TRUSTWORTHY ARTIFICIAL INTELLIGENCE

MAIN TECHNICAL COMMITTEES ON QUANTUM TECHNOLOGY STANDARDIZATION

- International level –

ISO/IEC JTC 1/SC 42 – Artificial Intelligence

Scope

Standardization in the area of Artificial Intelligence

- Serve as the focus and proponent for JTC 1's standardization program on Artificial Intelligence
- Provide guidance to JTC 1, IEC, and ISO committees developing Artificial Intelligence applications

- European level –

CEN/CLC/JTC 21 – Artificial Intelligence

Scope

The JTC shall produce standardization deliverables in the field of Artificial Intelligence (AI) and related use of data, as well as provide guidance to other technical committees concerned with Artificial Intelligence.

The JTC shall also consider the adoption of relevant international standards and standards from other relevant organisations, like ISO/IEC JTC 1 and its subcommittees, such as SC 42 Artificial intelligence.

The JTC shall produce standardization deliverables to address European market and societal needs and to underpin primarily EU legislation, policies, principles, and values.

ETSI/TC SAI – Securing Artificial Intelligence

Scope

The aim of Technical Committee Securing Artificial Intelligence (TC SAI) is to develop technical specifications that mitigate against threats arising from the deployment of AI, and threats to AI systems, from both other AIs, and from conventional sources. Whilst in the short to medium term the focus of TC SAI will be on the application of Machine Learning (ML) the group shall also give guidance and evaluation reports to ETSI and its stakeholders on the wider developments of AI.

TC SAI addresses 4 main aspects of AI security standardisation:

1. Securing AI from attack e.g. where AI is a component in the system that needs defending.
2. Mitigating against AI e.g. where AI is the ‘problem’ (or used to improve and enhance other more conventional attack vectors),
3. Using AI to enhance security measures against attack from other things e.g. AI is part of the ‘solution’ (or used to improve and enhance more conventional countermeasures),
4. Societal security and safety aspects of the use and application of AI.

RELEVANT PUBLISHED STANDARDS ON TRUSTWORTHY AI

ISO/IEC JTC 1/SC 42

ISO/IEC TR 24028:2020	Artificial intelligence — Overview of trustworthiness in artificial intelligence
ISO/IEC TR 24027:2021	Artificial intelligence (AI) — Bias in AI systems and AI aided decision making (adopted as CEN/CLC ISO/IEC/TR 24027:2023)
ISO/IEC TR 24029-1:2021	Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview (adopted as CEN/CLC ISO/IEC/TR 24029-1:2023)
ISO/IEC 24029-2:2023	Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods
ISO/IEC 23894:2023	Artificial intelligence — Guidance on risk management (adopted as EN ISO/IEC 23894:2024)
ISO/IEC 25059:2023	Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems
ISO/IEC TR 5469:2024	Artificial intelligence — Functional safety and AI systems
ISO/IEC TS 8200:2024	Artificial intelligence — Controllability of automated artificial intelligence systems

ETSI/TC SAI

ETSI TR 104 067 V1.1.1 (2024-04)	Securing Artificial Intelligence (SAI); Proofs of Concepts Framework
ETSI TR 104 225 V1.1.1 (2024-04)	Securing Artificial Intelligence TC (SAI); Privacy aspects of AI/ML systems
ETSI TR 104 032 V1.1.1 (2024-02)	Securing Artificial Intelligence (SAI); Traceability of AI Models
ETSI TR 104 031 V1.1.1 (2024-02)	Securing Artificial Intelligence (SAI); Collaborative Artificial Intelligence

RELEVANT ONGOING PROJECTS ON TRUSTWORTHY AI

ISO/IEC JTC 1/SC 42

ISO/IEC CD TS 6254	Artificial intelligence — Objectives and approaches for explainability of ML models and AI systems
ISO/IEC TS 12791	Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks
ISO/IEC DIS 12792	Information technology — Artificial intelligence — Transparency taxonomy of AI systems
ISO/IEC AWI TS 17847	Artificial intelligence — Verification and validation analysis of AI systems
ISO/IEC AWI TS 22440	Artificial intelligence – Functional safety and AI systems — Requirements
ISO/IEC AWI TS 29119-11	Software and systems engineering — Software testing — Part 11: Testing of AI systems

CEN/CLC/JTC 21

prCEN/CLC/TR	AI Risks - Check List for AI Risks Management
prEN	AI-enhanced nudging
prEN	Artificial Intelligence trustworthiness framework

ETSI/TC SAI

DTR/SAI-009 (TR 104 222)	SAI Mitigation Strategy report
DTS/SAI-006 (TS 104 033)	AI Computing Platform Security Framework
DTR/SAI-0015 (TR 104 066)	Securing Artificial Intelligence TC (SAI); Security Testing of AI
DTR/SAI-001 (TR 104 029)	Global AI Security Ecosystem
DTR/SAI-002 (TR 104 030)	AI Critical Security Controls
DTS/SAI-006 (TS 104 033)	AI Computing Platform Security Framework