Practical Application and Limitations of AI Certification Catalogues in the Light of the AI Act

Gregor Autischer[1*], Kerstin Waxnegger[2] and Dominik Kowald[1,2*]

[1]Graz University of Technology, Graz, Austria.
[2]Know Center Research GmbH, Graz, Austria.

*Corresponding author(s). E-mail(s):
gregor.autischer@student.tugraz.at; dkowald@know-center.at;
Contributing authors: kwaxnegger@know-center.at;

Abstract

In this work-in-progress, we investigate the certification of AI systems, focusing on the practical application and limitations of existing certification catalogues in the light of the AI Act 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 publicly accessible 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. 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 anymore and highlighted the importance of complete system documentation. Finally, we identify some limitations of the certification catalogues used and proposed ideas on how to streamline the certification process.

Keywords: Algorithmic Auditing, Artificial Intelligence, Certification Catalogues

1 Introduction

Artificial intelligence (AI) has evolved, over several decades, to complex machine learning (ML) algorithms and neural networks that are common-place today. However, in recent years, AI systems were increasingly integrated into our daily lives, moving

from specialized research labs to mainstream applications [1]. Today AI is utilized in many domains such as healthcare [2], human resources [3], finance [4] as well as in prominent technologies such as recommender systems for e.g., social networks [5, 6] or online marketplaces [7, 8]. AI now plays a significant role in shaping our interactions with technology and informing decision-making processes across various sectors [9]. The rapid proliferation of AI applications has raised concerns about safety, privacy, fairness, and further ethical implications [10, 11]. In response to these challenges, AI governance has become an increasingly prominent focus for legislators and policymakers worldwide [12]. The European Union's AI Act, for instance, represents a landmark piece of legislation that aims to establish a comprehensive regulatory framework for AI systems [13]. Similar initiatives are underway in other countries and regions. This reflects a growing global consensus on the need for AI governance [14]. Organizations and policymakers are increasingly emphasizing the need for AI system certification to achieve regulatory compliance and foster confidence in AI systems. Since the EU AI Act entered into force on August 1, 2024, this is now more important than ever. The certification of AI systems is a complex challenge, distinctly different from traditional software certification. While conventional software certification primarily focuses on functionality and security, AI certification must address a broader spectrum of concerns, including prediction accuracy, fairness, transparency, and other ethical considerations [15].

With this work-in-progress, we aim to bridge the gap between theoretical certification frameworks and their practical application. We attempt to certify an existing open-source AI system using current certification frameworks and document the challenges encountered along the way. We primarily use the Fraunhofer AI Assessment Catalogue published in Poretschkin et al. [16]. However, 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 potential AI certification process.
- Discover the limitations where sample certifications encounter challenges.

Therefore, we provide a comprehensive walk-through of an AI certification in practice. Our work starts with essential background information on selected AI regulations and existing certification catalogues. We then detail the specific AI system chosen for certification, and the core of this work documents our AI certification attempt. We offer an analysis of what aspects of certification were achievable and which proved problematic. We conclude this paper by describing these findings and by offering recommendations for future developments in AI certification schemes.

2 Current State of AI Regulation

Rapid advancement and widespread adoption of AI in various industries require the development of comprehensive regulatory frameworks, and increasing AI deployment in more critical areas appears to require stricter regulation [17]. Some bodies like the US Food and Drug Administration have already approved certain AI applications,

particularly in medicine [18]. In the European Union, some medical applications also got approval, one example is the ChestLink software, that automatically reports chest x-rays that are classified as normal. Systems such as this set an important precedent for other medical and potentially nonmedical systems in the future [19]. Recently, lawmakers have recognized the urgent need for comprehensive regulation. New regulatory frameworks intend to encompass not only individual industries, but also to create universal rules that ensure consistency and fairness. Worldwide, several organizations are advancing policy initiatives [14]. Now, we describe some key efforts in this area.

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, potentially setting global standards for AI management. The AI Act uses a risk-based approach to achieve the following goals [20]:

- Ensure AI safety and compliance: guarantee that AI systems in the EU are safe and adhere to laws protecting fundamental rights and EU values, safeguarding users' privacy and preventing discrimination.
- Certainty for AI investment: establish a clear legal framework to foster innovation and investment in AI, providing businesses and investors with regulatory clarity.
- Enhance governance: strengthen enforcement mechanisms to effectively apply existing laws on fundamental rights and AI safety requirements.
- Unify AI market: foster a single, cohesive market for trustworthy AI applications across the EU, preventing fragmentation through harmonized regulations.

2.1.1 Scope

The AI Act covers various actors and scenarios within the AI ecosystem [20]. The regulation applies to providers, deployers, importers, and distributors of AI systems, as well as product manufacturers incorporating AI systems into their products, regardless of their location, if the AI system or its output is used within the European Union. The AI Act explicitly excludes, among others, AI systems developed purely for scientific research and development. It also excludes AI systems published under open-source licences. However, the regulation applies if an organisation brings an open-source system to market or uses it as a prohibited AI system, a high-risk AI system or an AI system with special transparency obligations. The AI Act imposes obligations across the entire AI value chain, ensuring a comprehensive approach to AI governance and safety within the European market.

2.1.2 Definition of AI

The AI Act adopts a broad and technology-neutral definition of AI systems, focusing on their functional characteristics rather than specific technologies or methods. According to the AI Act, an AI system is defined as a machine-based system designed to operate with varying levels of autonomy, which potentially exhibits adaptiveness

after deployment, and that is capable of generating outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments based on input it receives [20]. This definition emphasizes two key elements: 'inference' and 'autonomy', which distinguish AI systems from traditional software with predetermined outputs. The AI Act takes a broad approach to stay relevant as technology rapidly advances. It encompasses both the core AI system and its surrounding code, recognizing the complexity of AI applications. This definition aligns with recent conceptualizations of AI and moves away from earlier, more restrictive definitions tied to specific technologies [21] (e.g., ML). The AI Act creates a framework that can accommodate current and future AI technologies, ensuring its long-term applicability in regulating the AI landscape.

2.1.3 AI Risk Categories

The AI Act categorizes AI systems into various risk levels and imposes corresponding requirements, with stricter regulations for higher-risk applications [20]. Key categories include:

1. Prohibited AI systems.
2. High-risk AI systems.
3. General-purpose AI systems.
4. AI systems with special transparency obligations.
5. Limited-risk AI Systems.

Prohibited systems include, but are not limited to, systems that manipulate behaviour, exploit vulnerabilities, or create facial recognition databases from untargeted scraping. The AI Act defines high-risk AI systems as those that pose significant risks to health, safety, or fundamental rights, and that are either used as products (or components of products) covered by specific EU legislation and require third party certification, or that are listed in Annex III of the AI Act [20]. Annex III includes several areas such as biometric identification, emotion recognition systems, management of critical infrastructure, education, employment, and law enforcement. General-purpose AI systems face some regulation, but it is less stringent than for high-risk systems. The AI Act also provides an exception for certain AI systems that, despite falling under Annex III categories, may not be considered high-risk. This is the case if they perform narrow procedural tasks or do not significantly influence human decision-making, provided they do not involve profiling of natural people. AI systems with special transparency obligations, according to Article 50, require clear disclosure when users interact with AI or encounter AI-generated content. Limited-risk AI systems are only subject to voluntary codes of conduct for ethical and responsible use, according to Article 95.

2.2 EU Artificial Intelligence Liability Directive

Complementing the AI Act, the European Commission proposed the AI Liability Directive in September 2022. This directive aims to modernize and enhance the EU's liability framework for AI systems [22], and also to ensure that individuals who suffer damages from AI systems receive equivalent protection to those harmed by other

forms of technology. By standardizing liability regulations across the EU, the directive aims to avoid legal discrepancies and guarantee uniform protection for those impacted by AI-related damages. The AI Act and other EU measures coordinate with the AI Liability Directive, which addresses only non-contractual liability claims, encompassing a wide array of potential AI-related harms [23]. This comprehensive strategy seeks to balance the protection of victims with the encouragement of AI innovation, reducing legal uncertainties and promoting the responsible advancement of AI technologies within the EU.

2.3 Other Global Initiatives

While the EU has developed one of the most comprehensive regulatory efforts, other countries and regions have also proposed AI regulations. The following examples represent some relevant international regulatory efforts. For instance, the US has discussed a Blueprint for an AI Bill of Rights, which outlines principles for the design and deployment of AI systems, although they currently only have a patchwork of state laws [24]. The UK has discussed a sector-led approach to AI regulation, and Japan has developed AI Guidelines emphasizing a multi-layered governance framework [25]. These initiatives do not have legal enforceability. They still highlight the global acknowledgment of the necessity for AI regulation and certification efforts. The current state of AI regulation evolves rapidly, with the EU taking a leading role by introducing the AI Act.

3 Certifying AI

Organizations use certification as an important and established tool to prove that technical systems meet certain standards or regulations. Successful certifications play a vital role in proving a system's compliance with applicable norms. It also plays a crucial role in establishing trust in a system among its users [26]. For traditional software projects, where every block of code can undergo review line by line, companies have long-established certification processes [15].

Organizations and regulatory bodies are still in the early stages of certifying AI applications. This is partly because policymakers and other actors have just recently begun developing comprehensive legal frameworks for AI. A prime example is the aforementioned EU AI Act. Despite ongoing efforts by international standardization organizations like ISO, IEC, and IEEE to create guidelines and standards, they have yet to fully establish a certification process [14]. These efforts aim to address the distinct challenges presented by AI technologies. However, the absence of comprehensive legal frameworks has hindered the development of robust certification processes for AI. Even with recent legislative developments, the rapidly evolving nature of AI technology continues to pose significant challenges for creating and maintaining effective certification standards [27]. This dynamic landscape requires certification processes that are both adaptable and rigorous, capable of evolving alongside the technology they aim to regulate.

Certifying AI systems also presents unique challenges due to their complex and often opaque nature. Often, humans do not directly program logical rules to model

decision-making processes. Instead, AI systems use various methods to analyse and interpret data, learning their own rules for decision-making processes. Humans do have multiple ways to influence and understand how the system makes decisions and how good the outcome is. However, the process lacks complete transparency. This complexity necessitates a different approach to certification, emphasizing the need to establish a comprehensive framework that accounts for the inherent opacity and adaptability of AI technologies [15]. Moreover, AI systems can evolve and change their behaviour over time by retraining with new data. This continuous learning process adds another layer of complexity to certification, as regulators and developers must continuously monitor and re-evaluate systems to ensure they comply with ethical standards and avoid biases. Addressing these dynamic aspects is essential to develop robust certification practices that can keep pace with the rapid evolution of AI [28].

3.1 AI Certification Catalogues

CEN, CENELEC and ETSI are leading European standardization bodies, they bridge the gap between EU regulations and practical certification frameworks designed to evaluate and certify AI systems. They integrate these guidelines with European legislative priorities, and ensure consistency across the European standardization landscape [29]. Several organizations have created catalogues and guidelines to evaluate, test, and certify AI systems. Prominent examples include the Fraunhofer AI Assessment Catalogue by Poretschkin et al. [16], the white paper 'Trusted Artificial Intelligence' by Winter et al. [15], and the white paper 'Auditing Machine Learning Algorithms' by the supreme audit institutions of various countries [30]. These frameworks provide distinct methodologies and list criteria to certify AI applications, addressing aspects such as fairness, autonomy and control, transparency, reliability, safety and security, and data protection. With this work, we focus primarily on the Fraunhofer Certification Catalogue and how it applies to a concrete AI application.

3.1.1 Fraunhofer AI Assessment Catalogue

The Fraunhofer AI Assessment Catalogue emphasizes the necessity of implementing stringent quality standards to ensure AI systems are reliable, safe, aligned with societal values and compliant with the law, particularly in sensitive application contexts [16]. The catalogue identifies several key challenges in assessing and ensuring AI quality. These include the complex value chain involved in AI development and the difficulty in explaining the inner workings of AI models. The authors of the catalogue argue that these challenges necessitate a systematic approach to quality implementation in AI development and highlight the importance of unbiased expert assessment in establishing trust in AI applications. A significant focus of the catalogue lies in operationalizing quality requirements for AI. While there are established guidelines for reliable AI, the catalogue points out that the specifics of their practical application remain largely unclear. The paper proposes a risk-based AI assessment approach and introduces an AI assessment catalogue. This catalogue provides a structured approach for certifying AI applications across different dimensions of trustworthy AI: fairness, autonomy and control, transparency, reliability, safety and security, and data protection. The

proposed framework offers a procedure for developing safeguarding arguments for AI applications. It tries to support developers and operators of AI systems in meeting regulatory requirements. It introduces a step-by-step process that this work uses when trying to certify an AI System.

3.1.2 Trusted Artificial Intelligence

This white paper published by TÜV Austria and Johannes Kepler University Linz wants to outline a structured approach to certifying ML applications [15]. It describes key ML principles and discusses relevant aspects and challenges in the context of certification. It emphasizes that while ML systems are complex, they are not black boxes, but rather white boxes whose operations can be analysed in detail. The paper introduces a certification approach for ML applications, focusing initially on supervised learning tasks with low-risk potential. It focuses mainly on technical aspects and cybersecurity measures. The paper presents a summary of the certification framework. However, since the actual catalogue remains inaccessible to the public, auditors cannot use it directly to certify an AI model. Instead, it serves as a reference point, highlighting key areas to consider during the certification process.

3.1.3 Auditing Machine Learning Algorithms

A collaboration of European public auditing institutions released this white paper. It highlights the increasing use of AI and ML in public services, emphasizing the need for new certification methodologies [30]. It identifies several risks, including an over-focus on numerical metrics at the expense of compliance and fairness, miscommunication between product owners and developers, over-reliance on external expertise, and uncertainty regarding personal data use. To address these challenges, the paper proposes a certification framework covering the entire AI application lifecycle. The audit areas focus on data understanding, model development, performance, and ethical considerations such as explainability and fairness. To aid in this process, the paper introduces a helper tool, in the form of a spreadsheet. Auditors can use it to prepare and conduct AI audits efficiently. The authors stress that specialized knowledge and skills are required for ML certifications. They emphasize that the proposed audit catalogue and helper tool should be continuously refined and updated. Ultimately, this paper aims to provide guidance and good practices to enable auditors to navigate different parts of the certification process.

4 Our AI Application: Facial Emotion Recognition

To undertake a certification process, the first requirement is a system that either requires certification or is eligible for it. The system should incorporate an AI component, ideally leveraging ML techniques. Furthermore, the AI component must be integrated into a larger, comprehensive system, as certification typically applies to entire systems rather than isolated components. Specifically, the Fraunhofer Certification Catalogue mandates a well-defined assessment object, which requires the AI component to be part of a larger, integrated system [16].

Addressing these challenges requires the auditor to identify and select a suitable system that meets these criteria. The system should include a ML-based AI component, and should demonstrate sufficient complexity and integration to justify certification. The objective is to find a system where the AI component plays a critical role in the overall functionality, thereby making the certification process relevant and meaningful. In essence, our aim is to identify a system that incorporates a ML-based AI module within a broader architecture.

4.1 Our Decision to Use the EmoPy Framework and RIOT Project for AI Certification

After careful consideration, we decided to use the EmoPy Framework [31] and its implementation within the RIOT Project [32] for this sample certification. Several key factors influenced this choice, to ensure that both the framework and the project are suitable for the certification process. Firstly, the EU AI Act's relevance significantly influenced our decision. Emotion recognition, the primary focus of the EmoPy Framework, aligns with the potential coverage of the EU AI Act. While open-source models like EmoPy are generally exempt, this system may fall under the AI Act's scope if it was brought to market as a high-risk AI system, a prohibited AI system or an AI system with special transparency obligations.

Another critical factor was the open-source nature and transparency of EmoPy. EmoPy is fully open-source, enabling an in-depth look into the technical details and making the system fully transparent, which should make the certification process possible. The codebase of EmoPy is relatively small and manageable, making it easier to understand and verify. Yet, it is sufficiently large to make a certification worthwhile. Additionally, the framework includes thorough documentation, with multiple articles and resources that describe the model selection process. This level of documentation is critical for reproducibility attempts of the AI models [33, 34], and with this also for the certification process, as it provides clear insights into the design and functionality of the model, facilitating a thorough evaluation. While solid documentation is essential for any certification, it is especially critical for our sample certification, since no company or active development team is managing the project anymore. Although no active development team provides ongoing support, contacting authors and lead developers of the EmoPy framework articles and code, proved beneficial. They kindly addressed questions about the framework and the RIOT setup. Beyond this input, we needed to extract all necessary information from the provided documentation. This reliance on static resources poses a limitation compared to standard certification processes, where ongoing interactions with active developers should be possible.

From a technical perspective, we considered the suitability of the EmoPy framework for sample certification a key factor. Its technical characteristics and comprehensive documentation make it well-suited for the certification process. Additionally, the integration of EmoPy within the RIOT project provides a complete system context, which is essential for certification. Traditional certification catalogues often struggle to validate standalone ML models. However, the RIOT project offers a comprehensive framework where the AI component is embedded within a broader system. This integration is critical, as certification typically applies to an entire system rather than

Fig. 1 RIOT installation. This image shows the RIOT Art installation in New York City in 2018. One participant is standing in front of the screen taking part in the experience [35].

individual components. All these factors influenced our choice to use the EmoPy framework and the RIOT project for our sample certification. Key considerations included its potential relevance to the EU AI Act, its open-source nature ensuring transparency, and its suitability for a sample certification. Additionally, the comprehensive system context that the RIOT project provides reinforced this decision, enabling an effective and thorough certification process. It is also important to note that both EmoPy and RIOT were not originally intended for certification by their creators. However, they serve well for this academic exercise, as the focus is on the certification procedure rather than the system's quality or real-world applicability.

4.2 Brief Overview of the EmoPy Framework

The EmoPy framework provides multiple neural network architectures for facial expression recognition, including ConvolutionalNN, TransferLearningNN, and ConvolutionalLstmNN. These architectures vary in complexity, with the ConvolutionalNN beingthe simpleste, and TransferLearningNN (using Google's Inception-v3) being the most complex. The authors experimented with different architectures and found that the ConvolutionalNN provides the best overall performance [35]. The EmoPy documentation suggests using two publicly available datasets for training and evaluation: the Microsoft FER2013 dataset and the Extended Cohn-Kanade dataset. The key goals of the EmoPy project are to provide free, open-source, and easy-to-use facial expression recognition capabilities, and to advance research in this field by making the models and datasets publicly available. The Framework was also used for the emotion recognition in the RIOT Project.

4.3 Brief Overview of the RIOT Project

In a nutshell, RIOT is a live-action film that dynamically responds to emotions, utilizing facial emotion recognition technology to guide viewers through an ongoing

dangerous riot. The experience allows the audience's emotions to drive the narrative of the film in real-time [32]. The project began during artist Karen Palmer's 2017-2018 residency at Thoughtworks, where she developed a new iteration of the emotion analysis engine and the RIOT user experience. The RIOT installation has since been showcased at various events and festivals.

The RIOT experience integrates the EmoPy framework and its pretrained emotion recognition model into its system to respond to participants' emotional states during the live-action film sequence. The characters and narrative adapt to the viewer's detected emotions, creating an immersive, multisensory experience that, according to its creators, enhances cognitive skills and self-awareness [36]. The system setup is shown in Figure 1. A participant stands in front of the screen to watch the experience. At different intervals, the mounted webcam captures the person's face and predicts their emotion. Based on the detected emotion, the film progresses differently, creating an interactive experience [35].

4.4 The AI Application to be Certified

For this sample certification, the RIOT installation provides the complete system context for the certification process. The EmoPy framework, which provides ML-Models to detect emotions, serves as the core framework within this installation. Our certification process focuses on validating the AI system within the RIOT context. The surrounding code, setup, and information that comprise the entire art installation are relevant, as we cannot certify the AI system independently of these components. However, this approach has shortcomings, as not all the required information is available. Our approach allows for an exploration of the certification process while acknowledging its limitations and academic nature. To facilitate the certification process, we prepared a complete overview of all available information about the AI application, which is summarized below. We also compiled a complete technical overview of the system, that we certified.

4.4.1 Compilation of System Resources

RIOT Installation

- GitHub repository of the RIOT art installation [37]
- Article on the RIOT art installation [32]
- TED Talk on the RIOT art installation [38]
- Article that describes different art installations (one of them is the RIOT art installation) [39]
- Article on Karan Palmer (the artist behind the RIOT Art Installation) [40]
- Short description of the RIOT art installation by Karan Palmer [36]
- Video that showcases and describes the RIOT art installation [41]

EmoPy Framework

- GitHub repository of the EmoPy Framework [31]
- Article describing the EmoPy Framework and technical decisions that were made in more detail [35]

- Article describing in more detail decisions that were made regarding the architecture of the Emotion recognition model [42]
- Python documentation of the EmoPy Framework [43]
- GitHub repository of the FER+ dataset [44]
- Extended Cohn-Kanade dataset [45]

The AI Application that is Certified

- The AI Application is not the entire RIOT art installation, but only the system subset that receives centred images from the webcam and returns the emotion predictions.
- Since the pretrained model and the target emotions used during training are not disclosed, we assume that the EmoPy ConvolutionalNN model is used without any alterations. Furthermore, we assume that the target emotions are anger, fear, calm, and surprise. We presume that the model with these features is the model used in the installation.
- We treat all items listed above as if they are part of a proper documentation of the system. We also treat items that are articles or videos as part of the documentation resources that we can use for the AI certification process.

4.4.2 Technical Overview

The AI Application that we attempted to certify has the following specifications: The developers trained the ConvolutionalNN from the EmoPy Framework, as this architecture had the best performance during early testing. The layers and parameters of the ConvolutionalNN can bee seen in Table 1.

Layer Name	Type	Output Shape	Parameters
conv2d_1	Conv2D	(None, 61, 61, 10)	170
conv2d_2	Conv2D	(None, 58, 58, 10)	1610
max_pooling2d_1	MaxPooling2D	(None, 58, 29, 5)	0
conv2d_3	Conv2D	(None, 55, 26, 10)	810
conv2d_4	Conv2D	(None, 52, 23, 10)	1610
max_pooling2d_2	MaxPooling2D	(None, 52, 11, 5)	0
flatten_1	Flatten	(None, 2860)	0
dense_1	Dense	(None, 4)	11444

Table 1 Summary of model layers and parameters. [35]

In total the model has 15640 trainable parameters and has 4 outputs to be able to classify 4 emotions. The model architecture consists of multiple convolutional and max-pooling layers, which gradually reduce the spatial dimensions while extracting hierarchical features. A final dense layer connects the extracted features to the output layer for classification of the emotions: anger, fear, calm or surprise.

To train the model, two datasets were used. The FER2013 dataset contains over 35,000 facial expression images across 7 emotion classes, while the Cohn-Kanade dataset includes 327 facial expression sequences. However, for this application only the subset of four emotions (anger, fear, calm and surprise) were used. The number

of images for each of these emotions can be seen in Table 2. To increase the size and suitability of the training and validation datasets, the authors applied data augmentation techniques. The training process begins by splitting the dataset into training and validation subsets. During training, the network weights are iteratively adjusted to minimize the loss between predicted and labelled emotions. To mitigate overfitting, the authors monitored the gap between training and validation accuracy. This approach prevents overfitting and ensures the model generalizes well by using unseen validation data during training. They measured performance using training and validation accuracy, to help refine the models [31].

Emotion	FER2013 Quantity	Cohn-Kanade Quantity
Anger	4953	45
Fear	5121	25
Calm	6198	0
Surprise	4002	83

Table 2 Emotion distribution comparison between FER2013 and Cohn-Kanade datasets for used emotion subset [35].

The developers tested the model and also analysed the confusion matrices to gain insights into the model's performance [35]. Some confusion matrices are also published [31]. The ones relevant to this application with the used emotion subset for both the FER2013 and the Cohn-Kanade Dataset are also published, and for brevity we combined them into one confusion matrix that represents the AI-systems performance for our certification attempt. We considered that the Cohn-Kanade dataset is approximately 100 times smaller than the FER2013 dataset. The resulting matrix can be seen in Figure 2. The data used during training to create this matrix consists only of the discussed datasets (FER2013 and Cohn-Kanade). However, the final model was trained on these datasets with additional data generated through data augmentation.
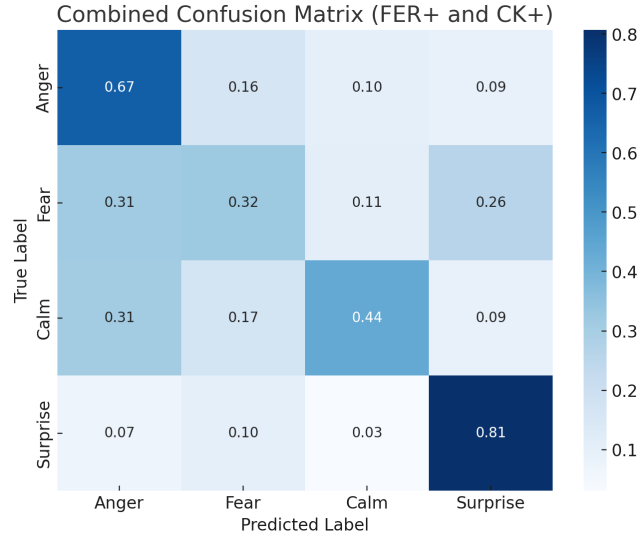

Fig. 2 Confusion matrix of the trained ConvolutionalNN [35].

5 AI Certification Approach

We apply the Fraunhofer AI Certification Catalogue to an existing AI system to explore and evaluate the certification process. We chose the facial emotion recognition component of the RIOT art installation, which uses the EmoPy framework. We selected this system because it is open-source, appears well-documented, and is integrated into a larger application context. Additionally, as we have discussed previously, it is potentially covered by the EU AI Act. In this work-in-progress, we use the Fraunhofer Catalogue as the primary certification framework due to its comprehensive nature and full public availability. The catalogue provides a structured approach to AI certification, addressing multiple dimensions of risk. The Fraunhofer Catalogue outlines the certification process, which involves these key steps [16]:

1. **First Step:** get an overview of the System (AI Profile (PF)) and define the AI-System and the boundaries to the surrounding system.
2. **Second Step:** define the life cycle of the AI application.
3. **Main Step:** get an overview over all the risk dimensions.
   (a) **Protection requirements analysis:** determine which risk dimensions apply.
   (b) **Risk Analysis:** for each applicable dimension:
      (i) Risk analysis and objectives.
      (ii) Criteria for achieving objectives.
      (iii) Measures.
      (iv) Overall assessment of a risk area.
      (v) Summary of each dimension.
      (vi) Cross-dimension assessment.
4. Drawing conclusions and making a certification decision based on the success of the cross-dimensional assessment.

After completing the certification process with the Fraunhofer Catalogue, we analyse and address several key aspects of the process, and we draw conclusions about the challenges we encounter. In this paper, we explore the challenges of selecting an appropriate AI application for certification. We also evaluate how effectively the Fraunhofer certification process worked for this specific case, and highlight its strengths and areas for improvement. Furthermore, we present a comparative analysis, examining how the other two introduced catalogues differ from the Fraunhofer approach and how they could potentially enhance or complement the certification process. Lastly, we evaluate the limitations of this approach. Our analysis addresses the constraints of the chosen AI system and certification catalogue, as well as the applicability of the findings to other AI applications and certification scenarios. With this comprehensive evaluation, we aim to provide valuable insights into the practical implementation of AI certification processes and contribute to the ongoing discourse on AI certification.

5.1 Before the Certification Process

Before starting the certification process, we completed several preparatory steps. First, we selected the AI application. We forked the GitHub repository of the EmoPy project and identified the correct dependencies to enable a detailed examination of

the codebase. We conducted an extensive research phase, focusing on both the EmoPy framework and the RIOT installation. This research resulted in the creation of the system summary, as previously presented. We created this basis, which combines elements of both the EmoPy framework and the RIOT installation, to facilitate the certification process. This step was necessary to provide a complete system context for certification, as certifying standalone ML models can be challenging with the Fraunhofer certification catalogue and would differ significantly from a real-world certification. During the preparation phase, we identified certain information gaps in the original documentation. To make the certification process feasible, we closed these gaps with some assumptions and additional details. We made these additions, based on reasonable interpretations of the available information and common practices in AI development, and kept them to a minimum to enable the certification process.

5.2 AI Profile

The Fraunhofer Catalogue outlines the first formal step in the certification process: completing the AI Profile. This step was straightforward due to the thorough research we conducted in the preparation phase. The AI Profile offered a structured overview of the system's functionality, intended application context, and key characteristics.

5.3 Life Cycle of the AI Application

Following the AI Profile, we conducted the life cycle overview. Although not explicitly stated as a distinct step in the Fraunhofer Catalogue, we found it beneficial to gain a thorough understanding of the AI system's development and operation stages. The AI life cycle encompasses all the stages an AI system undergoes, from planning and development to deployment, operation, ongoing maintenance, and potentially continued model training, ensuring trustworthiness and compliance throughout its use. We adapted the questions for this life cycle overview from a table in the Fraunhofer Catalogue, covering aspects such as data acquisition, model development, and operational considerations.

5.4 Protection Requirement Analysis

The protection requirement analysis serves as an important first step in the certification process, identifying the risk dimensions that require more in-depth analysis. This analysis involves evaluating the potential impact of the AI system across various dimensions such as fairness, reliability, and data protection. We examined all dimensions and identified several with medium risk. A summary of the Protection Requirement Analysis can be seen in Table 3. For the purposes of this work, we selected two of the required risk dimensions, namely reliability and fairness, for detailed analysis. This selection helps us to focus the sample certification and manage the scope of the certification process. Exploring these two dimensions is also sufficient to understand the certification procedure and draw the appropriate conclusions.

Dimension	Risk Category	Reasoning
Fairness	Medium	Processes personal data (facial images), output linked to personal behaviour but has limited impact on rights.
Autonomy and Control	Medium	Temporarily influences user perception during the interactive film experience.
Transparency	Low	Lack of transparency does not affect safety or usefulness significantly.
Reliability	Medium	Misclassification can degrade user experience but does not cause major harm.
Safety and Security	Low	No risk of physical injury or significant financial damage from malfunction.
Data Protection	Medium	Uses personal data (live facial images), though not particularly sensitive or secret data.

Table 3 Protection Requirements Analysis Summary

5.5 Risk Analysis

The risk analysis forms the core of the certification process. We carried out this step by working through a questionnaire from the Fraunhofer Catalogue. The questions addressed different aspects of the selected risk dimensions. For each dimension, we covered topics such as data quality, model design, testing procedures, and operational considerations. A summary of the conclusion of our Risk Analysis for our investigated dimension can be seen in Table 4. Following the individual dimension analyses, we conducted a cross-dimensional assessment to identify potential trade-offs or interactions between the examined dimensions. This step is crucial to ensure a complete understanding of the AI system's performance and risks. In our sample certification, we were unable to conclude the certification attempt with a positive result, but found significant shortcomings of the AI Application.

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.

Table 4 Certification Summary for Evaluated Dimensions

In the following section, we discuss in detail the challenges we encountered during this certification process. Based on these experiences, we draw conclusions. Additionally, we discuss comparisons with two other certification catalogues to provide a broader perspective on AI certification methodologies.

6 Results and Main Findings

We performed the certification of the chosen facial emotion recognition system. The core of the certification process with the Fraunhofer Catalogue involves the Protection

Requirement Analysis and the Cross-dimensional assessment. Performing the certification allows us to highlight the challenges that arise from using this certification approach in general. It also brings attention to issues that make a work like this more difficult, such as finding a suitable AI system to certify in the first place. This work did not prove that the system complies with today's regulations, nor was that its intent. Our certification process itself did not cover all the necessary dimensions required to certify the system fully. The conclusion of the certification suggests that even the dimensions we closely examined were not sufficient for a successful certification. In a real certification scenario, we would communicate these shortcomings to the development team, so they can address the issues and allow the system to be certified. If full certification were feasible, the process would demonstrate the AI system's compliance with specific standards outlined in the Fraunhofer Catalogue. We describe all the challenges and potential improvements found in the following paragraphs.

6.1 Selecting the AI System

In a conventional certification scenario, the process of selecting an AI system for certification is usually not a consideration, as the system to be certified is predetermined by the organization seeking certification. However, for the purposes of this work, the selection of an appropriate AI system represented a crucial first step that significantly influenced the subsequent certification process and what can potentially be learned from it. The selected system provided a mostly robust foundation for the certification effort. Its existing application context, and good documentation of both the AI model and its surrounding system, enabled meaningful progress through the certification process. During the certification process, we saw the importance of considering both technical factors and solid documentation when choosing an AI system for certification.

6.1.1 Initial Considerations and Challenges

The selection of an appropriate AI system requires careful consideration of multiple interconnected factors. While an initial approach might suggest identifying and selecting a standalone AI model or neural network, this proves insufficient when considering the comprehensive requirements of a certification processes. Particularly, the Fraunhofer Catalogue, which we primarily used in this paper, takes an extensive look at how to define the system and its boundaries with other software components. This definition cannot be found with a standalone AI model.

6.1.2 Context and Embedding Requirements

The certification process inherently demands a broader contextual framework than what might be immediately apparent. Rather than existing in isolation, the AI system must be embedded within a larger operational context and demonstrate clear use case applications. While it would be theoretically possible to construct artificial use cases for the certification purpose, such an approach could result in a suboptimal certification scenario. In this paper, therefore, we have chosen an AI application, which has already established a real-world application context. This characteristic proved invaluable, as it enabled a more natural translation into a certifiable system, providing the

necessary surrounding information to support a comprehensive certification approach. The existence of this practical context significantly enhanced the certification process's authenticity and relevance.

6.1.3 Documentation Requirements

Documentation emerged as another critical factor in the selection process, on two distinct levels. First, the AI model itself must be thoroughly documented, providing technical specifications and operational parameters. Second, and equally important, the surrounding system infrastructure must be comprehensively documented to make certification feasible. This documentation requirement significantly narrows the field of suitable candidates for certification studies. A particular challenge encountered in this work relates to the absence of a development team. When selecting an existing model for certification, there is typically no active development team invested in the certification process. This situation creates a significant constraint. Without the ability to request additional documentation or engage in an iterative process with developers, the available documentation must be sufficiently comprehensive from the outset. Any information gaps that arise during the certification process cannot be supplemented or clarified.

6.2 The Certification Process

We used the Fraunhofer catalogue as the primary basis for this certification due to its comprehensive and detailed nature. Although we considered other catalogues, such as the TÜV catalogue, they presented significant limitations. The incomplete publication of the TÜV catalogue made it unsuitable for use in the certification process. The Auditing Machine Learning Algorithms catalogue, while fully published and potentially suitable for certification, employs a substantially different approach compared to the Fraunhofer catalogue. Its less step-by-step nature potentially presents additional challenges for those with limited certification experience.

6.2.1 Main Challenges During the AI Certification Process

The system's documentation

We encountered several key challenges during the certification process. One fundamental challenge is that the system's documentation was not originally intended for certification purposes. Additionally, the development process did not require extensive and detailed documentation. This limitation created occasional gaps in documentation that would be essential for a complete certification. In some instances, we made adequate substitutions beforehand to create a more realistic certification scenario. The documentation of a system is key to certification. Our choice of a publicly available system that was not intended for certification has its shortcomings. This choice makes a sample certification, such as the one we attempted here, more difficult and potentially less meaningful.

No active development team

The absence of an active development team emerged as a critical limitation in the certification process. Without ongoing development support, we could not implement the typical feedback loop, where certification findings would normally lead to documentation improvements and system adjustments. In a standard certification scenario, identified gaps or shortcomings trigger an iterative process of enhancement, with the development team actively working to make the system more certifiable. However, in this paper, we had to evaluate the system purely based on its existing documentation and state. Therefore, we could only have two possible outcomes: either certifiable or not certifiable with the available materials.

This limitation became more complicated by the fact that the system was originally developed several years ago, and the entire development team had moved on from the project. The lead developer generously provided time to answer questions. However, because the project is old, certain details became less accessible or clear over time. This combination of inactive development and the system being old created a static evaluation scenario, rather than the dynamic, iterative process that typically characterizes successful certification efforts where development teams actively work towards certification compliance.

6.2.2 Specific Observations on the Fraunhofer Catalog

The implementation of the Fraunhofer catalogue revealed several notable characteristics and challenges. The catalogue's documentation-centric approach makes it nearly impossible to use for code-only projects, as it focuses exclusively on documentation rather than direct code examination. While code can inform the certification process and documentation creation, the catalogue never directly addresses or describes code. The catalogue's high specificity and detail provide comprehensive coverage, reducing the likelihood of overlooking critical aspects. However, this thoroughness occasionally results in similar or nearly duplicate questions, increasing the time required for certification completion. The strong documentation focus means less direct attention to mathematical or technical system operations. While the neural network structure remains important for documentation purposes, the Fraunhofer Catalogue only requires it to be examined implicitly rather than explicitly. This approach can be advantageous when dealing with proprietary information, as documentation alone might suffice for a potential certification. A particular strength of the Fraunhofer catalogue lies in its clear differentiation between the AI model, system, and embedding code, which proves crucial in determining certification scope and requirements. This distinction helps ensure appropriate certification coverage.

TÜV Catalogue

One notable limitation of the Fraunhofer catalogue is the lack of guidance on how to answer the posed questions. In this regard, a more technology-centric catalogue like the one from TÜV could provide valuable complementary guidance. The TÜV catalogue, despite its publication limitations and restricted focus on ML and supervised learning systems, offers useful insights into the technical aspects of ML systems operations.

Auditing Machine Learning Algorithms Catalogue

The Auditing Machine Learning Algorithms Catalogue presents a markedly different structural approach compared to Fraunhofer's. Its topic-based organization consolidates related questions – for instance, grouping all data-related questions together – contrasting with Fraunhofer's distributed approach where data-related questions appear across various subsections. This structural difference complicates the potential combination of these catalogues. However, the auditing catalogue's reduced duplication could potentially streamline the certification process.

6.3 Learnings and Recommendations

The certification process revealed several significant insights regarding both methodological approaches and practical certification challenges.

6.3.1 Catalogue-specific Observations

The Fraunhofer catalogue, while demonstrating robust effectiveness, revealed both strengths and limitations in practical application. Its exhaustive and detailed nature ensures comprehensive coverage, but is time intensive. This thoroughness, while beneficial for certification rigour, needs to be balanced against practical time constraints in real-world scenarios. The evaluation of alternative catalogues provided additional insights. The TÜV catalogue's incomplete publication status rendered it unsuitable for standalone certification efforts. The Auditing Machine Learning Algorithms catalogue showed promise for certification purposes, potentially offering a more streamlined approach compared to the Fraunhofer methodology. However, its less structured nature suggests a need for deeper AI system expertise. But it might potentially be a faster certification process while maintaining quality standards.

6.3.2 Limitations

Several key limitations emerged during the certification process. The absence of an active development team is limiting, as it prevented the implementation of the typical feedback loop essential for certification refinements. This limitation transformed the certification process into evaluation rather than an iterative improvement process, highlighting the importance of ongoing development support for successful certification efforts. Documentation gaps cannot be addressed through subsequent submissions. This emphasized the importance of comprehensive initial documentation of the chosen system.

Any actors performing a future sample certification of systems that are not actively developed should keep this in mind, and they should choose a system with the most comprehensive documentation.

6.3.3 Recommendations for future Sample Certifications

The experience gained from this study suggests several crucial considerations for future certification efforts. For AI application selection, auditors should identify AI systems that exist within a broader application setting with surrounding code. The AI system

itself, as well as the application setting and surrounding code, should be extensively documented. The ideal certification candidate should have an active development team willing to engage in the certification process.

A practical recommendation that emerged from this study was the value of creating a centralized archive of all available information and documentation before initiating the certification process.

6.3.4 Recommendations for Real-World Applications

Our findings yield several practical recommendations for real-world certification implementations. The Fraunhofer catalogue, while highly detailed and extensive, requires significant time investment for thorough completion. However, its precision and comprehensiveness make it a valuable tool for certification processes. Particularly noteworthy are the initial sections of the catalogue, specifically the AI lifecycle overview, which prove especially effective in providing auditors with comprehensive insights into the AI system's general functionality. This initial overview serves as an excellent starting point for any certification process. Although we did not extensively examine the Auditing Machine Learning Algorithms Catalogue in this work, its different structural approach suggests potential for more efficient certification processes. It could offer auditors a faster path to system certification while maintaining appropriate quality.

7 Conclusion and Future Research Directions

This work-in-progress presents our ongoing research into the practical application of AI certification catalogues, providing insights into both the capabilities and limitations of current certification approaches, as well as the challenges we encountered in certifying an open-source AI system. 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 in some areas.

In future works, other approaches and certification catalogues should be considered, to potentially streamline the process. The certification process highlights the critical importance of complete system documentation and active engagement from the development team. The absence of these elements can negatively impact the certification process and even make certification infeasible. During the certification attempt, we identified the key characteristics an AI system must possess to be adequately certified. Before selecting a public AI system for a sample certification, auditors should consider factors like thorough documentation and accessibility to the development team. These findings directly address the initial research objectives by identifying both useful aspects of this certification approach and areas for improvement. They also provide practical insights into the certification process and its limitations. Future work in this field could focus on developing more flexible certification methodologies that accommodate various system states and development scenarios.

References

[1] Costa, C.J., Aparicio, M.: Applications of Data Science and Artificial Intelligence. Applied Sciences **13**(15), 9015 (2023) https://doi.org/10.3390/app13159015 . Number: 15 Publisher: Multidisciplinary Digital Publishing Institute

[2] Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J.: Ai in health and medicine. Nature medicine **28**(1), 31–38 (2022)

[3] Broek, E., Sergeeva, A., Huysman, M.: When the machine meets the expert: An ethnography of developing ai for hiring. MIS quarterly **45**(3) (2021)

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

[5] Kowald, D., Dennerlein, S., Theiler, D., Walk, S., Trattner, C.: The social semantic server: A framework to provide services on social semantic network data. In: 9th International Conference on Semantic Systems, I-SEMANTICS 2013, pp. 50–54 (2013). CEUR

[6] Lacic, E., Kowald, D., Seitlinger, P.C., Trattner, C., Parra, D.: 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, pp. 4–9 (2014). Association of Computing Machinery

[7] Lacic, E., Kowald, D., Parra, D., Kahr, M., Trattner, C.: Towards a scalable social recommender engine for online marketplaces: The case of apache solr. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 817–822 (2014)

[8] Lacic, E., Kowald, D., Eberhard, L., Trattner, C., Parra, D., Marinho, L.B.: Utilizing online social network and location-based data to recommend products and categories in online marketplaces. In: Mining, Modeling, and Recommending'Things' in Social Media: 4th International Workshops, MUSE 2013, Prague, Czech Republic, September 23, 2013, and MSM 2013, Paris, France, May 1, 2013, Revised Selected Papers, pp. 96–115 (2015). Springer

[9] Kasinidou, M., Kleanthous, S., Busso, M., Rodas, M., Otterbacher, J., Giunchiglia, F.: 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, pp. 24–30, New York, NY, USA (2024)

[10] Baum, K., Bryson, J., Dignum, F., Dignum, V., Grobelnik, M., Hoos, H., Irgens, M., Lukowicz, P., Muller, C., Rossi, F., Shawe-Taylor, J., Theodorou, A., Vinuesa, R.: From fear to action: AI governance and opportunities for all. Frontiers in Computer Science **5** (2023). Publisher: Frontiers

[11] Kowald, D., Scher, S., Pammer-Schindler, V., Müllner, P., Waxnegger, K., Demelius, L., Fessl, A., Toller, M., Mendoza Estrada, I.G., Šimić, I., *et al.*: Establishing and evaluating trustworthy ai: overview and research challenges. Frontiers in Big Data **7**, 1467222 (2024)

[12] Smuha, N.A.: From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. Law, Innovation and Technology **13**(1), 57–84 (2021) https://doi.org/10.1080/17579961.2021.1898300 . Publisher: Routledge _eprint: https://doi.org/10.1080/17579961.2021.1898300. Accessed 2024-10-28

[13] Mueck, M., Cadzow, S., Communications, C., Wood, S.: ETSI Activities in the field of Artificial Intelligence Preparing the implementation of the European AI Act - 1st Edition – December -2022. Technical report, ETSI (2022)

[14] Nad, T., Scher, S., Königstorfer, F.: Trustworthiness of AI. Technical Report NIST AI 100-1, SGS (June 2023)

[15] Winter, P.M., Eder, S., Weissenböck, J., Schwald, C., Doms, T., Vogt, T., Hochreiter, S., Nessler, B.: Trusted Artificial Intelligence: Towards Certification of Machine Learning Applications. arXiv. arXiv:2103.16910 [cs, stat] (2021)

[16] Poretschkin, D.M., Schmitz, A., Akila, D.M., Adilova, L., Becker, D.D., Cremers, D.A.B., Hecker, D.D., Houben, D.S., Rosenzweig, J., Sicking, J., Schulz, E., Voss, D.A., Wrobel, D.S.: AI Assessment Catalog. Technical report, Fraunhofer IAIS (February 2023)

[17] Pimentel, B.: Why AI still needs regulation despite impact (2024). https://legal.thomsonreuters.com/blog/why-ai-still-needs-regulation-despite-impact/ Accessed 2024-10-28

[18] Benjamens, S., Dhunnoo, P., Mesko, B.: The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. npj Digital Medicine **3** (2020) https://doi.org/10.1038/s41746-020-00324-0

[19] Saenz, A.D., Harned, Z., Banerjee, O., Abràmoff, M.D., Rajpurkar, P.: Autonomous AI systems in the face of liability, regulations and costs. npj Digital Medicine **6**(1), 1–3 (2023) https://doi.org/10.1038/s41746-023-00929-1 . Publisher: Nature Publishing Group. Accessed 2024-06-01

[20] European-Union: 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. Legislative Body: CONSIL, EP (2024). http://data.europa.eu/eli/reg/2024/1689/oj/eng Accessed 2024-07-20

[21] Fernhout, F., Duquin, T.: The EU Artificial Intelligence Act: our 16 key takeaways | Stibbe (2024). https://www.stibbe.com/publications-and-insights/the-eu-artificial-intelligence-act-our-16-key-takeaways Accessed 2024-08-07

[22] Madiega, T.: Artificial intelligence liability directive (2023)

[23] European-Union: Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496 Accessed 2024-11-16

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

[25] Digital, M..S. Culture: Establishing a pro-innovation approach to regulating AI (2022). https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement Accessed 2024-06-02

[26] Blösser, M., Weihrauch, A.: A consumer perspective of AI certification – the current certification landscape, consumer approval and directions for future research. European Journal of Marketing **58**(2), 441–470 (2023). Publisher: Emerald Publishing Limited

[27] Delgado-Aguilera Jurado, R., Ye, X., Ortolá Plaza, V., Zamarreño Suárez, M., Pérez Moreno, F., Arnaldo Valdés, R.M.: An introduction to the current state of standardization and certification on military AI applications. Journal of Air Transport Management **121**, 102685 (2024)

[28] Benjamin Fresz, Vincent Philipp Göbels, Safa Omri, Danilo Brajovic: The Contribution of XAI for the Safe Development and Certification of AI: An Expert-Based Analysis (2024). https://arxiv.org/html/2408.02379v1 Accessed 2024-11-16

[29] Hadrien Pouget: Standardsetzung | EU-Gesetz zur künstlichen Intelligenz (2024). https://artificialintelligenceact.eu/de/standardeinstellung/ Accessed 2024-11-16

[30] SAI-FI-DE-NL-NO-UK: Auditing machine learning algorithms. Technical report, Supreme Audit Institutions FI, DE, NL, NO, UK (February 2023)

[31] Angelica Perez, Julien Deswaef, Puneetha Pai: thoughtworksarts/EmoPy. Thoughtworks Arts. original-date: 2017-12-20T02:19:22Z (2021). https://github.com/thoughtworksarts/EmoPy Accessed 2024-07-23

[32] Thoughtworks: RIOT | Thoughtworks Arts (2018). https://thoughtworksarts.io/projects/riot/ Accessed 2024-10-29

[33] Semmelrock, H., Kopeinik, S., Theiler, D., Ross-Hellauer, T., Kowald, D.: Reproducibility in machine learning-driven research. arXiv preprint arXiv:2307.10320 (2023)

[34] Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., Kowald, D.: Reproducibility in machine learning-based research: Overview, barriers and drivers. arXiv preprint arXiv:2406.14325 (2024)

[35] Perez, A.: EmoPy: a machine learning toolkit for emotional expression (2018). https://www.thoughtworks.com/insights/blog/emopy-machine-learning-toolkit-emotional-expression Accessed 2024-10-29

[36] Palmer, K.: RIOT AI (2016). http://karenpalmer.uk/portfolio/riot/ Accessed 2024-10-29

[37] Thoughtworks: thoughtworksarts/riot. Thoughtworks Arts. original-date: 2017-11-22T15:59:17Z (2019). https://github.com/thoughtworksarts/riot Accessed 2024-10-29

[38] TED Residency: Karen Palmer: The film that watches you back (2018). https://www.youtube.com/watch?v=Rw8gLEkFdSw Accessed 2024-10-29

[39] Thoughtworks: Thoughtworks Arts Exhibition at SPRING/BREAK for Armory Week 2018 | Thoughtworks Arts (2018). https://thoughtworksarts.io/blog/thoughtworks-arts-exhibition-spring-break-armory-week/ Accessed 2024-10-29

[40] Thoughtworks: Karen Palmer Awarded Thoughtworks AI Residency | Thoughtworks Arts (2017). https://thoughtworksarts.io/blog/karen-palmer-ai-residency/ Accessed 2024-10-29

[41] Karen Palmer: RIOT Video (2017). https://www.youtube.com/watch?v=-BCny9SuI3A Accessed 2024-10-29

[42] Perez, A.: Recognizing human facial expressions with machine learning (2018). https://www.thoughtworks.com/insights/articles/recognizing-human-facial-expressions-machine-learning Accessed 2024-10-29

[43] Team, E.D.: Welcome to EmoPy's documentation! — EmoPy 1.0 documentation (2017). https://emopy.readthedocs.io/en/latest/ Accessed 2024-10-29

[44] Microsoft: microsoft/FERPlus. Microsoft. original-date: 2016-09-14T06:35:21Z (2023). https://github.com/microsoft/FERPlus Accessed 2024-10-29

[45] Cohn, J.: Resources – Jeffrey Cohn (2024). https://www.jeffcohn.net/resources/ Accessed 2024-10-29