



Enhancing Robustness and Security of Edge AI Systems for Safety-Critical Applications

WP5 – Impact Maximization: Dissemination, Exploitation, and Certification Roadmap

D5.3: K4AI Portal



This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101168067. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Document Information

GRANT AGREEMENT NUMBER	101168067	ACRONYM	GuardAI
FULL TITLE	Enhancing Robustness and Security of Edge AI Systems for Safety-Critical Applications		
START DATE	1 st October 2024	DURATION	36 months
PROJECT URL	https://www.kios.ucy.ac.cy/guardai/		
DELIVERABLE	D5.2: K4AI Portal		
WORK PACKAGE	WP5 – Impact Maximization: Dissemination, Exploitation, and Certification Roadmap		
DATE OF DELIVERY	CONTRACTUAL	05/2025	ACTUAL 05/2025
TYPE	Report	DISSEMINATION LEVEL	PU
LEAD BENEFICIARY	UCY-KIOS CoE		
RESPONSIBLE AUTHORS	Christos Kyrkou, Rafaella Elia, Antonis Savva		
CONTRIBUTIONS (FROM)	Christos Kyrkou (UCY-KIOS CoE), Rafaella Elia (UCY-KIOS CoE), Antonis Savva (UCY-KIOS CoE)		
ABSTRACT	<p>This deliverable presents the finalized structure and some preliminary content and functionality of the Knowledge Hub for Secure and Resilient AI (K4AI) Portal under the GuardAI project. The K4AI Portal will serve as a centralized digital knowledge base to support the development, evaluation and certification of secure and resilient Edge AI systems. It will provide access to open data, open-source code, best practices, and certification-oriented guidelines.</p>		

Document History

VERSION	ISSUE DATE	STAGE	DESCRIPTION	CONTRIBUTOR
V 0.1	05/02/2025	Draft	Table of contents	Christos Kyrkou (UCY-KIOS CoE), Antonis Savva (UCY-KIOS CoE)
V 0.2	28/04/2025	Draft	Initial Contributions	Christos Kyrkou (UCY-KIOS CoE), Rafaella Elia (UCY-KIOS CoE), Antonis Savva (UCY-KIOS CoE)
V 0.3	14/05/2025	Draft	Final Contributions	Christos Kyrkou (UCY-KIOS CoE), Rafaella Elia (UCY-KIOS CoE), Antonis Savva (UCY-KIOS CoE)
V 0.4	26/05/2025	Draft	Comments from internal reviewers	Stelios Erotokritou (EIGHT BELLS LTD), Joseph Karis (CYS)
V 0.5	26/05/2025	Draft	Edits after internal review process	Christos Kyrkou (UCY-KIOS CoE), Rafaella Elia (UCY-KIOS CoE), Antonis Savva (UCY-KIOS CoE)
V 1.0	30/05/2025	Final	Final edits	Christos Kyrkou (UCY-KIOS CoE), Rafaella Elia (UCY-KIOS CoE), Antonis Savva (UCY-KIOS CoE)

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© GuardAI Consortium, 2025

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised provided the source is acknowledged.

Glossary of Acronyms

Acronym	Meaning
AI	Artificial Intelligence
K4AI	Knowledge Hub for Secure and Resilient AI
SMEs	Small – Medium Enterprises
OJEU	Official Journal of the European Union
WPs	Work Packages

Table of Contents

Executive summary	6
1. Introduction	7
1.1. Purpose	7
1.2. Target Audience	7
1.3. Expected Impact	8
2. Structure and Functionality	9
2.1. Structure	9
2.2. Content Update	10
2.2.1. Update Responsibility	10
2.2.2. Update Frequency	10
2.2.3. Content Validation and Quality Assurance	10
3. Open Data and Open-Source Code	11
3.1. Open Data	11
3.2. Open-Source Code	12
4. Knowledge Hub	13
4.1. Core Principles in Edge AI	13
4.2. Research Publications	15
4.3. Technical Reports and Guidelines	18
5. Certification and Compliance	19
6. Best Practices	20
7. Conclusion	22

Executive summary

This document presents GuardAI's deliverable on Knowledge Hub for Secure and Resilient AI - K4AI Portal. The deliverable D5.3 presents the design, structure and functionality of the knowledge platform developed within the project. The K4AI Portal will serve as a central repository dedicated to advancing secure AI practices and facilitating the adoption of certified algorithms for Edge AI systems deployed in safety-critical environments. It will provide, comprehensive resources to address the key challenges related to assessing the security and resiliency of Edge AI technologies, which are often restricted by the absence of standardized evaluation criteria and specialized expertise. The portal offers a structured and collaborative knowledge space containing open data, open-source tools, core AI security principles, certification-aligned resources, and best practices for high-risk application domains. This document details the portal's structure, thematic pillars, and contribution to long-term sustainability and knowledge transfer.

1. Introduction

This section outlines the rationale for developing the K4AI portal and its positioning within the Guard AI dissemination and certification roadmap. By centralizing resources and offering a shared knowledge infrastructure, the portal enables more effective communication between AI developers, regulators, and industry stakeholders.

1.1. Purpose

The portal's primary goal is to provide a comprehensive and accessible environment for sharing, accessing and contributing knowledge on secure and resilient AI practices. It supports the implementation and validation of robust Edge AI systems by offering domain-specific guidelines, evaluation tools, and security principles aligned with emerging certification requirements. The purpose of the K4AI Portal is to act as a centralized Knowledge Hub advancing secure AI practices and encouraging the adoption of resilient algorithms in Edge AI applications.

1.2. Target Audience

The K4AI Portal targets a variety of stakeholders involved AI product companies (startups, SMEs), academic and research communities, certification bodies, standardization organizations, EU and national policy and regulatory bodies, corporate executives and strategists, and investors. Each group benefits from tailored resources aiming to support the development, evaluation, certification, and safe adoption of secure Edge AI systems. Table 1 presents the target groups and the relevance to the K4AI Portal.

Table 1: K4AI Portal - Target Audience.

Target Group	Relevance to K4AI Portal	Focus
AI Product Companies (startups, SMEs)	Startups and SMEs building Edge AI products will benefit from access to best practices, guidelines, and compliance insights, which are essential for secure and trustworthy AI deployments.	Access to technical knowledge, certification-readiness guides, and industry-specific recommendations.
Academic & Research Communities	Researchers and academic institutions will use the portal to stay informed on the latest advances in AI robustness, security, and certification methodologies.	Research publications, domain - specific guidelines, training resources, and collaboration opportunities.
Certification Authorities & Standardization Bodies	Certification professionals and experts in standardization will be able to find resources that align Edge AI development practices with emerging certification frameworks.	Knowledge base on certification practices, security standards mapping, and reference architectures.

EU and National Policy & Regulatory Bodies	Although not the primary users, regulatory and policymaking communities can use the portal's materials to understand technical trends and practical challenges in regulating Edge AI technologies.	High-level briefs, best practices for regulatory compliance, contributions to standardization discussion.
Corporate Executives & Strategists	Executives seeking to align their AI strategies with upcoming security and certification requirements can consult the portal for strategic insights into trustworthy Edge AI deployment.	Summarized best practices, impact studies, and certification-readiness indicators for business strategy.
Investors & Funding Bodies	Investors interested in AI startups and technologies can use the portal to evaluate the robustness, security practices, and market readiness of Edge AI solutions.	Sectoral analysis, innovation trends, and certification maturity insights supporting investment decisions.

1.3. Expected Impact

The K4AI Portal provides a unified framework for assessing safety, security, and robustness, to enhance communication among developers, regulators and end-users, enable consistent evaluation practices across industries, and promote the widespread adoption of secure Edge AI solutions. It will also serve as a foundation for certification initiatives, thereby supporting trust and accountability in high-risk sectors such as monitoring of critical infrastructures, transportation, and communication networks. Through the dissemination of robust principles, best practices, and certification guidance, K4AI is expected to:

- Enable broader understanding and adoption of robust AI security practices.
- Support harmonization of certification frameworks in Europe.
- Contribute to the development of resilient AI ecosystems.
- Strengthen stakeholder collaboration across disciplines.
- Facilitate post-project knowledge continuity and the long-term sustainability of GuardAI's outputs.

2. Structure and Functionality

The K4AI Portal is designed to provide a comprehensive suite of resources that will support the entire AI lifecycle, from research and development to deployment and certification.

2.1. Structure

The portal is organized into four main sections, each designed to meet the needs of a specific user group, from developers to regulators and end-users. These sections aim to foster a dynamic environment for collaboration, learning, and the integration of secure AI practices.

The four main sections of the K4AI portal are:

- **Open Data and Open-Source Code:** Access to datasets and source code for Edge AI, safety-critical domains.
- **Knowledge Hub:** Foundational principles, methodologies, and domain specific guidelines for building secure and robust AI systems.
- **AI Certification:** Roadmaps, templates, and supporting resources for aligning Edge AI solutions with certification standards.
- **Best Practices:** Practical guidelines, case studies and implementation frameworks across different industry sectors.

Each section is designed to integrate seamlessly with others, providing a user-centric interface that simplifies access to valuable information, tools, and resources. Figure 1 shows the main sections of the K4AI Portal, which are described in the following sections.

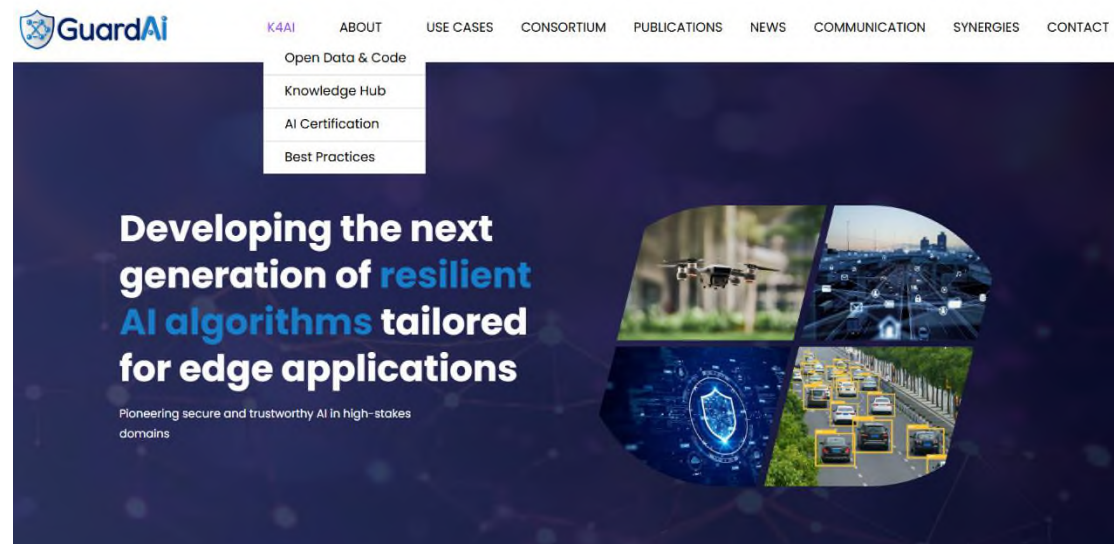


Figure 1: K4AI Portal - Main Sections.

2.2. Content Update

To ensure the continued relevance, accuracy, and impact of the K4AI Portal, a structured content update has been established. This plan outlines the responsible partners, and updated cycles.

2.2.1. Update Responsibility

Content development and curation are coordinated by the KIOS Research and Innovation Center of Excellence (UCY-KIOS CoE), as the lead beneficiary of Task 5.1. UCY-KIOS CoE is responsible for hosting the platform, managing editorial workflows, and ensuring content consistency.

Content development is supported by the GuardAI Consortium Members across multiple Work Packages (WPs):

- **Technical partners** contribute datasets, code and validation tools developed in WP2-WP4.
- **Dissemination and certification teams** ensure that reports, webinars, and certification resources remain up to date.
- **Standardization experts** (e.g. CYS) provide validation for content related to certification frameworks.

2.2.2. Update Frequency

To maintain this knowledge base, content will be reviewed and updated on a biannual basis during the project lifecycle. This includes:

- Addition of new publications, webinars, or technical reports.
- Updates to standardisation and certification mapping as regulatory frameworks evolve (e.g. EU AI Act).
- Update of new datasets and code repositories.

2.2.3. Content Validation and Quality Assurance

All content on the K4AI portal is reviewed by UCY-KIOS CoE before publication. The review checks that the material is relevant, properly attributed, and meets open-access and ethical standards. Content is also tagged by topic or domain to make it easy to find. This process helps maintain high quality while ensuring quick and efficient updates.

3. Open Data and Open-Source Code

This section of the K4AI Portal serves as a practical resource base to support experimentation, benchmarking, and the secure development of Edge AI systems. By sharing project-relevant datasets and curated software tools, this section helps promote reproducibility and community-driven development in secure Edge AI. **Error! Reference source not found.** represents the Open Access section of the K4AI Portal, which includes the Datasets and Source Code.

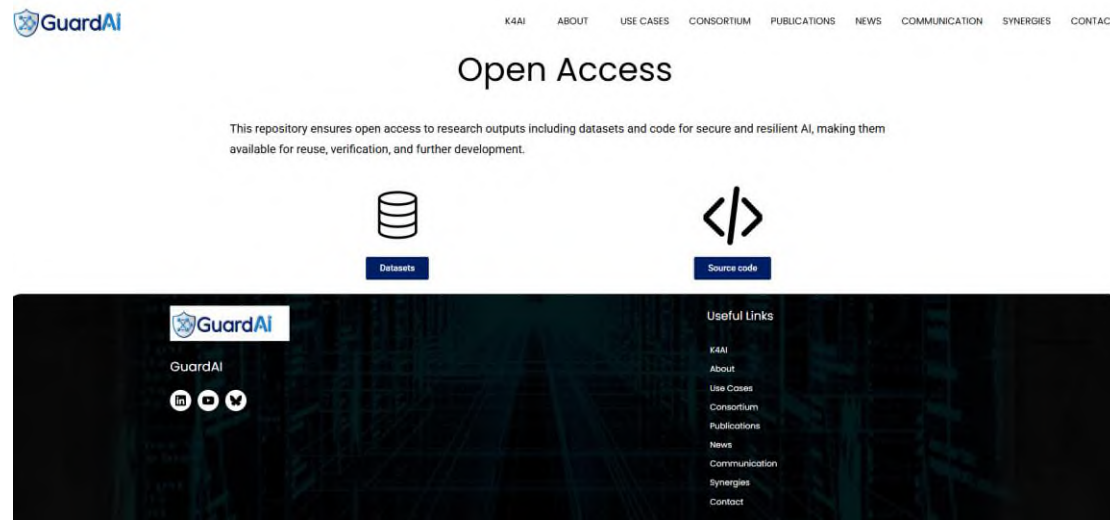



Figure 2: Open Access Section: Open Data and Open-Source Code.

3.1. Open Data

A key section of the K4AI Portal is the promotion of open data, enabling all stakeholders to benefit from high-quality, publicly available datasets. The portal hosts a collection of datasets that are relevant to the testing and certification of Edge AI technologies. These include both synthetic and real-world datasets, as well as curated datasets sourced from publicly available repositories and related EU projects. These datasets support a variety of tasks such as model training, validation, and performance evaluation under realistic and diverse conditions. This section of the K4AI portal is a key enabler for reproducibility, and collaboration in the development of secure and resilient Edge AI systems. Figure 3 shows the Open Access – Datasets subsection of the Open Data Section from the portal.



[K4AI](#)
[ABOUT](#)
[USE CASES](#)
[CONSORTIUM](#)
[PUBLICATIONS](#)
[NEWS](#)
[COMMUNICATION](#)
[SYNERGIES](#)
[CONTACT](#)

Open Access: Datasets


This repository ensures open access to research outputs including datasets and code for secure and resilient AI, making them available for reuse, verification, and further development.

Dataset	Partner	Purpose	Link
AIDerv2 (Aerial Image Dataset for Emergency Response Applications)	UCY-KIOS CoE	This dataset will be applied to assess the suitability of different methods for detecting objects of interest and understand whether additional contextual data is needed. Also, it will be used to evaluate how adversarial attacks affect model performance qualitatively.	Zenodo
Multi-Altitude Aerial Vehicles	UCY-KIOS CoE	This dataset will be used to train models capable of robustly detecting objects of interest at various altitudes on edge devices.	Zenodo
Small Object Aerial Person Detection	UCY-KIOS CoE	This dataset will aid in training models for robust object detection at various altitudes, with a focus on detecting small, hard-to-identify objects on edge devices.	Zenodo

Figure 3: Open Data Sub-section.

3.2. Open-Source Code

The Open-Source code section of the K4AI Portal provides a practical and actively maintained repository of accessible tools and frameworks. This repository is essential for enabling developers, researchers, and certification bodies to implement and validate trustworthy AI solutions. Figure 4 shows the Source code subsection from the Open Data section.



[K4AI](#)
[ABOUT](#)
[USE CASES](#)
[CONSORTIUM](#)
[PUBLICATIONS](#)
[NEWS](#)
[COMMUNICATION](#)
[SYNERGIES](#)
[CONTACT](#)

Open Access: Source-Code

This repository ensures open access to research outputs including datasets and code for secure and resilient AI, making them available for reuse, verification, and further development.

Category	Title	Description	Link
Adversarial Defences	Free Adversarial Training	This repository provides codes for training and evaluating the models on the ImageNet dataset. The implementation is adapted from the official PyTorch repository.	Github
Adversarial Defences	Friendly Adversarial Training	This repository provides codes for friendly adversarial training (FAT).	Github
Adversarial Defences	Pytorch Adversarial Training CIFAR	This repository provides simple PyTorch implementations for adversarial training methods on CIFAR-10.	Github

Figure 4: Open Source-Code Sub-section.

4. Knowledge Hub

The Knowledge Hub offers a wide array of information, resources, and guidelines related to Edge AI security and resilience. This section aims to guide developers, regulators, and researchers in understanding the principles, methodologies, and best practices needed to ensure the safe and responsible deployment of AI systems in high-risk sectors. Figure 5 below presents the view of the Knowledge Hub.

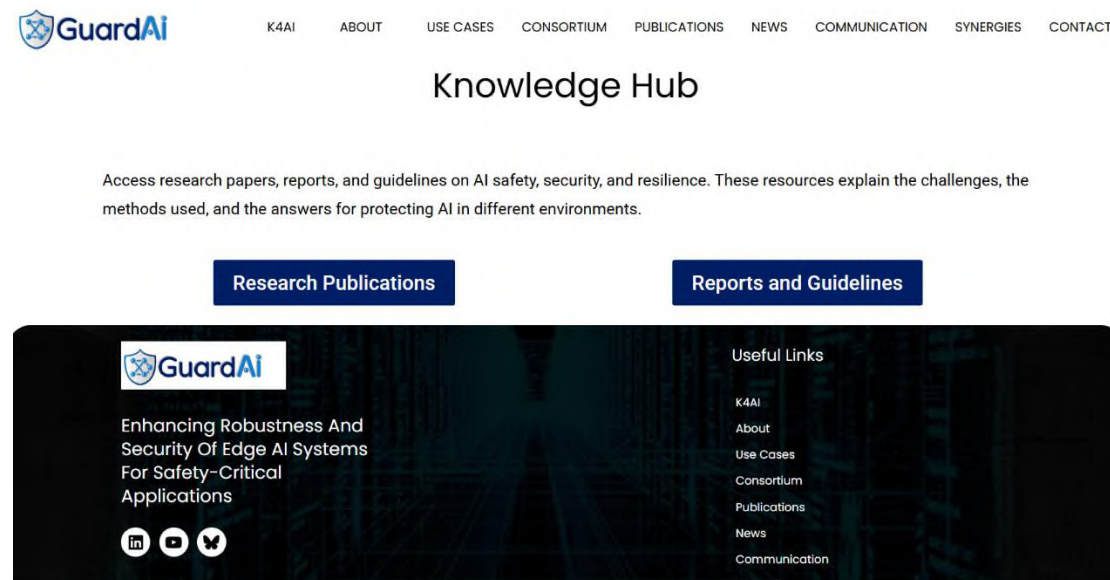


Figure 5: Knowledge Hub Section.

4.1. Core Principles in Edge AI

The Knowledge Hub emphasizes the foundational principles required to ensure the effectiveness, safety, and security of Edge AI systems. These core principles provide the basis for developing, deploying, and maintaining resilient and secure AI solutions. The principles are aligned with best practices in security, robustness, and resiliency.

- **Robustness** in Edge AI refers to the system's ability to function correctly and reliably, even when faced with unexpected situations, variations in input data, or perturbations in the environment. The Knowledge Hub will provide guidelines for building AI systems that:
 - **Perform consistently** despite changes in operating conditions, such as fluctuating data quality, environmental factors, or hardware malfunctions.
 - **Withstand adversarial inputs**, ensuring that AI models can maintain their decision-making integrity even when subjected to attacks like data poisoning or adversarial perturbations.
 - **Handle noisy and incomplete data**, which is often encountered in real-world applications, particularly in remote, edge environments where data transmission might be unreliable.

Developing robust AI systems requires extensive testing under stress scenarios, along with the adoption of methods like adversarial training and redundancy to minimize the impact of any failures.

- **Security** is a critical aspect of AI system development, particularly when AI is deployed in sectors like autonomous vehicles, and critical infrastructure. Secure Edge AI systems protect against both external and internal threats and safeguard sensitive data. The Knowledge Hub will provide resources to ensure AI systems are designed and operated with security at the forefront, covering:
 - **Data security**, including encryption, secure data storage, and secure transmission of information to ensure that sensitive data is protected from unauthorized access or tampering.
 - **Model security**, focusing on protecting AI models from adversarial attacks that aim to manipulate or mislead the model's decision-making process. This includes the use of robust ML techniques, secure model training, and real-time monitoring to detect potential vulnerabilities.
 - **Infrastructure security**, ensuring that the edge devices and networks supporting AI systems are resilient to cyberattacks, with mechanisms in place to prevent unauthorized access or malicious activities.

The Knowledge Hub will include security guidelines aligned with standards, helping stakeholders implement industry best practices for securing AI systems at all stages, from development to deployment.

- **Resiliency** refers to the ability of an Edge AI system to recover quickly from disruptions or failures, ensuring continuous and reliable operation in dynamic, high-risk environments. AI systems in safety-critical domains must be able to:
 - **Recover from failures** swiftly without compromising safety or service. This may involve backup mechanisms, failover strategies, and redundant systems that allow for quick restoration of functionality.
 - **Adapt to changing conditions** in real time, such as variations in data quality, environmental disturbances, or shifting operational parameters. Resilient systems can adjust their behavior or decision-making in response to these changes without affecting their core functionality.
 - **Maintain operational integrity** in the face of potential attacks, system failures, or unexpected environmental changes, ensuring that the AI system continues to perform safely and as expected, even when individual components experience issues.

To ensure resiliency, the Knowledge Hub will include strategies for continuous monitoring, stress testing, and self-healing mechanisms that can detect anomalies and automatically adapt to ensure system stability. The content is organized into two main pillars: Research Publications and Technical Reports and Guidelines.

4.2. Research Publications

This subsection provides access to peer-reviewed publications and project-supported research that explores key issues in Edge AI security and robustness. The Research Publications section is shown in Figure 6, while in Figure 7, Figure 9, Figure 8, and Figure 9, the three sub-categories are shown. To improve navigation and relevance, publications are grouped into three thematic categories:

- **Surveys**
Comprehensive review of current methodologies and challenges in AI robustness, Edge deployment, and Adversarial ML.
- **Attacks and Defenses**
Technical papers addressing known and emerging threats in AI systems alongside defense strategies such as adversarial training.
- **Ethics**
Research exploring ethical dimensions of AI development and deployment, including fairness, transparency, accountability, and human oversight – particularly in high-risk or safety-critical contexts.

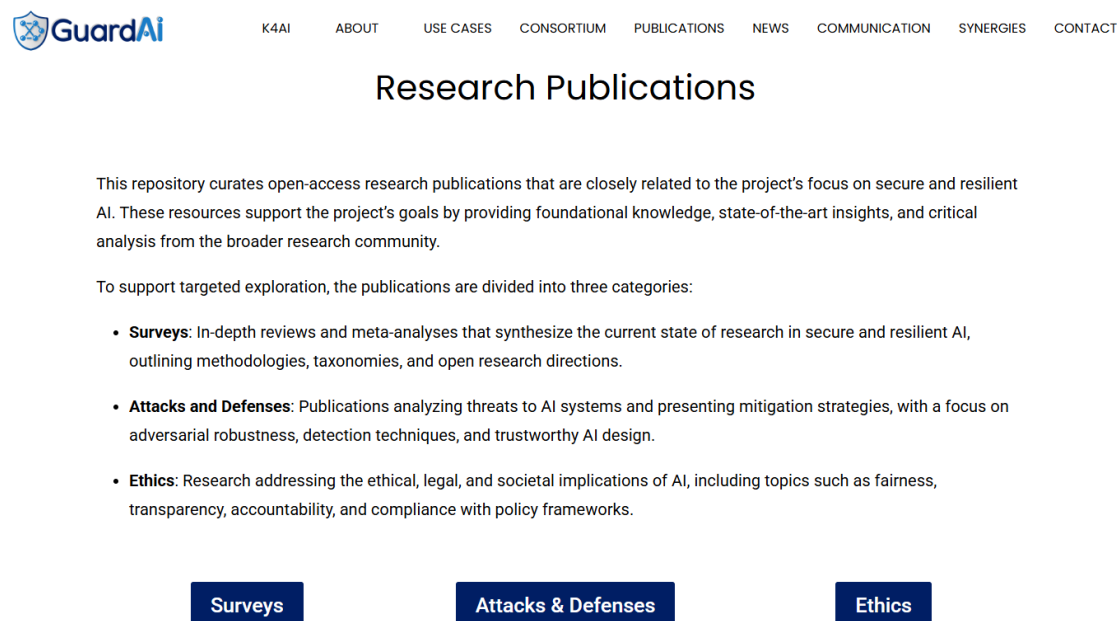


Figure 6: Research Publications Sub-section.

Research Publications: Surveys

This page features comprehensive surveys and review articles that map the evolving landscape of secure and resilient AI. These resources offer valuable overviews of existing methodologies, systematized taxonomies, and identified research gaps to guide further investigation and development.

Title	Publisher	Year	Description	Link
An Introduction to Adversarially Robust Deep Learning	IEEE Transactions on Pattern Analysis and Machine Intelligence	2024	This work presents a comprehensive survey of adversarial machine learning since 2013, covering key attack and defense strategies, taxonomies, and theoretical insights into adversarial robustness, fragility, and certification.	IEEE Explore
The Impact of Adversarial Attacks on Federated Learning: A Survey	IEEE Transactions on Pattern Analysis and Machine Intelligence	2024	This paper presents a hybrid deep learning framework that combines convolutional neural networks (CNNs) and transformers to enhance the accuracy and robustness of medical image segmentation,	IEEE Explore

Figure 7: Research Publications - Surveys Category.

Research Publications: Attacks & Defenses

This collection highlights research on adversarial threats and corresponding defense mechanisms in AI systems. It covers theoretical and applied work on attack vectors, vulnerability assessments, and approaches to building robust, resilient, and trustworthy AI solutions.

Adversarial Attacks			Adversarial Defenses	
Title	Publisher	Year	Description	Link
Hardening Interpretable Deep Learning Systems: Investigating Adversarial Threats and Defenses	IEEE Transactions on Dependable and Secure Computing	2024	This paper surveys adversarial attacks and defenses in machine learning-powered networks, categorizing attack methods and defense strategies, and highlighting challenges in balancing robustness, performance, and transferability.	IEEE Explore
Quantization Aware Attack: Enhancing	IEEE Transactions on Information	2024	This paper presents a comprehensive survey of adversarial machine learning, focusing on attacks and defenses across	IEEE Explore

Figure 8: Research Publications - Attacks and Defenses Category.

Research Publications: Ethics

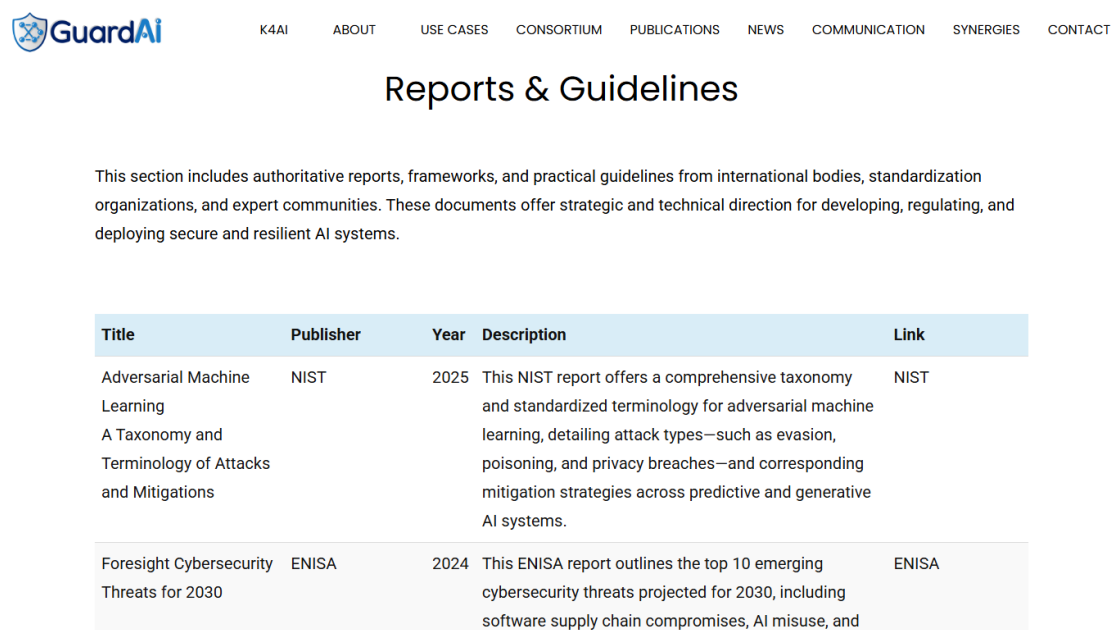
This section compiles research publications focused on the ethical, legal, and societal aspects of AI. Topics include algorithmic fairness, data privacy, transparency, and the responsible deployment of AI systems in compliance with emerging policy and regulatory frameworks.

Title	Year	Description	Link
COMPL-AI Framework: A Technical Interpretation and LLM Benchmarking Suite for the EU Artificial Intelligence Act	2025	This paper introduces COMPL-AI, a framework that translates the EU AI Act into measurable technical requirements for large language models and provides an open-source benchmarking suite to assess model compliance in areas like robustness, safety, and fairness.	arXiv
Towards a Privacy and Security-Aware Framework for Ethical AI: Guiding the Development and Assessment of AI Systems	2024	This paper presents a unified survey of adversarial machine learning attacks and defenses, emphasizing trends, challenges, and practical implications for secure AI deployment.	ACM Library

Figure 9: Research Publications - Ethics Category.

4.3. Technical Reports and Guidelines

This subsection includes high-impact materials such as whitepapers and domain specific guidelines. These documents offer actionable insights into secure AI system design, testing procedures, and resilient evaluation strategies. The reports cover a wide range of relevant themes, including the classification and mitigation of adversarial machine learning attacks, foresight into emerging AI-related cybersecurity threats, sector-specific security frameworks, and regulatory perspectives on the safe integration of AI into domains such as transportation and autonomous systems. Several documents focus on standardizing terminology and offering structured taxonomies to support a common understanding of AI threats and resilience strategies across sectors. Reports and Guidelines section is presented in Figure 10.



Title	Publisher	Year	Description	Link
Adversarial Machine Learning A Taxonomy and Terminology of Attacks and Mitigations	NIST	2025	This NIST report offers a comprehensive taxonomy and standardized terminology for adversarial machine learning, detailing attack types—such as evasion, poisoning, and privacy breaches—and corresponding mitigation strategies across predictive and generative AI systems.	NIST
Foresight Cybersecurity Threats for 2030	ENISA	2024	This ENISA report outlines the top 10 emerging cybersecurity threats projected for 2030, including software supply chain compromises, AI misuse, and	ENISA

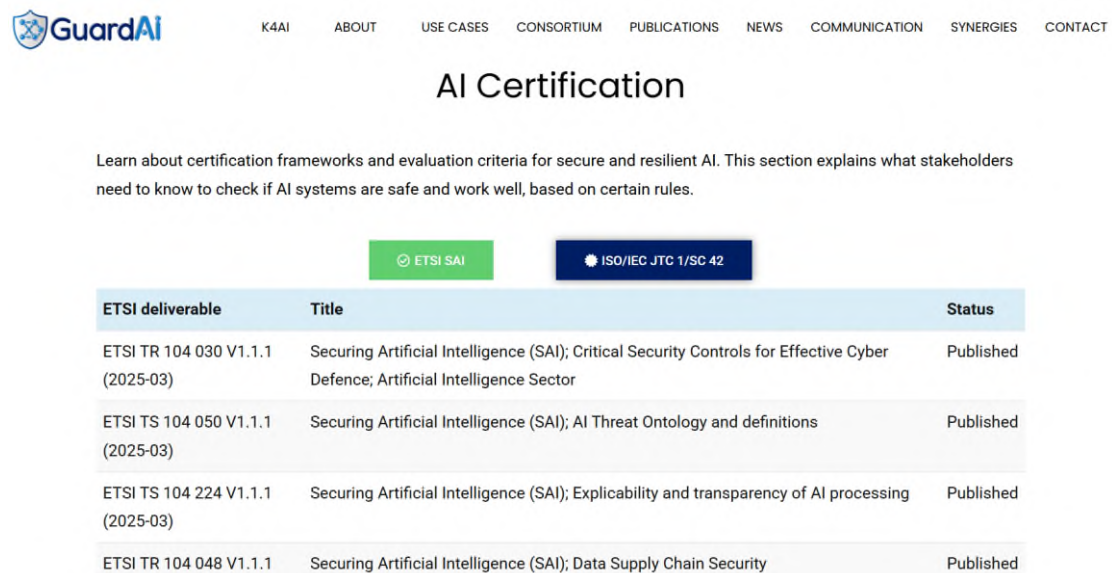
Figure 10: Reports and Guidelines Sub-section.

5. Certification and Compliance

To comply with the AI Act, manufacturers or providers of high-risk AI systems must comply with the harmonized European standards drafted exclusively by the European Technical Standardization Committee (CEN-CLC/JTC 21 “AI”). This committee is formally mandated by the European Commission to translate the AI Act’s essential requirements into technically precise standards and testable clauses. These standards, once cited in the Official Journal of the European Union (OJEU), provide a legal presumption of conformity for high-risk AI systems and thus play a central role in ensuring structured, consistent compliance across the EU.

By implementing the harmonized standards developed by CEN-CLC/JTC 21, organizations not only demonstrate compliance with the AI Act but also benefit from a well-defined, technically aligned route to certification. These standards cover areas related to risk management, data quality and governance, technical documentation, record keeping, transparency and provision of information to users, human oversight, accuracy, robustness and cybersecurity and secure lifecycle operations. For providers operating in critical or regulated domains, adopting these standards as part of their design and governance processes builds both legal defensibility and user trust. Thus, harmonized standards function not only as compliance instruments, but as strategic enablers for the safe, ethical, and lawful deployment of AI technologies across Europe.

The K4AI Portal contributes to the evolving certification landscape by offering structured, accessible resources that help stakeholders align their AI development processes with established and emerging standards. This section focuses on providing structured guidance for aligning Edge AI systems with existing and emerging certification frameworks.



ETSI deliverable	Title	Status
ETSI TR 104 030 V1.1.1 (2025-03)	Securing Artificial Intelligence (SAI); Critical Security Controls for Effective Cyber Defence; Artificial Intelligence Sector	Published
ETSI TS 104 050 V1.1.1 (2025-03)	Securing Artificial Intelligence (SAI); AI Threat Ontology and definitions	Published
ETSI TS 104 224 V1.1.1 (2025-03)	Securing Artificial Intelligence (SAI); Explicability and transparency of AI processing	Published
ETSI TR 104 048 V1.1.1	Securing Artificial Intelligence (SAI); Data Supply Chain Security	Published

Figure 11: Certification and Compliance Section.

6. Best Practices

The Best Practices section offers a curated collection of resources aimed at enhancing the security, robustness, and resilience of Edge AI systems in safety-critical applications. The portal provides educational resources designed to build awareness, promote secure AI development habits, and introduce key concepts related to AI robustness and adversarial resilience. Rather than prescriptive technical guidelines, this section focuses on knowledge transfer through accessible, practical examples and expert-led tutorials.

Key components of this section include:

- **Webinar and Presentations**
Access to sessions covering topics such as Adversarial AI, machine learning vulnerabilities, and strategies for ensuring AI robustness and reliability.
- **Case Studies and Examples**
Detailed analysis of specific scenarios, including demonstrations like the “one-pixel attack”, illustrating how minimal perturbations can significantly impact neural network predictions.
- **Educational Resources**
Courses and materials focus on deep neural network robustness, offering practical knowledge on handling realistic perturbations and enhancing system resilience.

The central feature of the section as mentioned above includes webinars, demonstration videos and presentations suggested by Consortium Members and collaborators. These materials focus on foundational topics in Edge AI, such as:

- Understanding vulnerabilities in machine learning systems.
- Introduction to threat modelling and risk assessment in AI pipelines.
- Architectural considerations for robustness in Edge AI environments.
- Visualizing and mitigating adversarial examples.

These resources aim to bridge the gap between theory and application by demonstrating how attacks work in practice, how system vulnerabilities can be exploited, and what measures can be taken to strengthen AI behavior in critical scenarios. As the GuardAI project progresses, the Best Practices section will serve as an entry point into the challenges of robust AI development and a growing repository of trusted learning content. The Best Practices section is shown in Figure 12.

Best Practices

Discover best practices, industry insights, and sector-specific approaches to AI safety and security. These guidelines support practitioners in adapting AI security measures to the needs of different application domains.

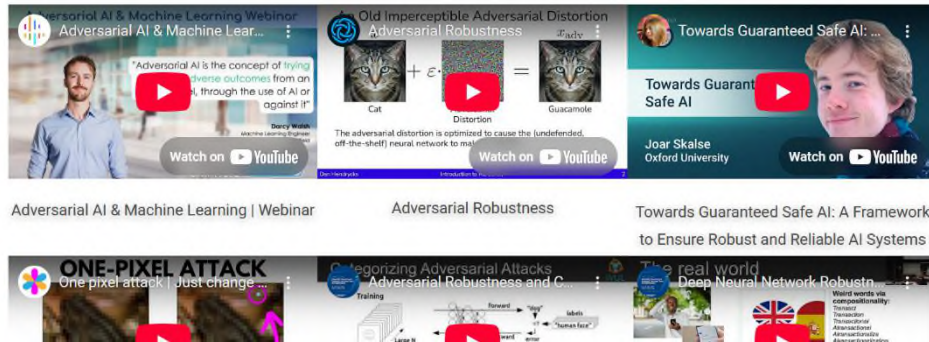


Figure 12: Best Practices Section.

7. Conclusion

The K4AI Portal is a foundational outcome of the GuardAI project, offering a structured, practical, and openly accessible platform to support the development, evaluation, and certification of secure Edge AI systems. By unifying essential tools, data and guidance, the portal empowers stakeholders to overcome current challenges in secure AI development and certification. It serves as a trusted resource hub for stakeholders seeking to implement and validate trustworthy AI solutions in real-world, safety-critical environments. Through expert contributions, shared resources, and links to ongoing webinars, and trainings, K4AI fosters a growing ecosystem towards common goals in AI trustworthiness and certification.

The portal will continue to evolve throughout the duration of the project, integrating new results and refining content based on feedback received from the users and based on stakeholder needs. The K4AI Portal is positioned to remain a key reference point for future efforts in certifying and deploying trustworthy Edge AI systems in high-risk environments.