*Legal Knowledge and Information Systems*
*G. Sileno et al. (Eds.)*


*doi:10.3233/FAIA230987*

# Tough Decisions? Supporting System Classification According to the AI Act

Hilmy HANIF [a], Jorge CONSTANTINO [a], Marie-Therese SEKWENZ [a], Michel VAN EETEN [a], Jolien UBACHT [a], Ben WAGNER [a], and Yury ZHAUNIAROVICH [a]

[a] *Delft University of Technology*

**Abstract.** 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, linking varying degrees of compliance obligations to the system's classification. 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 those shortcomings, we propose a Decision-Tree-based framework aimed at increasing robustness, legal compliance and classification clarity with the Regulation. Quantitative evaluation shows that our framework is especially useful to individuals without a legal background, allowing them to improve considerably the accuracy and significantly reduce the time of case classification.

**Keywords.** Artificial Intelligence, AI Act, AIA, Risk Classification, Compliance

## 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. However, it is imperative to acknowledge that while these technologies yield significant benefits, their (uncontrolled) use can also bring detrimental consequences. The European Artificial Intelligence Act (AI Act, AIA) represents a significant legislative effort by the European Union (EU) to govern AI systems [1]. It aims to ensure the trustworthiness of AI systems, aligning their deployment with fundamental rights and the European values [2,3]. 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. The AI Act establishes a four-level risk classification guideline for AI systems: *Unacceptable Risk (UR)*, *High-Risk (HR)*, *Limited Risk (LR)*, and *Minimal Risk (MR)*.

Most academic discussions surrounding the AI Act focus on its potential impact on specific industries or use cases. Lim et al. [4] apply the AI Act to cases such as accidents or incidents that emerged as social problems by AI, while van Dijck [5] assesses its use in the criminal justice system. Marano & Li [6] evaluate the potential risks associated with insurance, while Hupont et al. [7] discuss facial processing in the context of the AI Act. Veale and Zuiderveen Borgesius [8] overviewed the AIA draft and analyzed its po-

tential implications; Sovrano et al. [9] discuss the metrics used to measure explainability and AI Act. However, it has been shown that the classification criteria mentioned in the AI Act are themselves unclear [10,7]. Certain AI systems may fall under more than one risk category [7]. According to [10], 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. Besides, the inaccurate classification poses another risk to compliance and requires additional specialized knowledge in the field [11]. Moreover, even if an expert judgment is obtained, this opinion might be subjective, opaque to the public, and could deviate by court jurisprudence. Lim et al. [4] argue that clear AI systems' classification is necessary to manage separate regulatory legislation. Barkane [12] suggests that the proposed classification should be reconsidered since there are multiple exceptions and loopholes. The inherent interrelatedness and the numerous possibilities for classification are problematic [13]. Mökander et al. [14] highlight the need to translate vague concepts into verifiable criteria and to strengthen institutional safeguards concerning conformity assessments based on internal checks. All these reasons stress the point made by [10], calling for standardization and clearer guidance for AI systems classification.

Therefore, 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 risk categories in accordance with the AI Act. In a quantitative assessment, we present the utility of our framework, particularly for individuals not familiar with the law, enabling them to achieve a substantial enhancement in accuracy while significantly reducing case classification time. *We share the DT-based framework*[1] so that the community can validate, use and improve it.

## 2. Methodology

In this work, we followed a mixed-methods approach. For designing the DT-based framework, we employed Design Science Methodology (DSM) that enables the systematic study and creation of artifacts to address practical problems [15]. To evaluate the artifact, we designed and executed a protocol that allowed us to obtain both quantitative results of performance increase achieved with the framework.

**Framework Design.** We started the framework design process by doing a **thematic** analysis of the AIA draft [1]. The goal of this analysis, done following the recommendations proposed by Lindgren et al. [16], was to gain an understanding of the criteria allowing to attribute an AI system to a particular risk class. These criteria were crucial for generating the proposed 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

[1] https://drive.google.com/file/d/1ioWG2ceoV7XdTt6t84nYepGMwLOsMV-z/view

in a specific domain, such as education, the workplace, critical infrastructure, etc., 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 use cases or technologies where AI cannot be used. To simplify the usage of our framework, we also added additional information blocks that influence a decision-making process. Thus, our framework is not a decision tree but closely resembles it; therefore, we call it a DT-based framework.

**Evaluation Design.** For our evaluation, we selected 8 use cases, described in Table 1: 4 Obvious (OB), considered in the AI Act; and 4 Non-Obvious (NO), referenced in the literature as complicated cases (see references in the table).

**Table 1.** Use cases and the duration of their evaluation

| Case ID | Case Description | Case Cat. | Risk Class | Exp. Duration, s | | | |
|---|---|---|---|---|---|---|---|
| | | | | W/out DT | | With DT | |
| | | | | Mean | Med. | Mean | Med. |
| 1 | AI system to filter unwanted emails and keep them separated from useful ones to reduce time and effort | OB | MR | 83.5 | 78.5 | 102.9 | 66.6 |
| 2 | AI system uses emotion recognition to identify/recognize patient's emotions | OB | HR | 79.3 | 91.0 | 105.1 | 57.9 |
| 3 | AI system to measure a truck driver's fatigue and playing a sound to push them to drive longer [17] | NO | UR | 75.7 | 63.8 | 114.0 | 80.0 |
| 4 | AI systems designed for social robots for children with autism to capture their behavior to assist treatment [7] | NO | HR/LR | 102.6 | 90.3 | 131.1 | 58.8 |
| 5 | AI systems for automatic transcription or enhancement of speech [7] | NO | HR/MR | 87.4 | 69.3 | 104.3 | 93.5 |
| 6 | AI systems to assess recidivism risk by providing quantitative risk assessments [5] | NO | HR | 107.7 | 89.5 | 114.8 | 67.0 |
| 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 | 121.1 | 111.1 | 77.6 | 67.0 |
| 8 | AI system that automatically converses with people in place for a human being and can interact with them | OB | LR | 106.6 | 95.0 | 75.1 | 66.1 |

To participate in the experiment, we recruited 16 participants: 7 with legal (Legal) and 9 with technical background (Non-Legal). Table 2 provides details (background and familiarity with the AI Act) about each participant. We developed a protocol[2], and with each participant, we ran an evaluation session divided into three sections: two experimental 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. 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. In the second section, they utilized the proposed framework. 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 sessions were conducted through a video call and recorded.

The cases considered by participants without and with the DT-based framework rotated as showed in the "Use Cases per Resp[ondent]" column in Table 2. 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) increase in classification accuracy; 2)

[2] https://drive.google.com/file/d/1Ubl7vB39vMM-TU2nQWDCLhM7TZQ_tbd6/view

**Table 2.** Respondents, experiment design and duration of evaluation by each respondent

| Resp. ID | Background | AI Act Famil. | Use Cases per Resp. | | | | | | | | Exp. Duration, s | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | W/out DT | | | | With DT | | | | W/out DT | With DT |
| A | Legal | Yes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 57 | 113 |
| B | Non-Legal | No | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 388 | 251 |
| C | Legal | Yes | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 567 | 1586 |
| D | Non-Legal | No | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 742 | 600 |
| E | Non-Legal | No | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 342 | 219 |
| F | Legal | Yes | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 729 | 765 |
| G | Non-Legal | No | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 302 | 244 |
| H | Non-Legal | No | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 560 | 285 |
| I | Legal | Yes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 186 | 395 |
| J | Non-Legal | No | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 737 | 368 |
| K | Legal | Yes | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 89 | 180 |
| L | Non-Legal | No | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 485 | 209 |
| M | Non-Legal | No | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 189 | 209 |
| N | Legal | Yes | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 421 | 506 |
| O | Legal | Yes | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 89 | 190 |
| P | Non-Legal | Yes | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 228 | 480 |

increase of inter-rater agreement; 3) time savings. Note that in the case of accuracy, *we report the results only for the OB cases because they are mentioned in the AI Act and, thus, can be treated as ground truth.*

**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.

## 3. Evaluation Results

Table 3 reports the results of accuracy (only OB cases) 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 respondents and overall 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 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 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. First, the participants with a legal background tend to agree more often without the framework than if 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 to converge less often. Second, Non-Legal respondents agree on OB case classification more often with the framework than without it. 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.

**Table 3.** Improvement of Accuracy and Inter-rater Aggreement

| Case | 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 |

Finally, we also evaluated the time spent by the participants on the classification. Table 2 reports the duration of the first (without DT) and second (with DT) sections per participant. Surprisingly, as we can see from the table, participants who are familiar with the AI Act spent significantly less time ($p = 0.008$ using the Wilcoxon signed-rank test) on classifying cases without the framework than with it. This means that individuals familiar with the AI Act 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 spent significantly less time ($p = 0.016$) with the framework than without it. Thus, the proposed framework is more useful to the audience unfamiliar with the AI Act. Table 1 reports the mean and median time spent by all the participants per case. First, it can be seen that the amount of time spent per case is equal to 103.1 and 95.5 seconds, on average, with and without the DT-based framework, respectively. Thus, on average, the usage of the framework makes the classification a bit slower. However, as we have shown earlier, this difference is caused by the participants familiar with the AI Act, who make the classification intuitively and, thus, considerably faster.

Summing up, our evaluation shows that individuals outside the legal field who are not familiar with the AI Act 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. Discussion and Conclusion

In this work, we aimed to make the risk classification of AI systems under the AI Act more transparent, accurate, and available by translating legal terminology and context into accessible language and technical design, catering to individuals from diverse backgrounds. The evaluation shows that our framework provides benefits to its users: it enables people unfamiliar with the AI Act to reduce significantly the time required to classify a case and to improve the classification accuracy. However, it still has some limitations. First, in our framework, we used the terminology and contexts as defined in the AI Act, potentially making it less understandable for the general public. Therefore, we propose to investigate potential changes in the regulation regarding terminology, contexts, and language design in future work. Second, conducting research in more concrete domains and targeting specific user groups could provide valuable insights and contribute to better alignment with a user-centric approach, leading to even higher effectiveness of the framework within those industries or sectors. Furthermore, we experienced several challenges, e.g., how to isolate all potential factors influencing the evaluation results like the order of the cases, and their impact on a participant's final choice or the time spent to

classify a case. Similarly, the order of experiments may also affect the outcomes. Also, respondents learn during the evaluation that could influence the results. Additionally, the AI Act has also been evolving. Our research was conducted before the final amendments of the AI Act, and we see the need for future work addressing those changes [18].

## References


[1] 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; 2021. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

[2] White Paper on Artificial Intelligence: a European Approach to Excellence and Trust; 2020. Available from: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

[3] Explanatory Memorandum to COM(2021)206 - Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts; 2021.

[4] Lim E, Park H, Kim B, Kim B. Review of the Validity and Rationality of Artificial Intelligence Regulation: Application of the EU's AI Regulation Bill to Accidents Caused by Artificial Intelligence. The International FLAIRS Conference Proceedings. 2022 May;35.

[5] van Dijck G. Predicting Recidivism Risk Meets AI Act. European Journal on Criminal Policy and Research. 2022;28(3):407-23.

[6] Marano P, Li S. Regulating Robo-Advisors in Insurance Distribution: Lessons from the Insurance Distribution Directive and the AI Act. Risks. 2023;11(1).

[7] Hupont I, Tolan S, Gunes H, Gómez E. The landscape of facial processing applications in the context of the European AI Act and the development of trustworthy systems. Scientific Reports. 2022;12(1).

[8] Veale M, Borgesius FZ. Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International. 2021;22(4):97-112.

[9] Sovrano F, Sapienza S, Palmirani M, Vitali F. A Survey on Methods and Metrics for the Assessment of Explainability Under the Proposed AI Act. In: Legal Knowledge and Information Systems - JURIX 2021. vol. 346 of Frontiers in Artificial Intelligence and Applications; 2021. p. 235-42.

[10] Liebl A, Klein T. AI Act: Risk Classification of AI Systems from a Practical Perspective; 2023. Available from: https://www.appliedai.de/en/hub-en/ai-act-risk-classification-of-ai-systems-from-a-practical-perspective.

[11] Ruschemeier H. AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal. ERA Forum. 2023 Feb;23(3):361-76.

[12] Barkane I. Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance. Information Polity. 2022;27(2):147-62.

[13] Neuwirth RJ. Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA). Computer Law and Security Review. 2023;48:105798.

[14] Mökander J, Axente M, Casolari F, Floridi L. Conformity Assessments and Post-market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation. Minds and Machines. 2022;32(2):241-68.

[15] Johannesson P, Perjons E. An Introduction to Design Science. 2nd ed. Springer; 2021.

[16] Lindgren BM, Lundman B, Graneheim UH. Abstraction and Interpretation During the Qualitative Content Analysis Process. International Journal of Nursing Studies. 2020 8;108.

[17] European Artificial Intelligence Act: many procedural and substantive requirements; 2022. Available from: https://www.pwc.nl/en/insights-and-publications/themes/digitalization/european-artificial-intelligence-act-many-procedural-and-substantive-requirements.html.

[18] 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)); 2023.