AI Certification and Assessment Catalogues: Practical Use and Challenges in the Context of the European AI Act

GREGOR AUTISCHER, Graz University of Technology, Austria
KERSTIN WAXNEGGER, Know Center Research GmbH, Austria
DOMINIK KOWALD, Know Center Research GmbH & Graz University of Technology, Austria

Certifying artificial intelligence (AI) systems remains a complex task, particularly as AI development has moved beyond traditional software paradigms. We investigate the certification of AI systems, focusing on the practical application and limitations of existing certification catalogues, by attempting to certify a publicly available AI system. We aim to evaluate how well current approaches work to effectively certify an AI system, and how AI systems, that might not be actively maintained or initially intended for certification, can be selected and used for a sample certification process. Our methodology involves leveraging the Fraunhofer AI Assessment Catalogue as a comprehensive tool to systematically assess an AI model's compliance with certification standards, focusing on reliability and fairness. We find that while the catalogue effectively structures the evaluation process, it can also be cumbersome and time-consuming to use. We observe the limitations of an AI system that has no active development team any more and highlight the importance of complete system documentation. Finally, we identify some limitations of the used certification catalogues and propose ideas on how to streamline the certification process.

Keywords: Algorithmic Auditing, Artificial Intelligence, Certification Catalogues, Algorithmic Fairness, AI Reliability

Reference Format:
Gregor Autischer, Kerstin Waxnegger, and Dominik Kowald. 2025. AI Certification and Assessment Catalogues: Practical Use and Challenges in the Context of the European AI Act. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAF'25).* Proceedings of Machine Learning Research, 7 pages.

1 Introduction and Motivation

Artificial intelligence (AI) has evolved, over several decades, to complex machine learning (ML) systems. Recently, AI systems were increasingly integrated into our daily lives, moving from specialized research labs to mainstream applications [7]. Today, AI is utilized in many domains such as healthcare [30], human resources [41], industry [10], research [14], finance [6, 13], and in prominent technologies such as recommender systems for e.g., social networks [18, 22, 23]. AI now plays a significant role in shaping our interactions with technology and informing decision-making processes across various sectors [17]. The rapid proliferation of AI applications has raised concerns about safety, privacy, fairness, and further ethical implications [3, 21]. In response to these challenges, AI governance has become an increasingly prominent focus for legislators and policymakers worldwide [37]. The European Union's

Authors' Contact Information: Gregor Autischer, Graz University of Technology, Graz, Austria, gregor.autischer@student.tugraz.at; Kerstin Waxnegger, Know Center Research GmbH, Graz, Austria, kwaxnegger@know-center.at; Dominik Kowald, Know Center Research GmbH & Graz University of Technology, Graz, Austria, dkowald@know-center.at.

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

AI Act, for instance, represents a landmark piece of legislation that aims to establish a comprehensive regulatory framework for AI systems [24]. Since its introduction, certification of AI systems to prove compliance is now more important than ever. AI certification must address a broad spectrum of concerns, including prediction accuracy, fairness, transparency, and should consider aspects related to ethics and reproducibility [34, 35, 43].

In our work, we aim to bridge the gap between theoretical certification frameworks and their practical application. We attempt to certify an existing open-source AI system (emotion recognition in an art installation [26, 27]) using current certification frameworks. We primarily use the Fraunhofer AI Assessment Catalogue [29] and we draw comparisons to other catalogues. With this approach, we aim to:

- Identify which parts of the catalogue are most useful and if simplifications or refinements could be beneficial.
- Provide a more practical understanding of the AI certification process, focusing on fairness and reliability.
- Discover the limitations where sample certifications encounter challenges.

2 Background

The growing deployment of AI in critical areas underscores the need for stricter regulation, as seen in approved medical systems that set a precedent for future frameworks [28, 31]. Worldwide, several organizations are advancing policy initiatives [25]. The most comprehensive and important for our work is the EU AI Act.

2.1 EU Artificial Intelligence Act

The European Union's AI Act is currently the most significant and far-reaching regulatory initiative in the field of AI, and it took effect on 1st of August 2024. The impact of the AI Act will extend far beyond the EU's boarders [11]. The AI Act has broad coverage for AI systems used within the European Union. The AI Act defines AI systems broadly and technology-neutrally, emphasizing functional characteristics rather than specific technologies (i.e., spanning from basic statistics-based to complex neural network-based predictions) [11]. This approach moves away from earlier, restrictive definitions, ensuring the framework remains applicable to current and future AI developments [12]. The AI Act categorizes AI systems into various risk levels and imposes corresponding requirements, with stricter regulations for higher-risk applications [11]. Key risk categories include: prohibited AI systems, high-risk AI systems, general-purpose AI systems, AI systems with special transparency obligations, and limited-risk AI systems.

2.2 Other Initiatives

Internationally, countries like the US have discussed a Blueprint for an AI Bill of Rights [42]. Japan has discussed developing AI guidelines emphasizing a multi-layered governance framework [9], and has recently even introduced a bill on AI risk management [36]. The recent AI Action Summit in Paris, also reflects the global acknowledgment of the need for AI regulation [38].

2.3 Certifying AI

Certification plays a critical role in proving compliance and fostering user trust [5], but AI certification differs significantly from traditional software certification due to AI's inherent opacity [43]. CEN, CENELEC, and ETSI bridge the gap between EU regulations and practical certification frameworks [15]. Furthermore, organizations like

ISO, IEC, and IEEE are also working on standards, they have yet to establish a comprehensive process connecting certification and applicable law [25]. The rapidly evolving nature of AI technology poses major challenges for creating and maintaining effective standards [8]. This underscores the urgent need for a broad framework that addresses AI's adaptability and complexity [4]. Addressing these dynamic aspects is essential to develop robust certification practices that are compliant with the EU AI Act.

2.4 AI Certification Catalogues

Several organizations have created AI certification guidelines, including the "Guideline for Designing Trustworthy Artificial Intelligence" by Fraunhofer [29], the "Trusted Artificial Intelligence" initiative by TÜV [43], and the "Auditing Machine Learning Algorithms" proposal by supreme audit institutions in the EU [32]. In this work, we focus on applying the Fraunhofer AI Assessment Catalogue, which is fully published and provides a comprehensive, questionnaire-style process centered on documentation and evidence artifacts, unlike the high-level, partially unpublished TÜV catalogue or the also fully published EU auditing approach, which uses documentation and a Excel-based tool covering data understanding, model development, performance, and ethical considerations.

3 Methodology for Certification

We applied the Fraunhofer AI Certification Catalogue to an existing AI system to explore and evaluate the certification process. Our first important step was to find and define an AI system. We chose the facial emotion recognition component of the RIOT art installation [39], which uses the EmoPy framework [1] for facial emotion recognition. It uses this technology to interactively adapt the plot of a film [16]. We selected this system because it is open-source, appears well-documented, and is integrated into a larger application context. Additionally, emotion recognition is a high risk AI application according to the EU AI Act. First, we compiled a list of all available documentation, like the RIOT GitHub repository [40] or articles describing the ML component [26, 27], and filled in some information gaps beforehand to have a complete set of information for our certification attempt. We used the Fraunhofer Catalogue as the primary certification framework due to its comprehensive nature and full public availability. In practice, our certification began by defining the AI system and its boundaries, then outlining its life cycle. We specifically focused on the two risk dimensions we assessed to be of medium risk - fairness and reliability - which we identified as the most important in the protection requirement analysis. A summary of the Protection Requirement Analysis can be seen in Table 1, with the full table available in the Arxiv version of this article [2]. The Fraunhofer Catalogue functions like a structured questionnaire, guiding us through each certification step, leading to our certification decision [29]. After completing this certification process, we analysed and addressed its key aspects and drew conclusions about the challenges we encountered. We also examined how the two alternative catalogues we mentioned differ from, and potentially complement, the Fraunhofer approach. Finally, we discussed the limitations and applicability of these findings, acknowledging both the constraints of our chosen AI application, namely emotion recognition in an art installation, and the broader challenges of certifying AI systems.

This methodology can be replicated by other researchers by selecting different AI systems relevant to their context. Researchers could similarly compile detailed system documentation. While the core steps of the Fraunhofer Catalogue remain fixed, the specific system choice, documentation approach, and risk dimension prioritization can be adapted to accommodate various use-cases.

Table 1. Summary for Protection Requirements Analysis for Fairness and Reliability.

Dimension	Risk Category	Reasoning
Fairness	Medium	Processes personal data (facial images), output linked to personal behaviour but has limited impact on rights.
Reliability	Medium	Misclassification can degrade user experience but does not cause major harm.

4 Certification Results

We performed a sample certification of a facial emotion recognition system using the Fraunhofer Catalogue. By using the Fraunhofer questionnaire as structured guidance, we examined training data representativeness, AI component modelling choices, feature selection processes, label quality, and other relevant factors. This catalogue-driven review uncovered significant gaps in documented fairness definitions and target group specifications, among other shortcomings. Due to these gaps and the resulting lack of clarity around key fairness criteria, we did not select, apply, or evaluate any quantitative fairness metrics, e.g., based on accuracy differences or model calibration [19, 20, 33]. A summary of our results for the two key assessed dimensions - fairness and reliability - can be found in Table 2. A more detailed account of our findings is available in the accompanying paper [2].

Table 2. Certification Summary for Evaluated Dimensions.

Dimension	Summary of Risk Analysis
Reliability	The system performs well enough within its defined scope, but lacks complete documentation and robustness testing. It is certifiable with improvements in testing and documentation.
Fairness	Insufficient analysis of potential biases and discriminatory behaviour. No clear metrics or target groups defined for fairness. Not certifiable without significant further improvements.

5 Key Findings and Limitations

In the following, we discuss key aspects we found while performing our sample AI certification:

Which parts of the catalogue are most useful, and could there be simplifications or refinements? The AI lifecycle overview proved especially helpful for gaining a comprehensive understanding of the system's functionality, while the detailed risk assessment ensured thorough coverage of potential issues. However, the high level of detail occasionally led to nearly redundant questions, suggesting that a different structuring could streamline the process. Additionally, the catalogue's exclusive focus on documentation, rather than direct examination of code, can be a strength when dealing with proprietary systems, but it may benefit from supplementary technical checkpoints, such as those in the TÜV [43] or the "Auditing Machine Learning Algorithms" catalogues [32].

Providing a more practical understanding of the AI certification process. Selecting a system with solid documentation and a clear real-world context was crucial for a meaningful certification attempt. The Fraunhofer Catalogue's step-by-step, questionnaire-like structure was generally effective in guiding us from initial system definition through risk analysis to a cross-dimensional assessment, although the process was time-intensive due to its thoroughness. To focus, our study we did not fully assess all risk dimensions. Even among the ones we did examine (fairness and reliability), we encountered shortcomings. Most notably, the system's fairness dimension failed to meet the certification criteria due to significant gaps in identifying potential biases and discrimination risks, coupled with the absence of clearly defined metrics (see also the full Arxiv version of this article [2]).

Discovering the limitations where sample certifications encounter challenges. Key limitations were that the system was no longer actively developed and that it was never initially built for certification. Therefore, the system had insufficient documentation in places, notably in the fairness dimension. In our case, these information gaps were not resolved by developers, as the system was no longer in active development. In a typical real-world scenario, certification findings would lead to refinements, underscoring the importance of ongoing development support. The lack of an iterative feedback loop meant that shortcomings could not be resolved within this project, highlighting how real-world AI certifications must allow for continuous improvements and collaboration.

Summed up, the Fraunhofer Catalogue, with its strong emphasis on documentation, effectively pinpoints critical risks, but requires considerable time, comprehensive system documentation, and an active feedback process with developers. Our findings illustrate the practical challenges and underscore the need for refinements and ongoing collaboration to ensure robust, real-world AI certification.

6 Conclusion and Future Research Directions

We highlight the complexity and shortcomings found in a sample AI certification, focusing on fairness and reliability. The implementation of the Fraunhofer AI Assessment Catalogue demonstrated its effectiveness as a comprehensive certification tool, particularly in its systematic approach to evaluate AI systems. We also found that the approach is, at times, bulky and time-consuming. The process showed that lacking documentation or developer support undermines feasibility. Our findings meet the research objectives by showing strengths and weaknesses of the approach. In future works, other certification catalogues and methodologies should be considered or integrated to potentially streamline the process. Another research path is to continue the certification attempt by incorporating feedback directly into the AI system design, ultimately resulting in a fully certifiable AI application and allowing for more in-depth analysis of a complete certification cycle. Further research could also focus on developing more flexible certification methods that adapt to various system states and development scenarios.

Acknowledgments

Know Center is a COMET competence center that is financed by the Austrian Federal Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK), the Austrian Federal Ministry of Labour and Economy (BMAW), the State of Styria, the Steirische Wirtschaftsförderungsgesellschaft m.b.H. (SFG), the vienna business agency and the Standortagentur Tirol. The COMET programme is managed by the Austrian Research Promotion Agency FFG.. A full version of this article is available via Arxiv [2].

References

[1] Angelica Perez, Julien Deswaef, and Puneetha Pai. 2021. thoughtworksarts/EmoPy. https://github.com/thoughtworksarts/EmoPy original-date: 2017-12-20T02:19:22Z.

[2] Gregor Autischer, Kerstin Waxnegger, and Dominik Kowald. 2025. Practical Application and Limitations of AI Certification Catalogues in the Light of the AI Act. *arXiv preprint arXiv:2502.10398* (2025).

[3] Kevin Baum, Joanna Bryson, Frank Dignum, Virginia Dignum, Marko Grobelnik, Holger Hoos, Morten Irgens, Paul Lukowicz, Catelijne Muller, Francesca Rossi, John Shawe-Taylor, Andreas Theodorou, and Ricardo Vinuesa. 2023. From fear to action: AI governance and opportunities for all. *Frontiers in Computer Science* 5 (2023).

[4] Benjamin Fresz, Vincent Philipp Göbels, Safa Omri, and Danilo Brajovic. 2024. The Contribution of XAI for the Safe Development and Certification of AI: An Expert-Based Analysis. https://arxiv.org/html/2408.02379v1

[5] Myrthe Blösser and Andrea Weihrauch. 2023. A consumer perspective of AI certification – the current certification landscape, consumer approval and directions for future research. *European Journal of Marketing* 58, 2 (2023), 441–470.

[6] Longbing Cao. 2022. Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)* 55, 3 (2022), 1–38.

[7] Carlos J. Costa and Manuela Aparicio. 2023. Applications of Data Science and Artificial Intelligence. *Applied Sciences* 13, 15 (2023), 9015. https://doi.org/10.3390/app13159015

[8] Raquel Delgado-Aguilera Jurado, Xiaojie Ye, Vicent Ortolá Plaza, María Zamarreño Suárez, Francisco Pérez Moreno, and Rosa María Arnaldo Valdés. 2024. An introduction to the current state of standardization and certification on military AI applications. *Journal of Air Transport Management* 121 (2024), 102685.

[9] Media & Sport Department for Digital, Culture. 2022. Establishing a pro-innovation approach to regulating AI. https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement

[10] Tomislav Duricic, Peter Müllner, Nicole Weidinger, Neven ElSayed, Dominik Kowald, and Eduardo Veas. 2024. AI-Powered Immersive Assistance for Interactive Task Execution in Industrial Environments. In *ECAI 2024*. IOS Press, 4491–4494.

[11] European-Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)Text with EEA relevance. http://data.europa.eu/eli/reg/2024/1689/oj/eng Legislative Body: CONSIL, EP.

[12] Frederiek Fernhout and Thibau Duquin. 2024. The EU Artificial Intelligence Act: our 16 key takeaways | Stibbe. https://www.stibbe.com/publications-and-insights/the-eu-artificial-intelligence-act-our-16-key-takeaways

[13] Valentin Förster, Jürgen Fleiß, Dominik Kowald, and Viktoria HSE Robertson. 2025. Detecting Resale Price Maintenance with Unsupervised Machine Learning. *Available at SSRN 5130745* (2025).

[14] Armin Haberl, Jürgen Fleiß, Dominik Kowald, and Stefan Thalmann. 2024. Take the aTrain. Introducing an interface for the Accessible Transcription of Interviews. *Journal of Behavioral and Experimental Finance* 41 (2024), 100891.

[15] Hadrien Pouget. 2024. Standardsetzung | EU-Gesetz zur künstlichen Intelligenz. https://artificialintelligenceact.eu/de/standardeinstellung/

[16] Karen Palmer. 2017. RIOT Video. https://www.youtube.com/watch?v=-BCny9SuI3A

[17] Maria Kasinidou, Styliani Kleanthous, Matteo Busso, Marcelo Rodas, Jahna Otterbacher, and Fausto Giunchiglia. 2024. Artificial Intelligence in Everyday Life 2.0: Educating University Students from Different Majors. In *Proceedings of the 2024 on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024)*. New York, NY, USA, 24–30.

[18] Dominik Kowald, Sebastian Dennerlein, Dieter Theiler, Simon Walk, and Christoph Trattner. 2013. The social semantic server: A framework to provide services on social semantic network data. In *9th International Conference on Semantic Systems, I-SEMANTICS 2013*. CEUR, 50–54.

[19] Dominik Kowald, Gregor Mayr, Markus Schedl, and Elisabeth Lex. 2023. A study on accuracy, miscalibration, and popularity bias in recommendations. In *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer, 1–16.

[20] Dominik Kowald, Markus Schedl, and Elisabeth Lex. 2020. The unfairness of popularity bias in music recommendation: A reproducibility study. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42*. Springer, 35–42.

[21] Dominik Kowald, Sebastian Scher, Viktoria Pammer-Schindler, Peter Müllner, Kerstin Waxnegger, Lea Demelius, Angela Fessl, Maximilian Toller, Inti Gabriel Mendoza Estrada, Ilija Šimić, et al. 2024. Establishing and evaluating trustworthy AI: overview and research challenges. *Frontiers in Big Data* 7 (2024), 1467222.

[22] Dominik Kowald, Paul Seitlinger, Simone Kopeinik, Tobias Ley, and Christoph Trattner. 2013. Forgetting the words but remembering the meaning: Modeling forgetting in a verbal and semantic tag recommender. In *International Workshop on Mining Ubiquitous and Social Environments*. Springer, 75–95.

[23] Emanuel Lacic, Dominik Kowald, Paul Christian Seitlinger, Christoph Trattner, and Denis Parra. 2014. Recommending Items in Social Tagging Systems Using Tag and Time Information. In *In Proceedings of the 1st Social Personalization Workshop co-located with Hypertext'14*. ACM, 4–9.

[24] Markus Mueck, Scott Cadzow, Cadzow Communications, and Suno Wood. 2022. *ETSI Activities in the field of Artificial Intelligence Preparing the implementation of the European AI Act - 1st Edition – December -2022*. Technical Report. ETSI.

[25] Tomislav Nad, Sebastian Scher, and Florian Königstorfer. 2023. *Trustworthiness of AI*. Technical Report NIST AI 100-1. SGS.

[26] Angelica Perez. 2018. EmoPy: a machine learning toolkit for emotional expression. https://www.thoughtworks.com/insights/blog/emopy-machine-learning-toolkit-emotional-expression

[27] Angelica Perez. 2018. Recognizing human facial expressions with machine learning. https://www.thoughtworks.com/insights/articles/recognizing-human-facial-expressions-machine-learning

[28] Brandon Pimentel. 2024. Why AI still needs regulation despite impact. https://legal.thomsonreuters.com/blog/why-ai-still-needs-regulation-despite-impact/

[29] Dr Maximilian Poretschkin, Anna Schmitz, Dr Maram Akila, Linara Adilova, Dr Daniel Becker, Dr Armin B Cremers, Dr Dirk Hecker, Dr Sebastian Houben, Julia Rosenzweig, Joachim Sicking, Elena Schulz, Dr Angelika Voss, and Dr Stefan Wrobel. 2023. *AI Assessment Catalog*. Technical Report. Fraunhofer IAIS.

[30] Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J Topol. 2022. AI in health and medicine. *Nature medicine* 28, 1 (2022), 31–38.

[31] Agustina D. Saenz, Zach Harned, Oishi Banerjee, Michael D. Abràmoff, and Pranav Rajpurkar. 2023. Autonomous AI systems in the face of liability, regulations and costs. *npj Digital Medicine* 6, 1 (Oct. 2023), 1–3. https://doi.org/10.1038/s41746-023-00929-1 Publisher: Nature Publishing Group.

[32] SAI-FI-DE-NL-NO-UK. 2023. *Auditing machine learning algorithms*. Technical Report. Supreme Audit Institutions FI, DE, NL, NO, UK.

[33] Sebastian Scher, Simone Kopeinik, Andreas Trügler, and Dominik Kowald. 2023. Modelling the long-term fairness dynamics of data-driven targeted help on job seekers. *Scientific Reports* 13, 1 (2023), 1727.

[34] Harald Semmelrock, Simone Kopeinik, Dieter Theiler, Tony Ross-Hellauer, and Dominik Kowald. 2023. Reproducibility in machine learning-driven research. *arXiv preprint arXiv:2307.10320* (2023).

[35] Harald Semmelrock, Tony Ross-Hellauer, Simone Kopeinik, Dieter Theiler, Armin Haberl, Stefan Thalmann, and Dominik Kowald. 2025. Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine* 46, 2 (2025), e70002.

[36] The Yomiuri Shimbun. 2025. Japan Cabinet OK's Bill on AI Risk Management, Innovation; Govt, Ruling Parties Aim to Pass Bill in Current Diet Session. *The Japan News (Online)* (2025). https://japannews.yomiuri.co.jp/politics/politics-government/20250228-241005/

[37] Nathalie A. Smuha. 2021. From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. *Law, Innovation and Technology* 13, 1 (March 2021), 57–84. https://doi.org/10.1080/17579961.2021.1898300

[38] techUK. 2025. What were the outcomes of the Paris AI Action Summit? *techUK (Online)* (2025). https://www.techuk.org/resource/what-were-the-outcomes-of-the-paris-ai-action-summit.html

[39] Thoughtworks. 2018. RIOT | Thoughtworks Arts. https://thoughtworksarts.io/projects/riot/

[40] Thoughtworks. 2019. thoughtworksarts/riot. https://github.com/thoughtworksarts/riot original-date: 2017-11-22T15:59:17Z.

[41] Elmira Van den Broek, Anastasia Sergeeva, and Marleen Huysman. 2021. When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS quarterly* 45, 3 (2021).

[42] White-House. 2022. Blueprint for an AI Bill of Rights. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

[43] Philip Matthias Winter, Sebastian Eder, Johannes Weissenböck, Christoph Schwald, Thomas Doms, Tom Vogt, Sepp Hochreiter, and Bernhard Nessler. 2021. Trusted Artificial Intelligence: Towards Certification of Machine Learning Applications. arXiv:2103.16910.