Editorial: AI Certification, Fairness and Regulations (CERT AI 2026)

Bernhard Nessler[†] **Michal Lewandowski**[†] **Simon Schmid**[†] **Gregor Aichinger**[†] **Iana Kazeeva**[†] **Rania Wazir**[‡]

[†]Software Competence Center Hagenberg (SCCH) [‡]leiwand.ai

Abstract

The Certification of Artificial Intelligence Systems (CERT AI) workshop, held under the title *AI Certification, Fairness and Regulations* at the Austrian Symposium on Artificial Intelligence, Robotics and Vision (AIRoV) 2026, provides a forum for discussing how artificial intelligence (AI) systems can be evaluated, documented, and monitored in ways that support certification and trustworthy deployment. Its central concern is the operationalisation of certification: translating legal, ethical, and organisational requirements into measurable, testable, and statistically valid properties of AI systems. This editorial introduces the motivation, scope, and programme of the workshop. We argue that certification cannot be reduced to a final compliance check, but must be understood as a lifecycle-oriented process connecting intended purpose, application-domain definition, risk management, technical evidence, fairness assessment, privacy protection, explainability, robustness, standardisation, conformity assessment, and post-market monitoring. The workshop contributions illustrate this breadth, ranging from statistical application-domain modelling and General Data Protection Regulation (GDPR)-compliant machine learning to multi-user conversational agents, safety architectures for autonomous surface vessels, explainable model selection, anthropomorphic terminology, and the ongoing standardisation process in the European Committee for Standardization and European Committee for Electrotechnical Standardization (CEN/CENELEC). Together, these perspectives point toward a practical research agenda for functional trustworthiness: AI systems should be certified not only by what they claim to do, but by what can be verified about their behaviour in the contexts in which they are deployed.

1 Introduction

Artificial intelligence systems are increasingly used in domains where errors, bias, opacity, or operational instability can affect safety, fundamental rights, economic participation, and public trust. As part of AIRoV 2026, the workshop contributes to the symposium objective of bringing together the Austrian artificial intelligence, robotics, and vision communities [2]. The European Artificial Intelligence Act (AI Act) has strengthened the need for systematic approaches to trustworthy AI by establishing a risk-based regulatory framework for AI developers and deployers [6, 3]. In parallel, harmonised standards are being developed to translate legal requirements into a technical language that can be used by developers, auditors, notified bodies, and market-surveillance authorities [4].

The resulting challenge is not only regulatory. It is also scientific and engineering-oriented. Many requirements associated with trustworthy AI, such as robustness, transparency, human oversight, fairness, cybersecurity, data quality, and post-market monitoring, cannot be assessed meaningfully without a precise account of the AI system's intended purpose and operational context. A certification claim therefore depends on evidence: What is the system expected to do? Under which

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

assumptions? For which users? In which environment? With which failure modes? With which statistical confidence? And how is continued compliance monitored once the system changes or is used in a changing environment?

The CERT AI workshop addresses this translation problem. It focuses on the operationalisation of AI certification, fairness, and regulation by asking how abstract legal and ethical requirements can become measurable, testable, and statistically valid system properties across the AI system lifecycle [1]. The 2026 edition builds on the workshop's earlier role as an interdisciplinary forum and places particular emphasis on *functional trustworthiness*: the alignment between intended purpose, operational context, and verifiable performance and risk criteria.

2 Workshop Scope

The workshop brings together technical researchers, legal scholars, certification experts, and practitioners concerned with conformity assessment. Its scope covers both methodological and application-oriented questions. On the methodological side, central topics include certification procedures for machine learning and AI systems, statistical testing, sequential testing, online hypothesis testing, multiple-testing control, benchmarking, reproducibility, auditability, and lifecycle risk management. On the technical side, the workshop addresses robustness, adversarial resilience, safety evaluation, explainability, transparency verification, privacy-preserving AI, differential privacy, and bias or discrimination assessment. On the regulatory side, it considers standardisation, harmonised standards under the AI Act, conformity assessment procedures, and post-market surveillance [1].

This breadth reflects a practical fact: certifiability is not a single property of an algorithm. It emerges from the relation between the model, data, system architecture, intended use, organisational controls, monitoring procedures, and the legal or sector-specific framework in which the system is deployed. A model may perform well on a benchmark but still fail to be certifiable if its application domain is vague, its data governance is insufficient, its human oversight concept is not operational, or its monitoring plan cannot detect relevant degradation after deployment.

3 Programme and Contributions

The CERT AI 2026 programme is organised as a three-hour workshop within AIRoV 2026. It includes opening remarks, two invited talks, five contributed presentations, a general discussion, and closing remarks [1]. The programme structure is intentionally interdisciplinary: legal and privacy questions are discussed alongside technical certification methods, system-safety architectures, human–AI interaction, explainability, and AI terminology.

The first invited talk, *GDPR-compliant Machine Learning and the Use of Personal Data in Large Language Models*, by Gregor Aichinger, addresses the interaction between machine learning practice and data-protection requirements. This topic is central for certification because many AI systems rely on personal or sensitive data during training, evaluation, adaptation, or monitoring. Compliance with the General Data Protection Regulation (GDPR) is therefore not an external legal add-on, but a design constraint for data pipelines, model development, deployment, and documentation [5].

The second invited talk, *The Current State of the AI Standardisation Process in CEN/CENELEC*, by Rania Wazir, connects certification research to the emerging standardisation landscape. Here, CEN/CENELEC refers to the European Committee for Standardization and the European Committee for Electrotechnical Standardization. This connection is critical because harmonised standards can provide a practical route from legal requirements to technical implementation and evidence generation. The European Commission has requested standardisation work in areas including risk management, data governance, record keeping, transparency, human oversight, accuracy, robustness, cybersecurity, quality management, and conformity assessment [4].

The contributed papers represent complementary views on certifiable AI. *Stochastic Application Domain Definition for Functional Trustworthiness Certification of AI Systems*, by Simon Schmid, focuses directly on the problem of defining the application domain in a way that supports statistical certification claims. This is a core issue for functional trustworthiness because system performance is only meaningful relative to the conditions under which the system is expected to operate.

Conversational Agents in Multi-User Environments, by Umut Tanriverdi and Tobias Halmdienst, addresses socially complex AI systems. Conversational agents are often evaluated through local response quality, but multi-user environments require additional evidence about consistency, social robustness, timing, interaction dynamics, and the ability to maintain plausible behaviour across several participants. This contribution highlights that certifiability also concerns interactional and behavioural properties, not only static predictive accuracy.

Safety Driven Hardware and Control Architecture for Automated Surface Vessel Systems, by Önder Hamamcioglu and Viktor Komyshan, turns the focus toward embodied and safety-critical AI-enabled systems. Automated surface vessels combine perception, control, hardware constraints, and environmental uncertainty. Certification in such systems must therefore integrate AI evaluation with system engineering, fault handling, operational design domains, and safety-oriented architecture.

Explainable Selection of Machine Learning Algorithms in Social Sciences, by Dijana Oreski, addresses explainability from a methodological perspective. In social-science applications, model choice can affect interpretability, validity, and downstream decisions. Explainable selection of machine learning algorithms is therefore relevant for certification because it supports transparent justification of why a model class is appropriate for a given empirical and operational context.

Anthropomorphic Terminology in Artificial Intelligence, by Iana Kazeeva, examines the language used to describe AI systems. This topic is directly relevant to regulation and certification because terminology can shape expectations, risk perception, documentation, and accountability. Claims that an AI system "understands", "reasons", or "decides" may obscure the actual technical mechanism and may lead to misleading communication unless such terms are carefully defined.

4 Emerging Themes

Several common themes connect the workshop contributions. The first is the need to define application domains precisely. Certification evidence is weak if it is disconnected from the environment in which an AI system is used. This is particularly important for systems that operate under distribution shift, interact with humans, or are embedded in larger socio-technical processes.

The second theme is statistical validity. Certification requires more than demonstration examples. It requires claims about performance, uncertainty, failure probability, bias, robustness, or degradation that are supported by suitable sampling assumptions, testing protocols, and monitoring strategies. This creates a bridge between machine learning evaluation, statistical quality assurance, and regulatory evidence.

The third theme is lifecycle orientation. AI systems may change through updates, retraining, new data streams, user adaptation, or environmental drift. For this reason, certification should be understood as an ongoing process. The AI Act explicitly connects high-risk AI systems to documentation, risk management, quality management, logging, human oversight, robustness, cybersecurity, and post-market monitoring obligations [6, 3]. The workshop therefore treats monitoring and auditability as central design concerns rather than late-stage compliance tasks.

The fourth theme is interdisciplinary translation. Legal concepts must become technical requirements, technical tests must become auditable evidence, and evidence must be interpretable by regulators, domain experts, developers, and affected stakeholders. This translation requires shared terminology, explicit assumptions, and methods that can be scrutinised across disciplinary boundaries.

5 Outlook

The operationalisation of AI certification remains an open and rapidly developing field. The next step is to move from general principles toward reusable methods, reference processes, and evidence patterns. Promising directions include statistically grounded application-domain definitions, reproducible certification benchmarks, structured assurance cases, continuous monitoring methods, fairness and robustness tests with explicit uncertainty estimates, privacy-preserving evaluation protocols, and documentation formats that connect model-level evidence to system-level risk claims.

For the AIRoV community, this workshop also highlights a broader opportunity. AI, robotics, and vision systems increasingly share certification challenges: perception models must be robust under changing conditions, robotic systems must remain safe under physical uncertainty, and interactive AI systems must behave reliably in social contexts. CERT AI therefore aims to support a continuing exchange between technical research, legal interpretation, standardisation, and industrial practice. The goal is not merely to make AI systems compliant after development, but to make certifiability a design principle from the beginning.

6 Conclusion

The CERT AI 2026 workshop positions certification as a practical bridge between trustworthy AI principles and deployable AI systems. Its central message is that trustworthy deployment requires operational evidence. Legal and ethical requirements must be translated into measurable properties; system behaviour must be evaluated within a defined context; and compliance must be maintained throughout the lifecycle. By bringing together contributions on application-domain definition, privacy, standardisation, conversational agents, safety architectures, explainable model selection, and terminology, the workshop provides a cross-disciplinary agenda for AI systems that can be tested, audited, monitored, and trusted in practice.

Acknowledgments

We thank the AIRoV 2026 organisers, invited speakers, authors, reviewers, and participants for supporting the CERT AI workshop. The organisational work of the SCCH contributors was connected to research activities funded by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK), the Austrian Federal Ministry of Labour and Economy (BMAW), and the State of Upper Austria in the frame of the SCCH competence center INTEGRATE [(Austrian Research Promotion Agency (FFG) grant no. 892418)] as part of the FFG COMET (Competence Centers for Excellent Technologies) Program, and by Upper Austria's #upperVISION2030 business and research strategy in the frame of the AI Engineering and Certification Center, no. Wi-2022-699557-Hub.

References

[1] AIRoV 2026. Ai certification, fairness and regulations, 2026.

[2] AIRoV 2026. Airov – austrian symposium on ai, robotics and vision 2026, 2026.

[3] European Commission. Ai act, 2026.

[4] European Commission. Standardisation of the ai act, 2026.

[5] European Parliament and Council of the European Union. Regulation (eu) 2016/679 on the protection of natural persons with regard to the processing of personal data, 2016.

[6] European Parliament and Council of the European Union. Regulation (eu) 2024/1689 laying down harmonised rules on artificial intelligence, 2024.