitly mentioned in the legislative text. Instead, the concept of explainable AI appears to be covered by multiple articles, such as ”Human Oversight” (art. 14), ”Transparency” (art. 13) and ”Record Keeping” (art. 12). These interdependencies rendered it difficult to define useful and distinguishable categories in preparation for RQ2 as described in subsection III-C. As a result, the categories and the respective names do not represent every individual requirement in the same way, returning software solutions with limited coverage in the systematic review process.

Other requirements or sets of requirements could not be covered by specific solutions due to their content being process-oriented or not specific to the AI systems special characteristic. Such process requirements need to be addressed by adequate management and governance methodologies (e.g., ”Risk Management System”, ”Technical Documentation”). The gap in corresponding software solutions also extends to sets of requirements not considered AI-oriented in the first place: Especially in the fields of ”Testing”, ”Record Keeping”, and some traditional IT-Security aspects, only few AI-specific technical solutions were found as result from the systematic review. This may indicate a demand for stronger synergies between AI-specific and general software engineering in non-functional software areas. End-to-end ML platforms address several aspects of the ML development cycle, including important non-functional aspects, and therefore, are able to cover more requirements than task-specific solutions.

Not only with regards to the level of detail of the legislative text, but also of the information and documentation of some technical solutions, limitations have become apparent. As the systematic review described in subsection III-C included both open-source as well as proprietary software solutions, the quality of sources available to comprehend their functionality varied widely. For a practical, detailed analysis of the requirements’ fulfilment, each solution would be required to be employed in the specific AI system for individual reviews in addition to technical tests. In some categories, software solutions are only applicable for specific types of ML models and data. For instance, IBM CNN-Cert is designed exclusively for certifying the robustness of Convolutional Neural Networks (CNNs), not any other neural network and ML models. While useful in the targeted cases, often reflecting technical development trends in the AI landscape, this limits the applicability of such solutions.

In addition to technology restrictions, most of the solutions solely address a certain functional or non-functional aspect, even within the assigned category, which would require combination with other solutions or manual implementation efforts to fulfill all given requirements.

Despite the limitations outlined, the results at hand are a useful foundation and guidance to understand the technical implications of the AI Act Proposal in the applicable areas and categories. For other categories, research demonstrated that further elaboration on the proposal itself, as well as case-specific evaluation for different applications and fields of AI will be necessary.

Finally, it should be noted that the scope of the act is substantially larger than the definition of obligations for the high-risk AI system itself. Only taking into consideration the rights and obligations of users, authorities, and other stakeholders will allow to estimate the total effort for AI system providers to comply with this law.

VI. Conclusion and Future Work

The main objective of this work was to analyze the legal obligations set out by the European Commission’s proposal for an Artificial Intelligence Act for their technological impact on high-risk AI systems in order to identify and evaluate technical solutions that assist in achieving compliance with these requirements.

As a result, an extensive set of 95 requirements has been derived from the legislative text along with an overview of ambiguous and vague terms or phrases which require specification in a revision of the draft. A list of 36 potentially suitable software solutions has been composed through a systematic review based on six technically relevant requirement categories. For each category, the three most scholarly mentioned solutions have been selected to evaluate their suitability to support compliance with the regulation when implemented in a specific AI system. For the majority of requirement categories, the mean requirement fulfillment scores is below 50%, indicating a considerable gap between current established solutions in the market and the scope of the AI Act Proposal. If unmet, the AI Act Proposal, irrespective of the appropriateness of its measures, may require a large technical effort for high-risk AI system providers to comply.

The results of this work can be considered a contribution to the joint effort of elaborating a technical specification, derived from the AI Act Proposal, which is explicitly envisioned and encouraged by the EU Commission [2]. As is the nature of a legislative proposal, the AI Act Proposal has drawn various criticism regarding some of its crucial aspects from several parties and stakeholders [49][50][51]. The research for this work has revealed some of those shortcomings, regarding lack of technical detail, interdependencies and ambiguities, and therefore confirmed part of the criticism. When revising the proposal to arrive at a final regulation, these aspects needs to be addressed thoroughly.

Until then, this work could potentially prove useful to the technical AI community in preparing for the binding impact of the regulation. The full results are available at [52] where it is sought to maintain and extend the requirements, software solutions, and evaluations as the legislative process progresses. For this purpose, contributions are highly welcomed. On the way to trustworthy AI, the technological feasibility of international regulations will be crucial to leverage the high potential of AI in a safe, ethical, and human-centered manner.

References

[1] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of ai ethics guidelines,” *Nature Machine Intelligence*, vol. 1, pp. 389–399, 2019.

[2] “Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence and amending certain union legislative acts,” European Commission, 2021.

[3] J. Wolff and N. Atallah, "Early gdpr penalties: Analysis of implementation and fines through may 2020," *Journal of Information Policy*, vol. 11, pp. 63–103, 2021.

[4] A. V. Joshi, *Machine learning and artificial intelligence*. Springer, 2020.

[5] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[6] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (xai)," *IEEE Access*, vol. 6, 2018.

[7] A. B. Arrieta, N. Daz-Rodrguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information Fusion*, vol. 58, 6 2020.

[8] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 7 2017, pp. 1885–1894.

[9] M. Bojarski, P. Yeres, A. Choromanska, K. Choromanski, B. Firner, L. Jackel, and U. Muller, "Explaining how a deep neural network trained with end-to-end learning steers a car," *CoRR*, 4 2017.

[10] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision-making and a right to explanation," *AI Magazine*, vol. 38, 10 2017.

[11] A. Chouldechova, "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments," *Big Data*, vol. 5, 6 2017.

[12] B. Kim, E. Glassman, B. Johnson, and J. Shah, "ibcm: Interactive bayesian case model empowering humans via intuitive interaction," *CSAIL Technical Reports*, 2015.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you? explaining the predictions of any classifier," *Tatra Mountains Mathematical Publications 74*, 2016.

[14] European Commission, "Excellence and trust in artificial intelligence," 2021. [Online]. Available: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en#building-trust-through-the-first-ever-legal-framework-on-ai

[15] "Annexes to the proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts," European Commission, 2021.

[16] K. Shahriari and M. Shahriari, "Ieee standard review ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems," in *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*. IEEE, 7 2017.

[17] ISO. Iso/iec jtc 1/sc 42 artificial intelligence. [Online]. Available: https://www.iso.org/committee/6794475.html

[18] ——. Iso/iec jtc 1 information technology. [Online]. Available: https://www.iso.org/isoiec-jtc-1.html

[19] UNESCO. (2020) Outcome document: first draft of the recommendation on the ethics of artificial intelligence. Ad Hoc Expert Group for the Preparation of a Draft text of a Recommendation the Ethics of Artificial Intelligence.

[20] Romecall. (2021) Rome call for ai ethics a human-centric artificial intelligence. [Online]. Available: www.romecall.org/

[21] ITU. (2021) Ai for good. [Online]. Available: https://aiforgood.itu.int

[22] UN Global Pulse. (2021) Expert group on governance of data and ai. [Online]. Available: https://www.unglobalpulse.org/policy/expert-group-on-governance-of-data-and-ai/

[23] World Economic Forum. (2021) Global future council on artificial intelligence for humanity. [Online]. Available: https://es.weforum.org/communities/gfc-on-artificial-intelligence-for-humanity

[24] OECD. Oecd principles on ai. [Online]. Available: https://www.oecd.org/going-digital/ai/principles

[25] Council of Europe. (2021) Cahai - ad hoc committee on artificial intelligence. [Online]. Available: https://www.coe.int/en/web/artificial-intelligence/cahai

[26] B. Nuseibeh and S. Easterbrook, "Requirements engineering: a roadmap," in *Proceedings of the Conference on the Future of Software Engineering*, 2000, pp. 35–46.

[27] T. D. Breaux, M. W. Vail, and A. I. Anton, "Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations," in *14th IEEE International Requirements Engineering Conference (RE'06)*, 2006, pp. 49–58.

[28] *ISO/IEC/IEEE International Standard - Systems and software engineering – Life cycle processes – Requirements engineering*, ISO/IEC/IEEE Std. 29148-2018, 2018.

[29] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos, "Systematic review in software engineering," *System Engineering and Computer Science Department COPPE/UFRJ, Technical techreport ES*, vol. 679, p. 45, 2005.

[30] *ISO/IEC/IEEE International Standard - Systems and software engineering – Software life cycle processes*, ISO/IEC/IEEE Std., 2017.

[31] *IEEE Standard for System and Software Verification and Validation*, IEEE Std. 1012-2012, 2012.

[32] *IEEE Standard for Software Reviews and Audits*, IEEE Std. 1028-2008, 2008.

[33] P. N. Otto and A. I. Anton, "Addressing legal requirements in requirements engineering," in *15th IEEE International Requirements Engineering Conference (RE 2007)*, 2007, pp. 5–14.

[34] N. Kiyavitskaya, A. Krausov, and N. Zannone, "Why eliciting and managing legal requirements is hard," in *2008 Requirements Engineering and Law*, 2008, pp. 26–30.

[35] N. Kiyavitskaya, N. Zeni, T. D. Breaux, A. I. Antn, J. R. Cordy, L. Mich, and J. Mylopoulos, "Automating the extraction of rights and obligations for regulatory compliance," in *International Conference on Conceptual Modeling*. Springer, 2008, pp. 154–168.

[36] T. Breaux and A. Antn, "Analyzing regulatory rules for privacy and security requirements," *IEEE Transactions on Software Engineering*, vol. 34, pp. 5–20, 2008.

[37] A. Siena, J. Mylopoulos, A. Perini, and A. Susi, "From laws to requirements," in *2008 Requirements Engineering and Law*, 2008, pp. 6–10.

[38] ——, "Designing law-compliant software requirements," in *Conceptual Modeling - ER 2009*, A. H. F. Laender, S. Castano, U. Dayal, F. Casati, and J. P. M. de Oliveira, Eds. Springer Berlin Heidelberg, 2009, pp. 472–486.

[39] J. C. Maxwell, A. I. Antn, and P. Swire, "A legal cross-references taxonomy for identifying conflicting software requirements," in *2011 IEEE 19th international requirements engineering conference*. IEEE, 2011, pp. 197–206.

[40] *ANSI / IEEE Standard 1002.1987: Standard Taxonomy for Software Engineering Standards*, ANSI / IEEE Std. 1002-1987, 1987.

[41] S. Sharma, J. Henderson, and J. Ghosh, "Certifai counterfactual explanations for robustness, transparency, interpretability, and fairness of artificial intelligence models," in *arXiv preprint arXiv:1905.07857*, 2019.

[42] J. Rauber, R. Zimmermann, M. Bethge, and W. Brendel, "Foolbox native: Fast adversarial attacks to benchmark the robustness of machine learning models in pytorch, tensorflow, and jax," *Journal of Open Source Software*, vol. 5, p. 2607, 2020.

[43] M.-I. Nicolae, M. Sinn, M. N. Tran, B. Buesser, A. Rawat, M. Wistuba, V. Zantedeschi, N. Baracaldo, B. Chen, and H. Ludwig, "Adversarial robustness toolbox v1. 0.0," in *arXiv preprint arXiv:1807.01069*, 2018.

[44] A. Boopathy, T.-W. Weng, P.-Y. Chen, S. Liu, and L. Daniel, "Cnn-cert: An efficient framework for certifying robustness of convolutional neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3240–3247.

[45] R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, and A. Mojsilovi, "Ai fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias," *IBM Journal of Research and Development*, vol. 63, pp. 1–4, 2019.

[46] J. Rauber. (2021) Foolbox. [Online]. Available: https://foolbox.jonasrauber.de/

[47] ——. (2020) foolbox.attacks foolbox 3.3.1 documentation. [Online]. Available: https://foolbox.readthedocs.io/en/stable/modules/attacks.html

[48] R. Hamon, H. Junklewitz, and I. Sanchez, "Robustness and explainability of artificial intelligence," Publications Office of the European Union, Tech. Rep., 2020.

[49] P. Glauner, "An assessment of the ai regulation proposed by the european commission," *arXiv preprint arXiv:2105.15133*, 2021.

[50] L. Floridi, "The european legislation on AI: a brief analysis of its philosophical approach," *Philosophy & Technology*, vol. 34, no. 2, pp. 215–222, Jun. 2021.

[51] European Trade Union Institute RPS Submitter and Aida Ponce del Castillo, "The AI regulation: entering an AI regulatory winter? why an ad hoc directive on AI in employment is required," *SSRN Electronic Journal*, 2021.

[52] GitHub. Ai act propsal results wwi2018e / technical requirements and viable solutions for high risk ai-systems: The european union artificial intelligence act proposal - technical requirements

and viable solutions for high-risk ai systems. [Online]. Available: https://github.com/AI-Act-Propsal-Results-WWI2018E/Technical-Implications-and-Viable-Solutions-for-High-Risk-AI-Systems-

Appendix
Table of Appendices