RESEARCH-ARTICLE

Software Systems Compliance with the AI Act: Lessons Learned from an International Challenge

TERESA SCANTAMBURLO, Ca' Foscari University of Venice, Venice, VE, Italy

PAOLO FALCARIN, Ca' Foscari University of Venice, Venice, VE, Italy

ALBERTO VENERI, Ca' Foscari University of Venice, Venice, VE, Italy

ALESSANDRO FABRIS, Max Planck Institute for Security and Privacy, Bochum, Nordrhein-Westfalen, Germany

CHIARA GALLESE, University of Turin, Turin, TO, Italy

VALENTINA BILLA

View all

Published: 29 July 2024

RAIE '24: 2nd International Workshop on Responsible AI Engineering
April 16, 2024
Lisbon, Portugal

Conference Sponsors:
SIGSOFT

https://doi.org/10.1145/3643691.3648589
ISBN: 9798400705724

Software Systems Compliance with the AI Act

Lessons Learned from an International Challenge

Teresa Scantamburlo
teresa.scantamburlo@unive.it
Ca' Foscari University of Venice
Italy

Paolo Falcarin
paolo.falcarin@unive.it
Ca' Foscari University of Venice
Italy

Alberto Veneri
alberto.veneri@unive.it
Ca' Foscari University of Venice
ISTI-CNR
Italy

Alessandro Fabris
alessandro.fabris@mpi-sp.org
Max Planck Institute for Security and Privacy
Germany

Chiara Gallese
chiara.gallese@unito.it
University of Turin
Italy

Valentina Billa
vbilla@aude.legal
Aude - In2Law
Italy

Francesca Rotolo
francesca.rotolo@unive.it
Ca' Foscari University of Venice
Italy

Federico Marcuzzi
federico.marcuzzi@unive.it
Ca' Foscari University of Venice
Italy

ABSTRACT

In this experience paper, we present the lessons learned from the First University of St. Gallen Grand Challenge 2023, a competition involving interdisciplinary teams tasked with assessing the legal compliance of real-world AI-based systems with the European Union's Artificial Intelligence Act (AI Act). The AI Act is the very first attempt in the world to regulate AI systems and its potential impact is huge. The competition provided firsthand experience and practical knowledge regarding the AI Act's requirements. It also highlighted challenges and opportunities for the software engineering and AI communities.

CCS CONCEPTS

• **Social and professional topics** → **Governmental regulations**; • **Computing methodologies** → **Artificial intelligence**; • **Security and privacy** → *Privacy protections*; • **Software and its engineering** → **Software creation and management**.

KEYWORDS

AI Act, Requirements Engineering, Legal Compliance, Requirements Validation, Conformity Assessment

ACM Reference Format:
Teresa Scantamburlo, Paolo Falcarin, Alberto Veneri, Alessandro Fabris, Chiara Gallese, Valentina Billa, Francesca Rotolo, and Federico Marcuzzi. 2024. Software Systems Compliance with the AI Act: Lessons Learned from an International Challenge. In *2024 International Workshop on Responsible AI Engineering (RAIE'24), April 16, 2024, Lisbon, Portugal.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3643691.3648589

RAIE '24, April 16, 2024, Lisbon, Portugal

ACM ISBN 979-8-4007-0572-4/24/04.
https://doi.org/10.1145/3643691.3648589

1 INTRODUCTION

The Artificial Intelligence Act [52], also known as the AI Act, is the world's first statutory law proposal for regulating AI systems. It aims to turn Europe into a global hub for trustworthy and responsible AI by defining rules governing the development, operations, and use of AI in the European Union. The AI Act aims to ensure that AI systems in the EU are safe and respectful of fundamental rights and values. Its application extends beyond EU-based organizations and regards any AI provider, importer, distributor, or authorized representative within the EU. The expected impact of the AI Act is enormous [13]. The envisioned fines for non-compliance can range from €10 million to €40 million or 2% to 7% of the global annual turnover, depending on the severity of the violation (see Art. 71 of the proposal). Moreover, other states may follow the EU example, setting up new AI regulations, and the United Nations moves towards a globally coordinated AI governance [50]. Therefore, it is imperative for software providers to understand and comply with the upcoming regulations.

The Grand Challenge competition [10, 49] was organized by the University of St. Gallen (Switzerland) to showcase how the AI Act can be implemented in concrete real-world applications. The event took place between Geneva and St. Gallen in July 2023 and involved twelve teams participating in a selection process from various parts of the world. The competition took inspiration from the DARPA Robotics Challenge organized in Los Angeles in 2015 [1] and represents a unique example in the domain of legal AI. The competition consisted of assessing four AI applications and reporting the assessment results in a document of up to 13 pages (approximately 3 pages per application). Each application was presented by the provider in a 30-minute presentation. The providers came from different sectors, such as telecommunication and transportation, and included

companies at different maturity stages ranging from start-ups to stable businesses. Teams had the opportunity to collect information in short interviews (about 15 minutes per application/provider) and, for some providers, reviewed brief documentation provided before the final. The assessment reports were evaluated by a jury composed of law experts; the reward for the winning team was a monetary prize of 100,000 euros (see Burri [11] for more details on the rule of competition).

We were selected as one of the 12 teams attending the competition. We participated with a multidisciplinary team, Synergy4AI, composed of computer scientists, lawyers, and ethics experts. Our involvement in the Grand Challenge allowed us to test the AI Act on real-life AI systems and to reflect on the priorities and difficulties of the tasks associated with the assessment processes. In this paper, we report our experience, describing our methodological approach and the obstacles we met during the assessment. Some difficulties are closely tied to the special conditions of the events, but others are intrinsically connected with the implementation and evaluation of legal requirements for complex socio-technical systems. The diversity that characterizes our team (in terms of work and training experience) stimulated a multiperspective discussion shedding light on distinct challenges and recommendations (technical, organizational, legal, etc.).

The rest of the paper is organized as follows. Section 2 briefly introduces the AI Act and essential legal requirements; Section 3 presents our approach and methodology; Section 4 highlights key challenges and provides recommendations for the legal and ethical assessment of AI systems; Section 5 summarizes the lessons learned and outlines future research directions.

2 THE AI ACT

In April 2021, the European Commission introduced the first EU regulatory framework for AI with the aim of overseeing the development and utilization of this groundbreaking technology. The AI Act is part of a broader EU strategy designed to enhance Europe's potential to compete globally in regulating the digital sector.

The Commission has been tasked with aiding the co-legislators in concluding the inter-institutional negotiations, commonly known as the "trilogue" [59]. Noteworthy negotiation phases include the proposal of a compromise text on the AI Act by the Council of the EU in November 2021. Note that the version approved by the European Parliament is the legal text that was used during the Grand Challenge. Once approved, this regulation could be the world's first legislation governing AI.

The AI Act proposal delineates four distinct risk categories and sets specific requirements accordingly. These categories are:

- Unacceptable Risk (Title II - Art. 5 and following);
- High Risk (Title III - Art. 6 and following);
- Limited Risk (Art. 52);
- Minimal Risk / No Risk;

For systems deemed to pose an unacceptable risk, which is outright prohibited, the Act provides explicit examples and exceptions, including the utilization of real-time remote biometric identification in public spaces (such as facial recognition), social scoring systems (classifying individuals based on behavior, socio-economic status, or personal characteristics), and the use of subliminal manipulation techniques targeting specific vulnerable groups (Art. 5).

High-risk systems are permitted, but due to their ability to negatively affect safety or fundamental rights, they must comply with multiple requirements and undergo a compliance assessment throughout their life cycle, including before and after being deployed. High-risk systems are divided into two categories (Art. 6, Annexes II and III):

- AI systems intended to serve as safety components in products covered by the legislation listed in Annex II, or subject to third-party ex-ante conformity assessment (e.g., toys, aviation, cars, medical devices, and lifts).
- Stand-alone AI systems with mainly fundamental rights implications, listed in Annex III, that will have to be registered in an EU database.

Examples of high-risk systems include those related to critical infrastructure management, systems in hiring processes or employee ratings, credit scoring systems, and systems with critical impact on law enforcement and interpretation of law. Companies developing or deploying high-risk AI systems must comply with various requirements, including having an appropriate risk management system, logging capabilities, and human oversight (see Chapter 2 of the AI Act).

Similar to the GDPR, proper data governance must be applied to users' (and, more broadly, data subjects') data, but, in comparison to the data protection regulatory framework, the AI Act goes further by requiring data governance for data used in the training, testing, and validation of AI systems. It also imposes controls to ensure the accuracy, safety, and robustness of AI systems. Additionally, it emphasizes transparency in their design, enabling users to interpret the system's output. These systems are subject to specific obligations, including the implementation of a quality management system, the conduct of conformity assessments, and the preparation of technical documentation.

Other systems are considered limited or minimal risk. For limited-risk systems, only minimal transparency requirements are necessary: i.e., users must be informed when interacting with AI systems generating content, allowing them to make informed decisions about continuing the usage of such systems. Examples include chatbots, texts, images, and videos (e.g., deep fakes), which are not inherently high-risk; however, users must be aware that they are AI-generated content (Art. 52). Minimal-risk or no-risk AI is allowed to be freely used. This includes applications such as AI-enabled video games or spam filters. For operators of "AI systems other than high-risk AI systems" (Art. 69), the implementation of an ethical AI Code of Conduct is recommended to encourage voluntary compliance with the requirements outlined in Title III, Chapter 2 "Requirements for high-risk AI systems".

A separate discussion would be needed for Generative AI, like GPT [51], as they appear to fall under the category of "general-purpose AI systems" provided in Art. 3, which a "general purpose AI system means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed". Due to the complexity of this technology and the different positions of the EU bodies, implementing a risk-based approach for these systems proves to be challenging [53]. The only requirement, set out in Art. 28b and Recital 60th, is that they adhere to general transparency obligations outlined in Art.

52(1) and additional specific requirements, such as disclosing AI-generated content, designing the model to prevent the generation of illegal content, and publishing summaries of copyrighted data used for training. However, the exact terms of regulation are the object of an intense debate among the Member States [6], and it is not known if a political agreement about these systems will be reached in the final text of the AI Act.

3 COMPLYING WITH THE AI ACT

To understand the implications of the EU AI Act on their products, AI providers must first assess the risk profile of their AI systems. Depending on such assessment, a system might be banned from operating in the European Union, or its risk might be considered high, limited, or minimal. As we outlined in Section 2, the risk class determines different requirements and obligations for the AI actors involved. In general, the higher the risk level of the initial assessment, the more information an organization must provide. AI operators are nevertheless encouraged to develop and use AI systems in accordance with ethical principles (see Art. 4a) regardless of the risk level of their AI applications. Until the regulation is under discussion, there is still uncertainty on what demonstrating compliance with the AI Act may look like. However, our experience with GDPR compliance can provide us with some suggestions for establishing a compliance framework.

3.1 Related works

To demonstrate compliance with the AI Act, principles and requirements must be translated into low-level engineering interventions. In other words, the compliance process needs to ensure that requirements are measurable, verifiable, and continuously monitored. To fulfill this task, AI actors can draw on a vast repertory of methods and standards spanning from software engineering practices to optimization and human-computer interaction design principles. With respect to this large body of work, we may distinguish three main areas of interventions: 1) frameworks aimed at implementing the conformity assessment of high-risk AI systems, 2) standards that can be considered for certifying compliance with AI Act requirements; and 3) methods and toolkits that have been developed in the past few years by the AI ethics community to support the operationalization of trustworthy AI principles such as fairness and transparency. A brief description of meaningful contributions to each area is presented in the following subsections.

3.1.1 Frameworks for conformity assessment. Conformity assessment procedures are structured processes that enable organizations to ensure and demonstrate adherence to specific principles and/or legal requirements. Cap-AI is a framework that offers guidance to develop an ethical assessment of AI systems in line with the AI Act [16]. The framework adopts an ethics-centered approach and articulates the assessment process at each stage of the AI life cycle. The assessment consists of three main components: 1) the internal review protocol, 2) a summary data sheet, and 3) an external scorecard. Similar attempts include the development of procedures aimed at checking whether the engineering processes involved in the AI system conform to certain ethical principles and responsible innovation practices. For example, SMACTR is an internal audit framework used by the industry encompassing distinct stages: Scoping, Mapping, Artifact Collection, Testing, Reflection, and Post-Audit [55]. More recently, other scholars explored existing tools and frameworks enabling continuous auditing of AI systems, in line with the AI Act provision for post-market monitoring [38]. Besides frameworks providing methods and techniques to assess various types of AI systems, other approaches have been proposed to target specific AI models and applications, such as Large Language Models [42], algorithmic recruitment [32] and medical applications [35, 47].

3.1.2 Standards for AI Act compliance. At the time of writing, no specific standard was introduced for the certification of AI systems. While a request for standardization is expected to be issued by the European Commission after the regulation comes into force (art. 40 AI Act), some organizations have started to explore which standards can be considered for AI. The AI Watch provided a preliminary roadmap of international standards in support of the requirements set out by the AI Act [31]. Other efforts seek to extend and adapt existing standards to the specifications of AI models [44]. Indeed, some of the already available standards, technical specifications, and technical reports proposed by the International Organization of Standardization (ISO) can be used to build new standards that are directly applicable to the AI Act. Notable examples of such publications are standards made for describing generic AI Framework [23] and technical specifications for the assessment of machine learning classification performance [28]. In addition, various technical reports have been published, and they are a starting point for the standardization of various aspects of an AI system, such as the implication on the safety of machinery [29], the usage of AI in healthcare [30], the evaluation of the robustness [26] and trustworthiness [25], the assessment of bias [24] and ethical or societal concerns [27].

3.1.3 Methods for Trustworthy AI. In response to the widespread call for moving Trustworthy AI from principles to practice, the community of AI researchers and practitioners has developed a variety of methods and tools for specific ethical and legal principles. In this short summary, we report some examples with reference to risk assessment, transparency, and fairness. To manage (foreseeable) risks, AI actors can draw on classical software risk management approaches [8], algorithmic impact assessment tools [56], or frameworks specifically developed for AI systems such as [2, 43]. To improve the interpretability of AI systems' output, as mandated by the AI Act [18], researchers have developed several methods such as learning algorithms that aim to balance models' accuracy and interpretability [17, 46] and techniques trying to "explain" the predictions of black-box ML models [36, 57].

Another significant area of transparency measures regards documentation methodologies for both data sets and the models. There are methodologies providing a standardized format to describe data sets in terms of relevant features [20], including schemes adapted to specific application areas, such as Natural Language Processing [5] or fairness concerns [15]. Other frameworks provide comprehensive information about data sets (in the form of labels or scores) combining the results of qualitative and quantitative analysis [19, 21]. Regarding AI algorithms, Mitchell *et al.* [39] presents a framework to report benchmarked evaluations of trained ML models in various conditions, and a group of the EU Joint Research Center presented

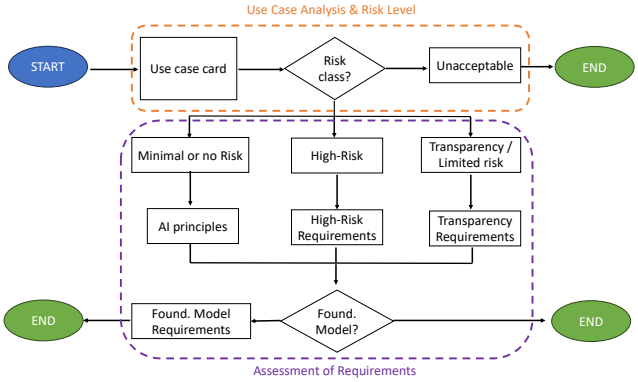

Figure 1: Flowchart of the first two steps of our assessment procedure, based on the AI Act version adopted at the challenge.

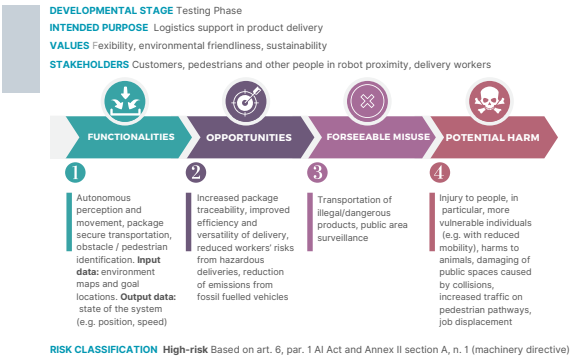

Figure 2: Prototype of a Use Case and Risk Profile Card for a delivery robot.

a template for documenting real-world, high-risk AI systems in line with information requirements set out in the AI Act [22].

The analysis of algorithmic (un)fairness and data bias (Art. 9 and 10) can be done by means of empirical tests on the data sets and the models [12]. Usually, these involve the adoption of one or more fairness metrics [37] but also a careful consideration of the underlying assumptions and modeling choices [40]. In the last few years, both industry and research institutions made available specific toolkits for fairness testing. Notable examples are [4, 7, 58] which offer Python and R packages to assess different types of fairness definitions in the entire application life cycle.

3.2 The Synergy4AI approach

During the challenge, we assessed real-world case studies and provided recommendations on plausibly achieving compliance with the AI Act. Our assessment approach was guided by a fundamental question: what steps should an AI provider take to adhere to the AI Act and, potentially, its underlying ethical principles? To address this question, we combined and adapted existing frameworks mentioned in the above section (see related works) with our own previous experience in GDPR and legal compliance. Our goal was twofold: first, to enable AI providers attending the challenge to consider AI Act requirements that were relevant to their AI applications; second, to provide them with actionable insights and specific steps to facilitate progress in the compliance process.

Our methodology relies upon AI system providers' document reviews, presentations, and face-to-face interviews. The assessment process structures around two key steps (see Figure 1 for a scheme): 1) the analysis and determination of the risk level of the AI system (see the orange dashed line in Figure 1); 2) the assessment of requirements applicable to the AI application based on the identified risk level (see the violet dashed line in Figure 1). The assessment process generates a report[1] composed of three parts:

The Use Case and Risk Profile Card. It contains salient information about the AI system and presents them in a structured, visual way. In particular, it conveys information about the status of the AI system (e.g. design or deployment), its intended purpose, the underlying values, and the stakeholders involved (directly or indirectly). Moreover, it briefly describes the functionalities of the systems, the gains it offers (i.e. the opportunities), possible misuses and negative effects (for a prototype of the Use Case and Risk Profile Card see Figure 2). This part builds upon [22] and suggests a preliminary standard to report critical information for assessing AI use cases through the lenses of the AI Act, selecting key information elements that could flexibly apply to all risk profiles. Given the tight time constraints of the challenge and the heterogeneity of providers interviewed, this part collects partial information and avoids the use of formal specifications as suggested by [22]. These pieces of information, combined with the team's expertise, determine the appropriate risk level of the AI system and, consequently, the most relevant legal and ethical requirements.[2]

The Requirements Assessment. The Requirements Assessment evaluates applicable requirements individually by using a checklist of items related to distinct risk levels (low or no risk, high-risk, transparency risk). Since the challenge used the proposal approved by the EU parliament, which added consideration for foundational models, the Requirements Assessment has an additional layer that assesses appropriate obligations for this peculiar case (Art. 28b). The checklist of items is inspired by questions elaborated for the internal review protocol put forward in the CapAI methodology [16]. Requirements are evaluated proportionally to their relevance for the assessed application, and some requirements involve more extensive consideration, such as the intersection with other legal requirements (GDPR, Copyright Law, the Machine Directive, etc.). In low-risk applications, a subset of items are considered based on the ethical principles that are most meaningful for the application at stake (Art. 4a AI Act).

The Recommendations. Ordered by urgency from the most to the least urgent, they conclude our assessment and involve different measures to support the AI providers based on the developmental stage of their AI system. Recommendations are classified into three main categories (legal, organizational, and technical). Examples of

[1]During the challenge all teams signed a Non-Disclosure Agreement with some AI providers. So we can disclose limited information in line with the rules of the competition.

[2]Note that no AI application was classified as prohibited.

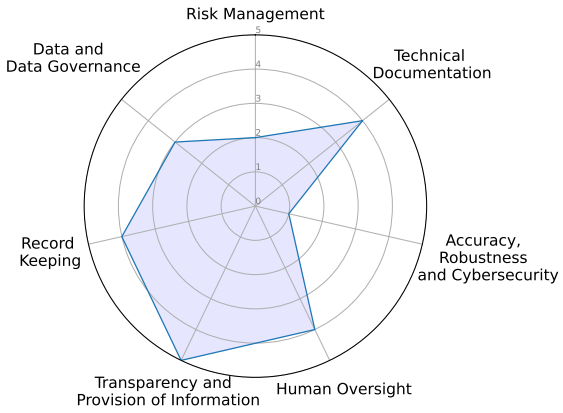

Figure 3: Fictional sample of the radar plot used to summarize the assessment for high-risk AI systems.

technical recommendations include: "Evaluate performance metrics (accuracy recall) with respect to geographical area for bias assessment" or "design a testbed for pedestrian detection system with attention to limited mobility subjects and minority groups."

The evaluation is expressed in two forms: on a 5-point categorical scale and qualitative descriptions. The categorical scale expresses the overall team's evaluation of the status of the AI system concerning requirements based on the review of checklist items. The categorical scale includes the following values ranging from 1 to 5, where 1 corresponds to a process seeking compliance that is "not initiated" (1), "initiated" (2), "underway" (3), "advanced" (4), "completed" (5). Scores are provided only for systems considered high-risk and presented in aggregate form by using a radar plot such as Figure 3.

The summary is divided into 6 main areas, namely: *Risk Management*, *Technical Documentation*, *Accuracy, Robustness and Cybersecurity*, *Human Oversight*, *Transparency and Provision of Information*, *Record Keeping*, *Data and Data Governance*. The subdivision of the requirements into 6 main areas reflects the main categorization present in the AI Act (from Art. 9 to Art. 15). This evaluation is complemented with textual information, adding more context and reasoning behind the evaluation of requirements or ethical principles. Scores and qualitative evaluation results from a process of consensus achievement within the team.

4 CHALLENGES

This section describes our experience in the Grand Challenge and the lessons learned while assessing the use cases from the participating industries about the different kinds of emerging challenges to implementing the AI Act compliance.

4.1 Our experience

Our team, Synergy4AI, is composed of six official members and two substitutes. It was diverse in terms of gender, age, professional experience, and disciplinary background. The team comprises experts in law (data protection, copyright, and corporate compliance), computer science and engineering (security, software engineering, algorithmic fairness, explainability), and computer ethics (ethics-based assessment, AI policy, data ethics). The team also includes two members with consultancy experience, with the aim to bring a pragmatic approach to the group and gain a better understanding of the business' needs.

Six months before the Grand Challenge, the team prepared for the competition in various ways. Team members met on a bi-weekly basis to delve into the topics of the AI Act and state-of-art methods seeking legal or ethical compliance of AI systems. Sessions were self-organized and enriched by the different expertise of each member. After an iterative process, the team identified the items to collect information on the AI use case and the questions to assess compliance with the AI Act. The preparation process was also characterized by an internal pilot of the assessment methodology on a real use case provided by the Italian branch of a multinational professional services firm. This exercise gave us the opportunity to gain feedback from an external entity and refine the methodology. In particular, thanks to the pilot, we adapted the checklist for the requirement assessments and clarified the task distribution during the interview with the AI provider.

We extensively relied on our experience with GDPR compliance due to the similarities with the AI Act, allowing us to combine the compliance procedures for both regulations:

(1) Risk and Impact Assessment: both regulations require several assessments to be performed (e.g., a Data Protection Impact Assessment, Fundamental Right Assessment, Transfer Assessment, Legitimate Interest Assessment);
(2) Data Governance and Cyber-security: both require appropriate measures to assess and protect the data;
(3) Transparency and Documentation: both GDPR and the AI Act mandate record keeping and extensive information to users/data subjects;
(4) Accountability and Human Oversight: both require appropriate procedures to guarantee a clear assignment of responsibilities and human supervision of automated systems;
(5) Organizational Measures: both regulations require the implementation of internal procedures and employee training to ensure continuous compliance;
(6) Legal Compliance: both require an assessment of the relevant legal framework and the applicable laws and regulations (e.g., the intersection with the Data Act, copyright laws, civil liability, sector legislation, etc.).

Our experience in legal compliance and consultancy also helped us during the interviews and in the relationship with technical experts and companies' management. This facilitated communication with the AI providers during and after the competition.

4.2 Legal Challenges

The proposal for AI regulation is a key component of the broader EU strategy aimed at bolstering Europe's ability to compete globally in overseeing the digital sector and other impactful domains associated with AI. In this context, some primary legal challenges arise from potential coordination issues with existing and forthcoming regulations, particularly those related to the protection and governance of both personal and non-personal data, as well as competition law and consumer law. Addressing challenges pertaining to the AI Act's content, a notable concern arose from the

expansive definition of AI systems. The rationale behind adopting such a comprehensive definition lies in prioritizing the "use case" over evaluating the specific technology behind the system. Essentially, the AI Act adopts a broad perspective on what qualifies as an AI system but adopts a more focused stance when considering its application. This approach aligns with the risk-based approach adopted by the AI Act. Nonetheless, this broad definition may introduce legal uncertainties for developers, operators, and users of AI systems, potentially granting excessive discretion to the private sector in determining what constitutes unacceptable, high, limited, or low levels of risk. This situation could be particularly detrimental to small and medium-sized entities and startups. Another legal challenge linked to the content of the AI Act pertains to "general-purpose AI systems" and the inherent difficulty in applying a risk-based approach to them. The challenge stems from the complexity of preemptively assessing the risks associated with these systems due to their potential for diverse purposes. Delegating the responsibility of defining permissible uses (along with the corresponding obligations for compliance) of potentially harmful systems to providers raises legitimate concerns.

4.3 Technical Challenges

Regulating a new technology from a technical point of view is always challenging, especially when we consider a fast-evolving technology such as the AI systems currently employed. Given the technical background of the members of our team, we can identify three main aspects of the technical challenges faced during the evaluation of the AI systems presented during the challenge, namely: *Software Quality Management*, *Fairness Management*, *Transparency and Explainability*.

Software Quality Management for compliance Organizations that develop, deploy, or use high-risk AI systems must have robust compliance procedures in place to ensure that they comply with the AI Act's requirements. This includes procedures for quality management, risk assessment, data management, training, auditing, and reporting. Quality and Risk Management processes for AI systems will play a fundamental part in complying with the AI Act, but despite recent advancements to improve the design process of an AI/ML system, the design techniques are still in their infancy. Design processes such as MLOps are not systematically employed yet, and data scientists are still managing ML workflows manually [34]. High-risk AI systems must be subject to rigorous testing and analysis to ensure reliability and safety for users. This can be a challenging task, as AI systems can be complex and opaque, making it difficult to identify and address potential defects with only a black-box testing approach on AI systems' APIs. As AI systems are constantly evolving, and new risks may arise as these systems become more sophisticated, organizations must be able to periodically adapt their compliance procedures to address these evolving risks.

Fairness Management. Non-discrimination is one of the key values upheld by the EU and the Member States' legislation; it has also been embedded into the AI Act, which highlights the need to mitigate biases in AI (see Art. 10). AI systems can perpetuate and exacerbate existing biases in data, leading to unfair and discriminatory outcomes. Ensuring fairness, which is context-dependent, requires careful consideration of the data used to train the system, the algorithms used for processing data, and the human oversight of the system's operation. Algorithmic fairness deals with equity and non-discrimination in AI [3]. The definition, detection, and prevention of undesirable biases in AI, however, depend on its application context. Socioeconomic status, for example, is important in medicine, where less wealthy patients have lower access to healthcare [48]. Speech impairments and accents are highly relevant in any application leveraging speech recognition (or speech-to-text), where unusual speech patterns are rare in training sets and more difficult to parse automatically [45]. Intersectional gendered and racial identities are especially relevant in computer vision systems, which often under-perform black women [9]. These biases evolve dynamically in response to audits and shifting incentives [54]. In hiring and recruitment, several attributes are explicitly protected by European law, including age, disability, gender, religion or belief, racial or ethnic origin, and sexual orientation [14]. A generic fairness approach poses the risk of missing critical vulnerabilities; domain-specific knowledge is required to identify, measure, and tackle specific biases.

Transparency and Explainability. High-risk AI systems must be transparent and explainable, meaning that users should be able to understand the system's decision-making processes: a challenging task, as AI systems can be complex and opaque. Even though a lot of academic research has been done in the context of Explainable Artificial Intelligence [41], the concept of "explanation of prediction" is still vague for ML practitioners and sometimes the consideration of what is a valid explanation or not is based on the individual perception. Therefore, there is the need to have a more precise operational and compliance-focused description of explainability, that can be seen as a new quality attribute to measure and improve during the system development.

4.4 Organizational and Cultural Challenges

Implementing the AI Act poses significant organizational challenges that companies must overcome to ensure compliance and ethical production and use of AI systems.

One key organizational challenge is represented by the lack of specialized human resources and by the consequent difficulties in establishing effective communication among legal and technical professionals, as well as aligning legal perspectives with the broader goals and operations of the company. Achieving synergy between experts from different fields, such as computer science, ethics, and law, is crucial, given that most of the problems addressed are multidimensional and require a deep understanding of the legal framework, the business, the social implications of the specific AI system and its technical foundations. This also reflects the increasing hybridization of roles, requiring experts to have not only traditional legal skills but also a deep understanding of technological aspects. For example, in order to perform the risk assessment, the interviews with stakeholders (e.g., executives, technicians, IT experts) must be very precise and focused on the most relevant questions, taking into account the peculiarities of the business sector in which the company operates.

Furthermore, integrating legal expertise with ethical considerations, such as performing the Fundamental Rights Impact Assessment, can only be effective if experts are involved in the process, a circumstance that is resource-consuming and might not be a

realistic option for smaller entities. Simplifying compliance processes poses another organizational burden. Companies must strike a balance between meeting legal requirements and addressing the operational and resource constraints within the organization. As we have suggested, an effective solution is to coordinate and integrate efforts, especially when multiple laws overlap, such as GDPR and the Medical Device Regulation. Legal departments need to align compliance efforts to avoid redundancies and optimize resources.

Moreover, the documentation of legal compliance related to multiple applicable laws might represent a challenge for less experienced companies. New techniques, such as legal design and specialized compliance software, aimed at synthesizing and simplifying legal requirements, might be of great help. This shift in approach ensures that compliance documentation is not only comprehensive but also accessible and understandable across various organizational levels.

Training is another topic of interest: as known, GDPR mandates effective employee training to ensure that compliance is performed at all organizational levels. A similar approach might be incorporated for the AI Act compliance: employees should be trained about relevant ethical and technical issues, such as debiasing techniques, human rights issues, post-market risk assessment, and quality management.

5 RECOMMENDATIONS AND CONCLUSIONS

The AI Act, like many regulatory requirements that the EU and other national institutions are working on introducing, can have a wide impact on the AI market [13] and on society as a whole. Many companies are preparing for the AI Act and, overall, their perception of the legislation is positive: the industry would like more precise regulations and more input from AI experts [33] to mitigate the risk that poorly planned investments in AI could damage their revenues and reputation. The recommended actions that companies can perform to prepare for the AI Act are:

(1) Make sure the legal department studies the AI Act text thoroughly, focusing on the definitions of the specific AI use case risk levels, to better understand the actions to take for particular use cases and consider the interactions with the existing legal framework.
(2) Integrate the AI Act compliance into preexisting compliance procedures and training programs to minimize the impact on the workload and resources;
(3) Ensure that all stakeholders are informed about the significance of the AI Act through training and awareness initiatives.
(4) Establish a well-defined policy identifying the authorized people and entities responsible for AI implementation decisions and specifying the restricted deployment areas for AI applications.
(5) Introduce a traceability framework throughout the whole AI design process to identify the individuals responsible for important decisions and describe the methodologies employed in the decision-making process.
(6) Exercise caution when considering use cases that involve the utilization of personal data, in accordance with GDPR or other relevant laws and regulations.
(7) Clearly document AI use cases, and establish systematic monitoring, via log-keeping and conduct periodic reviews to maintain compliance.
(8) Document all the dependencies and integration with third-party software and services, all copyrighted materials employed by AI systems (e.g., a "Software Bill of Materials"), for transparency and accountability.[3]

Despite the undeniable compliance burden, the Grand Challenge experience has taught us that the AI Act compliance can be carried out fruitfully if integrated into existing procedures and performed by experts in the field. While it will surely take time to adjust and train employees and executives, we are positive that, in the end, the AI Act will benefit the EU AI landscape and help mitigate biases, especially if adequately supported by EU internal technological infrastructures.

ACKNOWLEDGMENTS

This work was partially funded by the the European Union's Horizon Europe program with the DataCom Project (grant agreement no. 101108151) and project FINDHR (Fairness and Intersectional Non-Discrimination in Human Recommendation, grant no. 101070212). Views and opinions expressed are those of the author(s) only and do not reflect those of the EU or the European Commission. This work was partially funded by Ca' Foscari University's IRIDE program and by project iNEST (Interconnected NordEst Innovation Ecosystem), funded by PNRR, NextGenerationEU (ECS 00000043).

REFERENCES

[1] Defense Advanced Rsearch Projects Agency. 2023. *DARPA Robotics Challenge.* https://www.darpa.mil/news-events/drc-finals
[2] NIST AI. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0). (2023).
[3] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2023. *Fairness and Machine Learning: Limitations and Opportunities.* MIT Press.
[4] Rachel K. E. Bellamy, Kuntal Dey, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilovic, Seema Nagar, Karthikeyan Natesan Ramamurthy, John Richards, Diptikalyan Saha, Prasanna Sattigeri, Moninder Singh, Kush R. Varshney, and Yunfeng Zhang. 2018. AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. https://arxiv.org/abs/1810.01943
[5] Emily M Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics* 6 (2018), 587–604.
[6] Luca Bertuzzi. 2023. *AI Act: Spanish presidency makes last mediation attempt on foundation models.* https://www.euractiv.com/section/artificial-intelligence/news/ai-act-spanish-presidency-makes-last-mediation-attempt-on-foundation-models/
[7] Sarah Bird, Miro Dudík, Richard Edgar, Brandon Horn, Roman Lutz, Vanessa Milan, Mehrnoosh Sameki, Hanna Wallach, and Kathleen Walker. 2020. Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft, Tech. Rep. MSR-TR-2020-32* (2020).
[8] Barry W. Boehm. 1991. Software risk management: principles and practices. *IEEE software* 8, 1 (1991), 32–41.
[9] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Conference on Fairness, Accountability and Transparency, FAT 2018, 23-24 February 2018, New York, NY, USA (Proc. of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 77–91.
[10] Thomas Burri. 2023. A Challenge for Law and Artificial Intelligence. *Nature Machine Intelligence (accepted)* (2023).
[11] Thomas Burri. 2023. *The First University of St. Gallen Grand Challenge: The EU AI Act 2023–Rulebook 2.0.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4503312

[3] https://www.cisa.gov/sbom

[12] Zhenpeng Chen, Jie M Zhang, Federica Sarro, and Mark Harman. 2023. A Comprehensive Empirical Study of Bias Mitigation Methods for Machine Learning Classifiers. *ACM Transactions on Software Engineering and Methodology* 32, 4 (2023), 1–30.

[13] COM(2021) 206 final - SEC(2021) 167 final - SWD(2021) 85 final 2021. *Commission Staff Working Document Impact Assessment Accompanying 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.* Technical Report. European Commission, Bruxelles, BE.

[14] Alessandro Fabris, Nina Baranowska, Matthew J Dennis, Philipp Hacker, Jorge Saldivar, Frederik Zuiderveen Borgesius, and Asia J Biega. 2023. Fairness and Bias in Algorithmic Hiring. *arXiv preprint arXiv:2309.13933* (2023).

[15] Alessandro Fabris, Stefano Messina, Gianmaria Silvello, and Gian Antonio Susto. 2022. Algorithmic fairness datasets: the story so far. *Data Mining and Knowledge Discovery* 36, 6 (2022), 2074–2152.

[16] Luciano Floridi, Matthias Holweg, Mariarosaria Taddeo, Javier Amaya Silva, Jakob Mökander, and Yuni Wen. 2022. CapAI-A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. *Available at SSRN 4064091* (2022).

[17] Caro Fuchs, Uzay Kaymak, and Marco S Nobile. 2022. Building interpretable and parsimonious fuzzy models using a multi-objective approach. In *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 1–8.

[18] Chiara Gallese. 2023. The AI Act proposal: a new right to technical interpretability? *arXiv preprint arXiv:2303.17558* (2023).

[19] Chiara Gallese, Teresa Scantamburlo, Luca Manzoni, and Marco S Nobile. 2023. Investigating Semi-Automatic Assessment of Data Sets Fairness by Means of Fuzzy Logic. In *2023 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 1–10.

[20] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[21] Sarah et al. Holland. 2020. The dataset nutrition label. *Data Protection and Privacy* 12, 12 (2020).

[22] Isabelle Hupont Torres, David Fernández-Llorca, Sandra Baldassarri, and Emilia Gómez. 2023. *Use case cards: a use case reporting framework inspired by the European AI Act.* https://arxiv.org/abs/2306.13701

[23] ISO/IEC 23053:2022 2022. *Framework for Artificial Intelligence Systems Using Machine Learning (ML).* Standard. International Org. for Standardization, Geneva.

[24] ISO/IEC TR 24027:2021 2021. *Information technology – Artificial intelligence (AI) – Bias in AI systems and AI aided decision making.* Technical Report. International Organization for Standardization, Geneva, CH.

[25] ISO/IEC TR 24028:2020 2020. *Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence.* Technical Report. International Organization for Standardization, Geneva, CH.

[26] ISO/IEC TR 24029-1:2021 2021. *Artificial Intelligence (AI) – Assessment of the robustness of neural networks – Part 1: Overview.* Technical Report. International Organization for Standardization, Geneva, CH.

[27] ISO/IEC TR 24368:2022 2022. *Information technology – Artificial intelligence – Overview of ethical and societal concerns.* Technical Report. International Organization for Standardization, Geneva, CH.

[28] ISO/IEC TS 4213:2022 2022. *Information technology – Artificial intelligence – Assessment of machine learning classification performance.* Technical Specification. International Organization for Standardization, Geneva, CH.

[29] ISO/TR 22100-5:2021 2021. *Safety of machinery – Relationship with ISO 12100 – Part 5: Implications of artificial intelligence machine learning.* Technical Report. International Organization for Standardization, Geneva, CH.

[30] ISO/TR 24291:2021 2021. *Health informatics – Applications of machine learning technologies in imaging and other medical applications.* Technical Report. International Organization for Standardization, Geneva, CH.

[31] Soler Garrido et al. Josep. 2023. *AI Watch: Artificial Intelligence Standardisation Landscape Update.* Technical Report. Joint Research Centre (Seville site).

[32] Emre Kazim, Adriano Soares Koshiyama, Airlie Hilliard, and Roseline Polle. 2021. Systematizing Audit in Algorithmic Recruitment. *Journal of Intelligence* 9, 3 (Sept. 2021), 46. https://doi.org/10.3390/jintelligence9030046 Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[33] Fiona Koh, Kathrin Grosse, and Giovanni Apruzzese. 2024. Voices from the Frontline: Revealing the AI Practitioners' viewpoint on the European AI Act. In *Proc. Hawaiian International Conference on System Sciences (HICSS)*.

[34] Dominik Kreuzberger, Niklas Kühl, and Sebastian Hirschl. 2023. Machine Learning Operations (MLOps): Overview, Definition, and Architecture. *IEEE Access* 11 (2023), 31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138

[35] Xiaoxuan et al. Liu. 2022. The medical algorithmic audit. *The Lancet Digital Health* 4, 5 (May 2022), e384–e397. https://doi.org/10.1016/S2589-7500(22)00003-6

[36] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[37] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–35.

[38] Matti Minkkinen, Joakim Laine, and Matti Mäntymäki. 2022. Continuous auditing of artificial intelligence: a conceptualization and assessment of tools and frameworks. *Digital Society* 1, 3 (2022), 21.

[39] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*. 220–229.

[40] Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. 2021. Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application* 8 (2021), 141–163.

[41] Christoph Molnar. 2022. *Interpretable Machine Learning* (2 ed.). https://christophm.github.io/interpretable-ml-book

[42] Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. 2023. Auditing large language models: a three-layered approach. *AI and Ethics* (May 2023). https://doi.org/10.1007/s43681-023-00289-2

[43] Per Rådberg Nagbøl, Oliver Müller, and Oliver Krancher. 2021. Designing a risk assessment tool for artificial intelligence systems. In *International Conf. on Design Science Research in Information Systems and Technology*. Springer, 328–339.

[44] Domenico Natale. 2022. Extensions of ISO/IEC 25000 Quality Models to the Context of Artificial Intelligence. *Proceedings of IWESQ@ APSEC* (2022).

[45] Mikel K. Ngueajio and Gloria J. Washington. 2022. Hey ASR System! Why Aren't You More Inclusive? - Automatic Speech Recognition Systems' Bias and Proposed Bias Mitigation Techniques. A Literature Review. In *HCI International 2022 - Late Breaking Papers: Interacting with eXtended Reality and Artificial Intelligence - 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26 - July 1, 2022, Proceedings (Lecture Notes in Computer Science, Vol. 13518)*, Jessie Y. C. Chen, Gino Fragomeni, Helmut Degen, and Stavroula Ntoa (Eds.). Springer, 421–440. https://doi.org/10.1007/978-3-031-21707-4_30

[46] Harsha Nori, Samuel Jenkins, Paul Koch, and Rich Caruana. 2019. Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223* (2019).

[47] Lauren Oakden-Rayner, William Gale, Thomas A Bonham, Matthew P Lungren, Gustavo Carneiro, Andrew P Bradley, and Lyle J Palmer. 2022. Validation and algorithmic audit of a deep learning system for the detection of proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study. *The Lancet Digital Health* 4, 5 (May 2022), e351–e358. https://doi.org/10.1016/S2589-7500(22)00004-8

[48] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453.

[49] University of Saint Gallen. 2023. *First University of St. Gallen Grand Challenge: The EU A.I. Act 2023.* Retrieved November 3, 2023 from https://www.thegrandchallenge.eu/

[50] UN Office of the Secretary-General's Envoy on Technology. 2023. *High-Level Advisory Body on AI.* https://www.un.org/techenvoy/ai-advisory-body

[51] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[52] European Parliament. 2023. *EU AI Act: first regulation on artificial intelligence.* https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[53] European Parliament. 2023. *Parliament's negotiating position on the artificial intelligence act.* https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/747926/EPRS_ATA(2023)747926_EN.pdf

[54] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, Honolulu, HI, USA, January 27-28, 2019*, Vincent Conitzer, Gillian K. Hadfield, and Shannon Vallor (Eds.). ACM, 429–435. https://doi.org/10.1145/3306618.3314244

[55] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 33–44.

[56] Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker. 2018. Algorithmic Impact Assessments: A Practical Framework for Public Agency. *AI Now* (2018).

[57] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proc. of the 22nd ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*. ACM, San Francisco California USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[58] Pedro Saleiro, Benedict Kuester, Abby Stevens, Ari Anisfeld, Loren Hinkson, Jesse London, and Rayid Ghani. 2018. Aequitas: A Bias and Fairness Audit Toolkit. *arXiv preprint arXiv:1811.05577* (2018).

[59] Kai Zenner. 2023. The AI Act: all publicly available documents in September 2023. https://www.kaizenner.eu/post/aiact-part3