Navigating the EU AI Act Maze using a Decision-Tree Approach

Hanif, Hilmy; Constantino Torres, J.E.; Sekwenz, M.T.; van Eeten, M.J.G.; Ubacht, J.; Wagner, Ben; Zhauniarovich, Y.

DOI
10.1145/3677174

Publication date
2024

Document Version
Final published version

Published in
ACM Journal on Responsible Computing

Navigating the EU AI Act Maze using a Decision-Tree Approach

HILMY HANIF, TPM, TU Delft, Delft, Netherlands
JORGE CONSTANTINO, TPM, TU Delft, Delft, Netherlands
MARIE-THERESE SEKWENZ, TPM, TU Delft, Delft, Netherlands
MICHEL VAN EETEN, TPM, TU Delft, Delft, Netherlands
JOLIEN UBACHT, TPM, TU Delft, Delft, Netherlands
BEN WAGNER, TPM, TU Delft, Delft, Netherlands
YURY ZHAUNIAROVICH, TPM, TU Delft, Delft, Netherlands

The AI Act represents a significant legislative effort by the European Union to govern the use of AI systems according to different risk-related classes, imposing different degrees of compliance obligations to users and providers of AI systems. However, it is often critiqued due to the lack of general public comprehension and effectiveness regarding the classification of AI systems to the corresponding risk classes. To mitigate these shortcomings, we propose a Decision Tree–based framework aimed at increasing legal compliance and classification clarity. By performing a quantitative evaluation, we show that our framework is especially beneficial to individuals without a legal background, allowing them to enhance the accuracy and speed of AI system classification according to the AI Act. The qualitative study results show that the framework is helpful to all participants, allowing them to justify intuitively made decisions and making the classification process clearer.

CCS Concepts: • **Applied computing** → **Law**; • **Human-centered computing** → *Visualization techniques;*

Additional Key Words and Phrases: Artificial intelligence, AI act, AIA, risk classification, compliance

ACM Reference Format:
Hilmy Hanif, Jorge Constantino, Marie-Therese Sekwenz, Michel van Eeten, Jolien Ubacht, Ben Wagner, and Yury Zhauniarovich. 2024. Navigating the EU AI Act Maze using a Decision-Tree Approach. *ACM J. Responsib. Comput.* 1, 3, Article 21 (September 2024), 16 pages. https://doi.org/10.1145/3677174

1 Introduction

The prevalence of applications incorporating **Artificial Intelligence (AI)** is experiencing a notable rise. This surge is propelled by the advancements in computing technology, the refinement of algorithms, and the accessibility of extensive datasets, resulting in the pervasive integration of these technologies into nearly every facet of our lives. McKinsey's research underscores the

Authors' Contact Information: Hilmy Hanif, TPM, TU Delft, Delft, Netherlands; e-mail: HilmyHanif@student.tudelft.nl; Jorge Constantino, TPM, TU Delft, Delft, Netherlands; e-mail: J.E.ConstantinoTorres@tudelft.nl; Marie-Therese Sekwenz, TPM, TU Delft, Delft, Netherlands; e-mail: M.T.Sekwenz@tudelft.nl; Michel van Eeten, TPM, TU Delft, Delft, Netherlands; e-mail: M.J.G.vanEeten@tudelft.nl; Jolien Ubacht, TPM, TU Delft, Delft, Netherlands; e-mail: j.ubacht@tudelft.nl; Ben Wagner, TPM, TU Delft, Delft, Netherlands; e-mail: B.Wagner@tudelft.nl; Yury Zhauniarovich, TPM, TU Delft, Delft, Netherlands; e-mail: y.zhauniarovich@tudelft.nl.

ACM 2832-0565/2024/09-ART21
https://doi.org/10.1145/3677174

potential of AI, suggesting that generative AI alone could contribute to an annual economic value of up to $4.4 trillion globally [5]. However, it is imperative to acknowledge that while these technologies yield significant benefits, their (uncontrolled) use can also harm society.

The European **Artificial Intelligence Act (AI Act, AIA)** represents a significant first legislative effort by the **European Union (EU)** to govern AI systems [12]. This legislative framework aims to ensure the trustworthiness of AI systems, aligning their deployment with fundamental rights and the European values [11, 12, 16]. To achieve these goals, the AI Act follows a risk-based approach, delineating distinct governance measures tailored for specific defined risk classes of AI systems.

Unfortunately, since the proposal of the AI Act, researchers and practitioners pointed out that the classification criteria mentioned in the AI Act might be unclear and problematic [22, 28]. Certain AI systems may fall under more than one risk category [22]. According to the Initiative for Applied Artificial Intelligence [28], about 40% of corporate AI systems may experience classification ambiguity, wherein their associated risks remain unclear. This may result in added time and financial investments required for conformity assessments, potentially leading to delays in bringing AI products to market, slowing down the economy and societal benefits. The need to evaluate not only final products but every innovative idea of an AI system increases the costs even more. Moreover, even if an expert judgment is obtained, this opinion might be subjective, opaque to the public, and not shared by courts. Furthermore, legal interpretation and enforcement of such regulations require specialized knowledge, representing an additional challenge in implementing the AI Act [10, 17, 38]. It would require novel expertise and a base of examples and precedents from jurisprudence, which is hard to get if regulation is new and relevant case law still has to be developed over time. All these reasons stress the point made by Liebl and Klein [28], calling for standardization and clearer guidance for AI systems classification [24, 41].

In this work, we try to address this gap by answering the research question *whether it is possible to improve the classification of AI systems into risk categories in compliance with the AI Act.* In light of the widespread deployment of systems potentially falling within the scope of the AI Act, our primary research objective is to empower individuals, including those *lacking specialized expertise*, with a tool that can facilitate the task of *reliable* AI systems classification.

To answer the research question and reach our goal, we propose a **Decision Tree–based (DT-based)** framework that aims to enable individuals with different backgrounds (including non-legal) to classify AI systems into appropriate risk categories in accordance with the AI Act. Through a quantitative assessment, we highlight the exceptional utility of our framework, particularly for individuals lacking a legal background. It enables them to substantially enhance classification accuracy and significantly reduce case classification time. Concurrently, the outcomes of our qualitative study reveal that the framework garners appreciation from all participants, rendering the classification process more transparent and intuitively guiding decision-making. *We share our DT-based framework*[1] [20] so that the community can validate the results and improve the framework further. Preliminary results of this study are described in Reference [21].

The rest of the article is organized as follows. It starts with a comprehensive background section (Section 2) that provides an overview of the development of the AI Act, its risk classification guidelines, and the Decision Tree concept. The methodology section (Section 3) details the development of the DT-based framework and the steps to perform its evaluation, which quantitative and qualitative results are provided in Section 4. Subsequently, Section 5 critically analyzes the findings, highlighting strengths, limitations, and potential areas for improvement of the proposed framework. Special emphasis is put on highlighting the changes between the adopted version of

[1] https://data.4tu.nl/datasets/bf7013ce-54b5-43b9-b275-f6c44652534b

the AI Act and the draft version utilized in this study. Finally, Section 6 explores existing literature relevant to the topic, and Section 7 concludes.

2 Background

The AI Act. In April 2021, the European Commission proposed the AI Act (herein, *the AI Act draft*[2]) to establish rules for the development and use of AI technologies, which was adopted on March 13, 2024 (hereafter, *the adopted version of the AI Act*) [15] after a few years of negotiations with stakeholders. The AI Act aims to foster technological development while promoting the protection of fundamental rights and European values [3, 11]. The regulation's effects extend across AI system providers who either place or operate them within the EU, irrespective of their geographical location [12, 40].

To achieve its objectives, the AI Act adopts a risk-based approach, delineating distinct governance measures tailored for specifically defined risk classes of AI systems, which are defined by the adopted version of the AI Act as "*a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.*" Although other jurisdictions, e.g., Quebec [7], have also tried regulating the use of AI in society, the European AI Act is the first regulatory approach looking at the risks of AI systems [12].

Risk Classification. The AI Act draft established a four-level risk classification guideline for AI systems[3]: ***Unacceptable Risk (UR)***, ***High Risk (HR)***, ***Limited Risk (LR)***, and ***Minimal Risk (MR)***. AI systems violating fundamental rights and EU values, e.g., those using remote biometric identification or manipulating a person's behavior, belong to the *Unacceptable-risk* class. Any AI system falling under this category is strictly prohibited from being placed on the market, put into service, or used, with the exception of military purposes (see Art 2(3) AI Act [12]). *High-risk* AI systems are those that potentially pose significant risks to the health, safety, or fundamental rights of individuals. The classification of technology as high risk is determined not solely by the function an AI system performs but also by its intended purpose and operational methods. *High-risk* AI systems require an ex-ante third-party conformity assessment and are subject to more rigorous compliance obligations, like the use of risk management, demanded standards of data quality, technical documentation, human oversight, increased transparency efforts, robustness, accuracy, and security. *Limited-risk* AI systems are the ones that interact with human beings but do not belong to the first two categories. Such systems are mainly required to comply with transparency obligations. For instance, chatbots have to inform the user about the use of the AI systems in products or components (see Art 52 AI Act [12]). Applications categorized as *Minimal Risk* encompass all other systems that have been extensively deployed and constitute the majority of the AI systems with which we currently engage, e.g., spam filters or video games with AI-driven agents.

Decision Tree. A **Decision Tree (DT)** is a concept representing a series of choices and their outcomes. This concept has been around for over half of a century [4] and is currently known both as a well-established machine learning set of algorithms and as a visual form of representing decision-making processes. In this study, we refer to the latter notion. It can be imagined as a structure resembling an upside-down tree (from root to leaves), where each internal node corresponds to a test, each branch represents a potential outcome of that test, and each leaf exhibits an end result. A sample is probed against these tests, each of which either assigns it to a particular end result or

[2]Within this study, we used the draft version of the AI Act.

[3]Note that the adopted version of the AI Act has changed this classification, as discussed in Section 5.

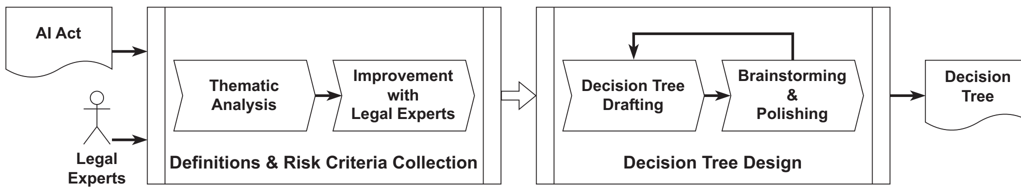

Fig. 1. Framework design procedure.

brings it to the next test down the tree. This abstraction is well-suited for handling multivariate, multiclass classification problems typical in decision-making scenarios [4, 6]. Moreover, decision trees are well-known for their simplicity and interpretability, making it apt even for inexperienced individuals to understand and explain the classification process.

3 Methodology

In this study, we employed a mixed-methods methodology. For designing the DT-based framework, we applied **Design Science Methodology (DSM)** that enables the systematic study and creation of artifacts to address practical problems of general interest [23]. The methodology encompasses identifying the requirements necessary to enhance AI systems classification, transforming these requirements into a framework that can further improve the classification process, and evaluating the framework by collecting feedback from AI experts. To evaluate the developed artifact, we designed and executed a protocol that allowed us to obtain both quantitative results of performance increase achieved with the framework and qualitative data estimating the acceptance of the framework by the interviewees.

3.1 Framework Design

To design the framework, we followed the procedure shown in Figure 1 that consists of two phases: (1) Definitions and Risk Criteria Collection and (2) Decision Tree Design.

The Definitions and Risk Criteria Collection phase starts with a **thematic** analysis of the AIA draft [12], the version of the proposal that was available at the start of our research (the beginning of 2023). The goal of this analysis was to gain an understanding of the criteria allowing one to attribute an AI system to a particular risk class and to collect relevant definitions and explanations. These components are crucial for generating the framework.

We performed the thematic analysis following the recommendations proposed by Lindgren et al. [30]. We chose this method because of its detailed exploration of the relationship between abstraction and interpretation within each stage of the analytical process. This enhances the credibility of our analysis compared to traditional thematic analysis, which is often criticized for its perceived lack of depth, scientific rigor, and evidential support. Following this method, we first analyzed the AI Act draft and selected all text passages associated with each risk class, including relevant definitions, explanations, descriptions and criteria. Second, we condensed these legislative passages condensed these chunks by removing repetitive and non-essential words while preserving the core content. Then, we labeled the condensed units with descriptive codes. Finally, we created risk-related categories by grouping codes related to a particular risk.

Note that this process is prone to inaccuracies, because specialized legal knowledge and practice are required to include the relevant details of the law. To mitigate this risk, we collaborated with two legal experts who helped us to resolve misunderstandings and clarify poorly understood contexts. In particular, the legal team helped the authors to clarify how relevant legal passages must be interpreted to model the tree correctly: e.g., how does the risk classification of high-risk systems work in relation to Annex III? Furthermore, their insights helped to interpret the AI Act to frame the questions of the proposed tree. For instance, the legal team made sure that fundamental legal principles addressed in the AI Act, the GDPR, or the Charter of Fundamental Rights were incorporated correctly into the articulation of the tree questions. Additionally, the legal team suggested how to improve the structure of the tree according to these linked legal sources.

During the second phase, we built a draft of the Decision Tree–based framework by organizing the risk-related information obtained during the previous phase. Then, we organized several brainstorming sessions, during which the authors of this article further elaborated, cleaned, and polished the obtained framework.

The final version of the framework consists of 20 questions organized into four pre-selected themes: *Protected Values*, *Objective/Intention*, *Domain*, and *Use-case/Technology*. The classification of AI systems under the AI Act aims to safeguard fundamental rights and Union values. Therefore, the *Protected Values* theme questions assist in excluding practices that are fundamentally prohibited. The goal of the *Objective/Intention* questions is to assess the intention and objectives of the proposed AI systems and their use. The *Domain* theme questions evaluate if an AI system is used in a specific domain, such as education, the workplace, critical infrastructure, and so on, that can put it under the High-risk category. Finally, the *Use-case/Technology* theme unites questions that check if an AI system uses a specific technology or is applied for a particular use case. These theme questions aim to list the use cases or technologies where AI cannot be used. To simplify the usage of our framework, we also added additional information blocks that influence the decision-making process. Thus, strictly speaking, our framework is not a decision tree but closely resembles it; therefore, we call it a DT-based framework.

To classify a case (an AI system), a decision-maker traverses the tree, responding to the questions encapsulated within each node and proceeding along the branch that corresponds to the most appropriate choice. The decision-maker starts from the root question, formulated as "Does it potentially cause significant harm to fundamental rights and Union values?" and follows the decision tree until reaching a leaf node representing the risk class. The longest path corresponds to 12, i.e., in the worst case, the decision-maker needs to answer 12 questions to make a conclusion.

3.2 Framework Evaluation

To evaluate our DT-based framework, we designed the following experiment. We selected several use cases of AI systems and asked participants to classify them into the corresponding risk categories. With each participant, we ran an evaluation session divided into three sections: two experimental parts and a semi-structured interview. During the first two sections, participants were tasked to classify a set of AI system use cases into four risk categories according to the AI Act draft. In the first section, they classified the AI systems without the aid of the DT-based framework, relying only on their interpretation of the AI Act draft. Prior to the interview, participants were sent an invitation email containing relevant details regarding the risk categories outlined in Articles 5, 6, 52, 69, and Annexes II and III of the AI Act. Furthermore, just before the start of the first section, participants were provided with the same information and allocated time to refresh their understanding. The interviewers refrained from discussing this information with the participants to prevent any potential bias. During the second section, the interviewees utilized the proposed framework to classify the cases. During the third section, they answered several semi-structured questions aimed at figuring out participants' opinions on the proposed framework and their understanding of the classification process. The one-on-one interview sessions, with a duration of approximately 60 to 90 min per respondent, were conducted online through a video call following the protocol[4] [20] and were recorded.

[4] https://data.4tu.nl/datasets/bf7013ce-54b5-43b9-b275-f6c44652534b

Table 1. Use Cases

Case ID	Case Description	Case Cat.	Risk Class
1	AI system to filter unwanted emails and keep them separated from useful ones to reduce time and effort	OB	MR
2	AI system uses emotion recognition to identify/recognize patient's emotions	OB	HR
3	AI system to measure a truck driver's fatigue and playing a sound to push them to drive longer [37]	NO	UR
4	AI systems designed for social robots for children with autism to capture their behavior to assist treatment [22]	NO	HR/LR
5	AI systems for automatic transcription or enhancement of speech [22]	NO	HR/MR
6	AI systems to assess recidivism risk by providing quantitative risk assessments [42]	NO	HR
7	AI system using remote biometric identification of political protesters creates a significant chilling effect on the exercise of freedom of assembly and association	OB	UR
8	AI system that automatically converses with people in place for a human being and can interact with them	OB	LR

For the evaluation, we selected eight use cases: four **Obvious (OB)** and four **Non-obvious (NO)**. Table 1 lists the selected use cases and their risk classes. The *Obvious* cases are *considered in the AI Act* itself. The *Non-obvious* cases are *found in the literature* [22, 37, 42], where they are described as *complicated cases.* The references to the corresponding articles are provided in Table 1. All cases are real-world examples.

To reach potential respondents, an open invitation for the interview was posted on LinkedIn and promoted within relevant research groups and mailing lists. Additionally, a referral strategy was implemented to expand the network of AI experts, particularly those with a legal background. The selection criteria for the respondents were as follows: (1) working in the AI-related fields, (2) residing in the EU region, and (3) employed by an organization or company within the EU. Over 40 personal invitations were sent via email and LinkedIn to individuals who met the aforementioned criteria.

Respondent selection for the interviews was carefully carried out to ensure a comprehensive and diverse spectrum of viewpoints. First and foremost, it was imperative that the respondents possessed expertise in AI-related fields to facilitate an effective evaluation of our DT-based framework. Second, respondents were deliberately selected from both legal and non-legal backgrounds to ensure a more comprehensive grasp of the classification process. The desired number of respondents must be a multiple of 8, selected to correspond with the total number of AI system use cases included in the interview session. This decision allowed us to ensure equal coverage and representation of each use case in the evaluation.

In the end, we managed to recruit 16 participants matching those criteria. Table 2 provides details about each participant.

To prevent potential bias, we rotated the cases considered by participants without and with the DT-based framework as shown in Table 3. This rotation allowed the respondents to classify the same cases with and without the help of the framework. Furthermore, it helped alleviate the "cold start" effect, which typically results in more time being required to process the first case compared to subsequent ones.

We evaluated our framework using three criteria: (1) an increase in classification accuracy; (2) an increase in the classification agreement between respondents (inter-rater agreement); (3) time savings.

The accuracy was measured by comparing all the respondents' responses in the experiment to the ground truth as presented in Table 1. Note that in this case, *we report the results only for the OB cases, because they are mentioned in the AI Act draft and, thus, can be treated as ground truth.*

Table 2. Respondents Profile

Resp. ID	Background	Context	Expertise	AI Act Famil.	Gender	Country
A	Legal	Academia	Human Rights AI and Policy	Yes	Female	Netherlands
B	Non-legal	Academia	AI Financial Technology	No	Female	Netherlands
C	Legal	Academia	Digital Services	Yes	Female	Netherlands
D	Non-legal	Industry	AI in Cybersecurity	No	Male	Netherlands
E	Non-legal	Academia	Generative AI	No	Male	Italy
F	Legal	Academia	Computer Vision Technology	Yes	Female	UK
G	Non-legal	Industry	AI Recommender Systems	No	Male	Germany
H	Non-legal	Academia	Cybersecurity/Ethics	No	Male	Netherlands
I	Legal	Academia	AI in Content Moderation	Yes	Female	Netherlands
J	Non-legal	Academia	Ethics in Autonomous Vehicles	No	Male	Netherlands
K	Legal	Academia	Biometric Identification Systems	Yes	Male	Italy
L	Non-legal	Industry	AI Medical Systems	No	Male	Netherlands
M	Non-legal	Academia	Ethics & Philosophy of AI Technology	No	Female	Netherlands
N	Legal	Academia	AI Conversational Systems	Yes	Male	UK
O	Legal	Academia	Human & Technology in AI	Yes	Male	Netherlands
P	Non-legal	Academia	NLP Researcher of Big Data for Intelligent Society	Yes	Female	UK

Table 3. Experiment Design

		Respondent ID															
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**
Case ID	**W/out DT**	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
		2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1
		3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2
		4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3
	With DT	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4
		6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5
		7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6
		8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7

To estimate the increase in classification agreement between respondents, we evaluate the increase of the inter-rater agreement. Inter-rater agreement refers to the level of consensus or consistency between two or more raters when assessing or categorizing the same set of data or subjects. It measures the extent to which different raters agree on their judgments or evaluations. In this work, we use *Krippendorff's alpha* [26] as an efficient tool for assessing agreement among raters [44]. This metric is suitable for this experiment as it accommodates multiple respondents, multiple subjects (case studies), and missing ratings (as not all participants classified all eight case studies using the DT-based framework) and is often used in studies with similar design [8]. According to Klaus Krippendorff [26], Krippendorff's alpha of 1 implies perfect rater agreement. Values above 0.8 suggest satisfactory agreement, while those between 0.67 and 0.8 indicate tentative conclusions. A score of 0 implies agreement as expected by chance. Scores below 0 show systematic disagreement among raters.

While recognizing the importance of the absolute values of this metric, our emphasis in this work is geared primarily toward its change. To assess the change in inter-rater agreement, we calculated this metric's absolute values for the results of the classification done with the help of the DT-based framework and without, and computed the difference.

So, as we recorded the interviews, we could also evaluate how much time each participant spent evaluating the cases with and without the framework. We used these values to calculate time savings.

Table 4. Improvements in Accuracy and Inter-rater Agreement

	Accuracy, %			Krippendorff's alpha					
	Legal	Non-legal	All	Legal		Non-legal		All	
				OB	NO	OB	NO	OB	NO
With DT:	78.6	66.7	71.9	0.43	0.12	0.45	−0.03	0.44	0.09
W/out DT:	71.4	61.1	65.6	0.51	0.20	0.27	0.16	0.32	0.23
Difference:	7.1	5.6	6.3	−0.08	−0.08	0.18	−0.19	0.12	−0.14

Ethics. We got ethics approval from our Institutional Review Board for this study. From all participants, we got explicit consent for the anonymized processing of the data.

4 Evaluation Results

4.1 Quantitative Results

Table 4 reports the results of accuracy (only OB cases, see Section 3.2 for details) and inter-rater agreement evaluation. Overall, the classification of AI systems using the DT-based framework demonstrated higher accuracy than without DT for all cases. Consequently, we can see that the classification accuracy for Legal, Non-legal and all respondents increases by 7.1%, 5.6%, and 6.3%, respectively. Interestingly, the increase for the respondents with a legal background is higher than for non-legal interviewees, suggesting that the proposed DT-based framework is useful in framing and structuring their knowledge. As expected, individuals with a legal background have categorized the cases more accurately both with and without our framework. This factor may be attributed to legal experts' familiarity with the legal principles, whereas non-legal respondents may be unfamiliar with them. However, this implies that the proposed decision tree has yet to effectively translate legal jargon into more plain language. In any case, the numbers show that the DT-based framework increases the accuracy of classification for both groups of participants, making it practically useful.

Focusing on the inter-rater agreement, the following observations should be mentioned. All recorded values of Krippendorff's alpha fall below the threshold required for drawing tentative conclusions regarding agreement. This suggests a lack of consensus among respondents and their respective sub-groups, underscoring the need for refinement in the classification criteria outlined in the AI Act draft. However, the participants with a legal background tend to agree more often without the framework than when it is used. We assume that they intuitively understand how to classify a particular case even without the framework, while its usage pulls them apart, making them converge less often. At the same time, Non-Legal respondents agree on OB case classification more often with the framework than without it, showing its usefulness for this group. As expected, the respondents have less agreement about the classification of NO cases compared to OB ones, both with and without the framework. That quantitatively confirms their confounding nature. Interestingly, the usage of the framework increased confusion among the interviewees regarding the NO cases.

The disagreement between respondents occurred regarding the Cases 2, 3, 4, 5, and 6, as detailed in Table 5. According to our observations, these disagreements are caused by several key frictions. The first is related to the assessment of whether an AI system violates fundamental rights and/or Union values, which is pivotal in determining whether it falls into the categories of Unacceptable or High Risk. The second discord is due to the polarity in evaluating whether an AI system distorts human behavior or benefits humanity. It is crucial for distinguishing between the Unacceptable-risk and High-risk classes. The third clash category appears thanks to the difference in the understanding of "vulnerable groups" mentioned in the AI Act draft. This difference contributes to

Table 5. Disagreement in Classifying the AI Systems with DT-based Framework

Case ID	Case cat.	Case Description	Disagreement
2	OB	AI system use emotion recognition system to identify/recognize patient's emotion	Nearly half of the respondents believed this system, designed to assist patients, would not significantly breach fundamental rights or Union values. This perception led them to categorize it as Limited Risk. However, the other half acknowledged this AI system potentially violates fundamental rights/Union values, but would also benefit patients, which led them to be classified as High Risk.
3	NO	AI system measuring truck driver's fatigue and playing a sound to push them to drive longer	Arguments supporting Unacceptable Risk centered on the belief that "pushing drivers to drive longer" violates human dignity and distorts human behavior. In contrast, other respondents contended that the AI system would benefit drivers and employers, making it suitable for work environments (Article 6: High Risk).
4	NO	AI systems designed for social robots for children with autism to capture their behavior to assist treatment	Most respondents perceived the AI system to pose significant harm to fundamental rights and/or Union values. In addition, Unacceptable-risk classification was rooted in the assumption that children with autism constitute a vulnerable group, making this AI system, which potentially involved biometric/emotion recognition systems to capture children's behavior, unsuitable for them. Conversely, the Limited-risk classification was derived from respondents' determination that the social robots and their associated benefits would not significantly harm fundamental rights and/or Union values.
5	NO	AI systems for automatic transcription or enhancement of speech	All respondents agreed that the case would not significantly harm human behavior, which led them to not categorize this AI system as Unacceptable Risk/High Risk. However, opinions differed on whether this case should adhere to transparency regulations, with some contending that the AI system's interaction with humans warranted transparency.
6	NO	AI systems to assess recidivism risk by providing quantitative risk assessments	The argument for classifying this case as Unacceptable Risk stems from using this AI system in law enforcement, a context mentioned in Article 5 of the AI Act. Legal respondents believed that recidivists are part of the vulnerable groups outlined in Article 5, which prohibits exploiting such groups. Conversely, others contended that the case should be classified as High Risk due to its applicability in supporting law enforcement authorities. One respondent emphasized the significance of technological details when quantitatively assessing recidivism risk. He believed the quantitative risk assessment, possibly involving biometric systems, was related to a vulnerable group and could cause potential misjudgment.

whether an individual or object within the AI system is classified as Unacceptable Risk or High Risk. Additionally, the question of transparency obligations emerged, prompting consideration of when an AI system should comply with such regulations, affecting whether it is categorized as Limited Risk or Minimal Risk. Last, assessing the foreseeable societal impact and use cases of AI systems also influences their classification.

Table 6 reports the duration of the experiment without (W/out DT) and with (With DT) the DT-based framework per each participant. Initially, we assumed that participants with the framework would spend less time on classification than without it. To test this assumption, we conducted a Wilcoxon signed-rank test with the null hypothesis positing that there is no statistical difference in the medians of the classification duration distributions with and without the DT-based framework, and the alternative that with DT people spend less time. We found no sufficient evidence to reject the null hypothesis ($statistic = 63$, $p = 0.41$). The data reveals that, on average, each participant spent approximately 103.1 s per case with the DT-based framework and 95.5 s per case without it. This indicates that the framework's usage leads to a slightly slower classification on average.

At the same time, the previous finding is not uniform for all the participants. As we can see from the table, participants who are familiar with the AI Act draft spent significantly more time ($statistic = 36$, $p = 0.004$ using the Wilcoxon signed-rank test) on classifying cases with the

Table 6. Duration of the Experiment without and with the DT-based Framework Per Each Participant (LE, Legal; NL, Non-legal)

	Respondent ID															
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**
AI Act Famil.	**Yes**	**No**	**Yes**	**No**	**No**	**Yes**	**No**	**No**	**Yes**	**No**	**Yes**	**No**	**No**	**Yes**	**Yes**	**Yes**
Background	**LE**	**NL**	**LE**	**NL**	**NL**	**LE**	**NL**	**NL**	**LE**	**NL**	**LE**	**NL**	**NL**	**LE**	**LE**	**NL**
W/out DT	57	388	567	742	342	729	302	560	186	737	89	485	189	421	89	228
With DT	113	251	1,586	600	219	765	244	285	395	368	180	209	209	506	190	480

framework than without it. This means that individuals familiar with the AI Act draft intuitively classify the cases much faster than if they need to follow the decision tree choices. At the same time, respondents not familiar with the AI Act draft spent significantly less time ($statistic = 1$, $p = 0.008$) with the framework than without it. Thus, the proposed framework is more useful to the audience unfamiliar with the AI Act draft (or the adopted version of the AI Act). We obtained similar results if we consider the legal and non-legal groups separately, although, in the case of non-legal audience, the p-value is much higher ($statistic = 7, p = 0.037$).

Summing up, our evaluation shows that individuals outside the legal field who are not familiar with the AI Act draft will derive the greatest advantage from utilizing our DT-based framework. With its assistance, they can significantly decrease case classification time while achieving nearly equivalent levels of accuracy and inter-rater agreement compared to legal experts who do not use the framework. Furthermore, we have demonstrated that the chosen non-obvious cases do present classification challenges.

4.2 Qualitative Results

The interviews showed that all respondents found the DT-based framework helpful in categorizing AI systems, citing various benefits associated with its use. First, it presents the information from the AI Act draft in a structured and more concise way. Respondent F noted that the decision tree offered a more concise representation of the AI Act draft compared to the actual text, eliminating unnecessary prompts and including only key points relevant to classifying AI systems, making it particularly beneficial for individuals who were not accustomed to reading legal texts: "*It is helpful because it just gives a much more concise framework than the actual text of the AI act, which can be useful for people who are not used to reading legal text.*" Respondent M echoed this sentiment, emphasizing the benefits of considering different circumstances instead of lengthy paragraphs: "*Decision tree is helpful, instead of more paragraph. More beneficial to read several circumstances.*" According to Respondent L, the framework facilitated a quick and comprehensive understanding of what is allowed and what is not, enabling a rapid overview of the AI Act draft: "*I like the fact that the things that are not allowed are on one side. So you can really easily see them, and I think it also allows you to skim the entire thing quite quickly.*" Second, the participants acknowledge that the framework simplifies the classification process. Respondent A stated that with the enforcement of the AI Act, such a model becomes increasingly important as it simplifies the classification process and logically assigns risk levels to AI systems: "*I think it was definitely a very nice research on interpreting the risk assessment model because I feel like we do need something like that[...]. So it really simplifies things a lot. It makes it easier. It helps you logically place your answer.*" Respondent B noted that the framework provided a clear and structured perspective on the regulation, making it easier to classify AI systems: "*As I explained before, this classification tree really helps me to make my way of thinking in a more structured way in comparison with just reading the acts[...]. This helps us to see the thing in more clear perspective and in a more structured way.*" Third, the framework increases awareness. Respondent C highlighted that the decision tree enhanced understanding of the AI Act draft and the associated risk classes. Respondent I highlighted that it increased awareness of the risks associated with AI systems, which might be overlooked when reading the AI Act draft articles. Respondents E and G, who possess technical expertise, emphasized the significance of understanding the AI Act draft through the decision tree to comprehend the boundaries of risks: "*So it helpful in a case for you as developer to understand on the boundaries of the risk.*"

The respondents also expressed some concerns regarding the framework. First, despite its usefulness in categorizing AI systems, Respondent A emphasized the importance of consulting legal experts to ensure that the output and explanations align with the AI Act draft. Respondent G echoed this sentiment, indicating that while they were satisfied with the decision tree, they would still consult the law to verify the correctness of the classification: "*I would rather consult with the legal team. Yeah, and not relying completely to the decision tree itself because it fails for some. Yeah, for some scope and for some edge cases.*" Second, Respondents A and F expressed concerns regarding the level of background knowledge required to comprehend the AI Act draft. This issue is particularly relevant for individuals with technical background who may not fully grasp the underlying context and reasons behind the regulations. Third, the classification of AI systems using the framework may be influenced by the personal biases of the respondents. The individual judgments and perspectives of legal and ethics domain experts can differ in their perception of AI systems based on their assessment of risks. For instance, Respondent B notes: "*So my decision could be biased. It could be influenced by my personal judgment.*"

Finally, the respondents provided some recommendations on how to improve the framework and its evaluation. First, to address the difficulty in understanding legal terms and definitions, it was advised to incorporate more concrete definitions and examples: "*You just need to add more text and maybe have some pointers to summary of the laws, but if it's not listing all possible use cases then it should be saying this clearly,*" Respondent P noted. Both legal and non-legal experts expressed difficulties in interpreting certain terms in the regulation, such as "significant harm" or "fundamental rights." For instance, Respondent F notes: "*Things like significant harm to fundamental rights, what does that mean? When does it become significant harm? [...] My guess is that there might be a bit more guidance on this.*" Second, the participants recommended including in the decision tree clearer indications of their progress or the specific assessment they are conducting. For instance, Respondents C and L suggested adding high-level information to inform users whether they are assessing, e.g., an Unacceptable or High-risk system. Third, there is a need to decouple the decision tree interpretation from an expert's background. For instance, during the interviews, several respondents with ethics expertise tended to classify all AI systems as prohibited. Finally, some respondents advised improving use case descriptions; otherwise, they have to make some assumptions. They suggested including information about the type of data used; social context; output; technical details; distribution of responsibility; purpose and intention: "*[...]what you need to know is the purpose of this system, the intention of the system and, after that, what is the output of this system and also what is the information that will be stored there?*" (Respondent O).

5 Discussion

Our research asserts a positive reception of the initiative aimed at enhancing the transparency and clarity of regulations. Such initiative entails translating legal terminology and context into accessible language, catering to individuals from diverse backgrounds. In our framework, we used the terminology and contexts as defined in the AI Act draft, potentially making it less understandable for the general public. However, even with this limitation, our framework was warm-welcomed by non-legal individuals. Still, in future work, it would be interesting to investigate if its performance and usefulness could be further improved by using adapted terminology, contexts and language.

Table 7. Key Changes in Annex III in the Adopted Version of the AI Act

Annex III	Key changes in the Annex	Description of key changes
1. Biometric systems	(a), (b), (c)	Remote biometric identification systems; (intended) biometric categorization; emotion recognition.
2. Critical infrastructure	(a)	Safety components of critical digital infrastructure.
3. Education and vocational training	(b), (c), (d)	Evaluate learning outcomes, steer the learning process; assessing the appropriate level of education; monitoring and detecting prohibited behavior of students during tests.
4. Employment	(a), (b)	Recruitment or selection and targeted job advertisements (including analyzing and filtering job applications), and evaluation; decisions affecting terms of the work, promotion and termination, allocate tasks and monitor and evaluate performance.
5. Access to essential services and benefits	(a), (b), (c), (d)	Evaluate the eligibility for public assistance benefits and services, grant, reduce, revoke, or reclaim such; evaluate the creditworthiness or establish their credit score (exception: financial fraud); evaluate and classify emergency calls or to dispatch, or establish priority in dispatching of emergency services (police, firefighters and medical aid, emergency healthcare patient triage systems); risk assessment and pricing in the case of life and health insurance.
6. Law enforcement	(a), (e)	Assess the risk of becoming a victim of criminal offenses; profiling in the course of detection, investigation or prosecution of criminal offenses.
7. Migration, asylum and border control	(d), (e)	Applications for asylum, visa and residence permits and associated complaints to the eligibility and assessment of the reliability of the evidence; migration, asylum and border control management detecting, recognizing or identifying natural persons (exception of verification of travel documents).
8. Justice and democratic processes	(a), (b)	Researching and interpreting facts and the law and in applying the law to a concrete set of facts or dispute resolution; influencing the outcome of an election or referendum or the voting behavior (exemption: persons are not directly exposed, e.g., tools used to organize, optimize and structure political campaigns).

Moreover, conducting research in more specific domains and targeting specific user groups could provide valuable insights and contribute to a more user-centric approach, thus leading to even higher effectiveness of the framework within specific industries or sectors.

While pursuing this study, we experienced several hurdles. We discovered that it is very challenging to isolate all potential factors influencing the evaluation results. For instance, the order of the cases may impact a participant's final choice and the time spent classifying a case. Similarly, the fact that all the respondents, at first, evaluate cases without the framework and then with it may also affect the results. Moreover, respondents also learn during the evaluation, and it is not clear how to confine this factor as well. In this work, we followed the initial protocol described in Section 3.2; however, we recommend researchers doing similar studies to pay particular attention to these aspects.

The AI Act has also been evolving [13]. Our research was conducted using the AI Act draft, which subjected our results to the legal framework proposed in this version. However, on March 13, 2024, the European Parliament agreed upon the adopted version of the AI Act [15]. According to the document [14], the risk classification within the AI Act changed significantly. First, while the draft included four risk classes, the adopted version simplifies the risk-based classification, proposing only three risk classes: *Unacceptable*, *High*, and *Limited Risk*. Correspondingly, the classification criteria were also updated significantly. For instance, Table 7 outlines significant changes between the two versions of the law in Annex III regarding *High-risk* classification.

Second, the adopted version introduces a new category of AI systems, ***General Purpose AI (GPAI)*** systems (see Articles 3(63), 3(66), and 51–54). For this new category, it creates a new risk class *Systemic Risk* (see Article 3(65) in conjunction with Article 3(64)), which further divides the

Table 8. Summary of Key Changes between the Draft and Adopted Versions of the AI Act Relevant to the DT-based Framework Design

Change category	Draft Version	Adopted Version
AI systems	AI system	AI system, General-Purpose AI system (GPAI) **Sources:** Articles 3(1), 3(63), 3(66), 51–55
GPAI	Not included	Transparency obligations, documentation, copyright protection, summary about model **Sources:** Articles 3(63), 3(66)
GPAI with Systemic Risk	Not included	Classification, high impact capabilities, model evaluation, risk mitigation, report, cybersecurity, code of practice **Sources:** Articles 3(64), 3(65), 3(67), 55
Risk classes	UR, HR, LR, MR **Sources:** Articles 5, 6, 52, 69, and Annex III	UR (prohibited practices), HR, LR, systemic risk (SR) **Sources:** Articles 3(2), 3(65), 5, 6, 51, 52, and Annex III

GPAI systems into *GPAI Models* and *GPAI Models with Systemic Risk* (see Article 51). These changes should help the regulation to embrace fast development of generative AI models, like OpenAI's ChatGPT [35], Google's Gemini [18], or Meta's LLama [32], which have appeared within several months after the draft version of the AI Act was published.

At the same time, the adopted version of the AI Act excludes open-source models from consideration under the law according to Articles 2(12) and 54(5), unless they belong to *Prohibited*, *High-risk* AI systems (Articles 5, 50) or *GPAI Systems with Systematic Risk* (Article 54(5)). Additionally, the adopted version of the AI Act also includes new and more detailed categories to define the application and scope of AI systems and their exemptions, e.g., Article 6(3) exempts AI systems that would not pose a significant risk of harming the health, safety, or fundamental rights.

Table 8 summarizes the main changes between the draft and the adopted version of the AI Act that would require adjustments to our DT-based framework. For example, the tree would need a novel root question distinguishing AI and GPAI systems. GPAI systems would additionally need to be differentiated between plain GPAI systems and those that pose a Systemic Risk according to Article 3(65) of the AI Act. Questions addressing *High-risk* systems would need to include the exemption of Article 6(3). Furthermore, questions querying the risk domains relating to Annex III would need to be adapted for the new text version (see Table 7).

Until the final version of the AI Act is published, its variants would most probably require updating our DT-based framework. Nonetheless, our research makes a significant contribution to the ongoing debate surrounding the classification of AI systems under the AI Act. Moreover, it advances the methodology on how to construct and evaluate similar frameworks.

6 Related Work

Understandably, the literature related to the AI Act is still in its early stages. Veale and Zuiderveen Borgesius [43] made one of the first attempts to overview the AIA draft and analyze its potential implications. At almost the same time, Ebers et al. [9] performed another analysis of the law, concentrating on key aspects related to prohibited-, high- and limited-risk categories. Additionally, they studied the AI Act in relation to the existing regulations, identifying the misalignments between them. Nevertheless, the majority of the academic discussions surrounding the AI Act concentrate on the potential impact the law might have on specific industries or use cases. For instance, Lim et al. [29] test the AI Act and its effects by applying it to several social issues that materialized as the result of AI usage. Van Dijck [42] assesses the AI Act's applicability within the domain of the criminal justice system in predicting recidivism risk. Marano and Li [31] evaluate the potential risks associated with the usage of robo-advisors in the insurance industry, while Hupont et al. [22] discuss facial recognition technology use case.

Another group of researchers argues that the current classification approach is vague and limited. Lim et al. [29] use several examples to show that clear classification of AI systems is hard and probably requires separate regulations for different domains. Barkane [3] suggests adapting the proposed classification process due to the various exceptions and ambiguities. Other voices in academia [34] see the problem in the four risk categories that are interrelated and offer various possible classification combinations for the same single case. The same issue of the interrelatedness of the AI Act's categories was also mentioned by Orlando [36], who stressed the blurring boundaries between *High-* and *Minimal-risk* systems. Hupont et al. [22] came to the same conclusion, showing on examples that many *High-risk* systems fall under both *High-* and *Limited-risk* categories.

Researchers have also been suggesting improvements related to the AI Act. Mökander et al. [33] underscore the necessity of converting vague and unclear definitions and concepts into transparent, formal and certifiable criteria. They prove that this transformation is vital to perform univocal conformity assessments in line with the regulation. Hacker [19] provides concrete recommendations on how to use personal data to train AI systems in accordance with the GDPR. Sovrano et al. [39] surveys methods and metrics that can be used to audit and assess the conformity of AI systems to the proposed regulation.

Some approaches, similar to our framework in terms of supporting the decision-making process, were also introduced within the context of other regulations, like the GDPR [10]. For instance, Agrawal et al. proposed legislative approaches to make the GDPR machine-readable [1] and introduced compliance assessment tools [2]. Leicht et al. developed frameworks to support privacy policy authors in the process of becoming compliant with the GDPR [27], while Kingston [25] considered solutions for how AI can be used to assist within compliance checks.

7 Conclusion

This work presents the design and evaluation results of the DT-based framework aimed at making the risk classification of AI systems, under the AI Act draft, more transparent, accurate, and available to the audience without a legal background. The quantitative evaluation shows that the individuals not familiar with the AI Act draft benefit the most from our framework usage: It enables them to reduce significantly the time required to classify a case and to improve the classification accuracy. At the same time, the qualitative study shows that all the participants found the framework useful.

References

[1] Sushant Agarwal, Sabrina Kirrane, and Johannes Scharf. 2017. Modelling the General Data Protection Regulation. In *Proceedings of the Internationales Rechtsinformatik Symposion (IRIS'17).*

[2] Sushant Agarwal, Simon Steyskal, Franjo Antunovic, and Sabrina Kirrane. 2018. Legislative compliance assessment: Framework, model and GDPR instantiation. In *Privacy Technologies and Policy*. Vol. 11079. Springer International Publishing, 131–149. https://doi.org/10.1007/978-3-030-02547-2_8

[3] Irena Barkane. 2022. Questioning the EU proposal for an artificial intelligence act: The need for prohibitions and a stricter approach to biometric surveillance. *Info. Polity* 27, 2 (July 2022), 147–162. https://doi.org/10.3233/ip-211524

[4] William A. Belson. 1959. Matching and prediction on the principle of biological classification. *Appl. Stat.* 8, 2 (1959), 65–75. https://doi.org/10.2307/2985543

[5] Michael Chui, Eric Hazan, Roger Roberts, Alex Singla, Kate Smaje, Alex Sukharevsky, Lareina Yee, and Rodney Zemmel. 2023. The Economic Potential of Generative AI: The Next Productivity Frontier. Retrieved 2024-03-11 from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

[6] G. M. Cramer, R. A. Ford, and R. L. Hall. 1976. Estimation of toxic hazard—A decision tree approach. *Food Cosmetics Toxicol.* 16, 3 (1976), 255–276.

[7] Sylvie Delacroix and Ben Wagner. 2021. Constructing a mutually supportive interface between ethics and regulation. *Comput. Law Secur. Rev.* 40 (Apr. 2021), 105520. https://doi.org/10.1016/j.clsr.2020.105520

[8] Natasha R. Donnolley, Georgina M. Chambers, Kerryn A. Butler-Henderson, Michael G. Chapman, and Elizabeth Sullivan. 2019. A validation study of the Australian maternity care classification system. *Women Birth* 32 (6 2019), 204–212. Issue 3.

[9] Martin Ebers, Veronica R. S. Hoch, Frank Rosenkranz, Hannah Ruschemeier, and Björn Steinrötter. 2021. The European Commission's proposal for an artificial intelligence act—A critical assessment by members of the robotics and AI law society (RAILS). *J* 4, 4 (2021), 589–603.

[10] European Commission. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance). Retrieved 2024-03-10 from https://eur-lex.europa.eu/eli/reg/2016/679/oj

[11] European Commission. 2020. White Paper on Artificial Intelligence: A European Approach to Excellence and Trust. Retrieved 2024-03-11 from https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

[12] European Commission. 2021. 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. Retrieved 2023-03-25 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

[13] European Commission. 2023. Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021–2021/0106(COD)). Retrieved 2024-03-25 from https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

[14] European Commission. 2024. Adopted version of the Artificial Intelligence Act. Retrieved 2024-03-18 from https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

[15] European Parliament. 2024. Artificial Intelligence Act: MEPs Adopt Landmark Law. Retrieved 2024-03-18 from https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law

[16] European Union. [n.d.]. Aims and Values. Retrieved 2024-03-11 from https://european-union.europa.eu/principles-countries-history/principles-and-values/aims-and-values_en

[17] Giulia Gentile and Orla Lynskey. 2022. Deficient by design? The transnational enforcement of the GDPR. *Int. Comp. Law Quart.* 71, 4 (Oct. 2022), 799–830.

[18] Google. 2023. Gemini. Retrieved 2024-03-12 from https://gemini.google.com/

[19] Philipp Hacker. 2021. A legal framework for AI training data—From first principles to the artificial intelligence act. *Law, Innovat. Technol.* 13, 2 (2021), 257–301.

[20] Hilmy Hanif, Jorge Constantino, Marie-Therese Sekwenz, Michel van Eeten, Jolien Ubacht, Ben Wagner, and Yury Zhauniarovich. 2024. Navigating the EU AI Act Maze Using a Decision Tree Approach: Decision Tree and Interview Protocol. Retrieved from https://data.4tu.nl/datasets/bf7013ce-54b5-43b9-b275-f6c44652534b/1. https://doi.org/10.4121/bf7013ce-54b5-43b9-b275-f6c44652534b.v1

[21] Hilmy Hanif, Jorge Constantino, Marie-Therese Sekwenz, Michel van Eeten, Jolien Ubacht, Ben Wagner, and Yury Zhauniarovich. 2023. Tough decisions? supporting system classification according to the AI act. In *Proceedings of the International Conference on Legal Knowledge and Information Systems*. IOS Press, 353–358. https://doi.org/10.3233/faia230987

[22] Isabelle Hupont, Songül Tolan, Hatice Gunes, and Emilia Gómez. 2022. The landscape of facial processing applications in the context of the european AI act and the development of trustworthy systems. *Sci. Rep.* 12, 1 (June 2022). https://doi.org/10.1038/s41598-022-14981-6

[23] Paul Johannesson and Erik Perjons. 2021. *An Introduction to Design Science* (2nd ed.). Springer International Publishing.

[24] Ronit Justo-Hanani. 2022. The politics of artificial intelligence regulation and governance reform in the european union. *Policy Scie.* 55, 1 (Mar. 2022), 137–159. https://doi.org/10.1007/s11077-022-09452-8

[25] John Kingston. 2017. Using artificial intelligence to support compliance with the general data protection regulation. *Artific. Intell. Law* 25, 4 (Dec. 2017), 429–443.

[26] Klaus Krippendorff. 2019. *Content Analysis: An Introduction to Its Methodology*. Sage Publications. https://doi.org/10.4135/9781071878781

[27] Jens Leicht, Maritta Heisel, and Armin Gerl. 2022. PriPoCoG: Guiding policy authors to define GDPR-compliant privacy policies. In *Trust, Privacy and Security in Digital Business*, Sokratis Katsikas and Steven Furnell (Eds.). 1–16. https://doi.org/10.1007/978-3-031-17926-6_1

[28] Andreas Liebl and Till Klein. 2023. AI Act: Risk Classification of AI Systems from a Practical Perspective. Retrieved 2024-03-11 from https://www.appliedai.de/assets/files/AI-Act_WhitePaper_final_CMYK_ENG.pdf

[29] Eugene Lim, Hyesung Park, Beopyeon Kim, and Beopyeon Kim. 2022. Review of the validity and rationality of artificial intelligence regulation: Application of the EU's AI regulation bill to accidents caused by artificial intelligence. In *Proceedings of the International FLAIRS Conference*, Vol. 35.

[30] Britt Marie Lindgren, Berit Lundman, and Ulla H. Graneheim. 2020. Abstraction and interpretation during the qualitative content analysis process. *Int. J. Nurs. Studies* 108 (Aug. 2020).

[31] P. Marano and S. Li. 2023. Regulating robo-advisors in insurance distribution: Lessons from the insurance distribution directive and the AI act. *Risks* 11, 1 (2023).

[32] Meta. 2023. Llama. Retrieved 2024-03-12 from https://llama.meta.com/

[33] Jakob Mökander, Maria Axente, Federico Casolari, and Luciano Floridi. 2021. Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed european AI regulation. *Minds Mach.* 32, 2 (Nov. 2021), 241–268. https://doi.org/10.1007/s11023-021-09577-4

[34] Rostam J. Neuwirth. 2023. Prohibited artificial intelligence practices in the proposed eu artificial intelligence act (AIA). *Comput. Law Secur. Rev.* 48 (2023), 105798.

[35] OpenAI. 2022. ChatGPT. Retrieved 2024-03-12 from https://chat.openai.com

[36] Alberto Orlando. 2022. AI for sport in the EU legal framework. In *Proceedings of the IEEE International Workshop on Sport, Technology and Research (STAR'22)*. 100–105.

[37] PwC. 2022. European Artificial Intelligence Act: Many Procedural and Substantive Requirements. Retrieved 2024-03-11 from https://www.pwc.nl/en/insights-and-publications/themes/digitalization/european-artificial-intelligence-act-many-procedural-and-substantive-requirements.html

[38] Hannah Ruschemeier. 2023. AI as a challenge for legal regulation—The scope of application of the artificial intelligence act proposal. *ERA Forum* 23, 3 (Feb. 2023), 361–376.

[39] Francesco Sovrano, Salvatore Sapienza, Monica Palmirani, and Fabio Vitali. 2021. A survey on methods and metrics for the assessment of explainability under the proposed AI act. In *Proceedings of the International Conference on Legal Knowledge and Information Systems (JURIX'21) (Frontiers in Artificial Intelligence and Applications, Vol. 346)*. 235–242.

[40] A. Zh. Stepanyan. 2021. European artificial intelligence act: Should Russia implement the same? *Kutafin Law Rev.* 8, 3 (2021), 403–422.

[41] Inga Ulnicane. 2022. *Artificial Intelligence in the European Union.* Routledge, 254–269. https://doi.org/10.4324/9780429262081-19

[42] Gijs van Dijck. 2022. Predicting recidivism risk meets AI act. *Eur. J. Crim. Policy Res.* 28, 3 (2022), 407–423.

[43] Michael Veale and Frederik Zuiderveen Borgesius. 2021. Demystifying the draft EU artificial intelligence act – analysing the good, the bad, and the unclear elements of the proposed approach. *Comput. Law Rev. Int.* 22, 4 (2021), 97–112.

[44] Antonia Zapf, Stefanie Castell, Lars Morawietz, and André Karch. 2016. Measuring inter-rater reliability for nominal data—Which coefficients and confidence intervals are appropriate? *BMC Med. Res. Methodol.* 16 (2016), 1–10.

Received 22 November 2023; revised 27 March 2024; accepted 3 May 2024