TRUSTWORTHY ARTIFICIAL INTELLIGENCE

"Trustworthy AI has **three components**, which should be met throughout the system's entire life cycle: (1) it should be **lawful**, complying with all applicable laws and regulations (2) it should be **ethical**, ensuring adherence to ethical principles and values and (3) it should be **robust**, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm."

High-Level Expert Group on AI "Ethics Guidelines for Trustworthy AI"

"Characteristics of trustworthiness include, for instance, reliability, availability, resilience, security, privacy, safety, accountability, transparency, integrity, authenticity, quality, usability."

ISO/IEC TR 24028:2020 Information technology -- Artificial intelligence -- Overview of trustworthiness in artificial intelligence

MAIN TECHNICAL COMMITTEES ON QUANTUM TECHNOLOGY STANDARDIZATION

- International level -

➢ ISO/IEC JTC 1/SC 42 – Artificial Intelligence

Standards	18	Projects	29	National delegates	22

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

7 Working Groups

JWG 2	Joint Working Group ISO/IEC JTC1/SC 42 - ISO/IEC JTC1/SC 7 : Testing of AI-based systems	WG 2	Data
JWG 3	Joint Working Group ISO/IEC JTC1/SC42 - ISO/TC 215 WG : AI enabled health informatics	WG 3	Trustworthiness
WG 1	Foundational standards	WG 4	Use cases and applications
		WG 5	Computational approaches and computational characteristics of AI systems

- European level -

➢ CEN/CLC/JTC 21 – Artificial Intelligence

Standards	2	Projects	15	National delegates	12

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.

4 Working Groups

WG 1	Strategic Advisory Group (SAG)	WG 3	Engineering aspects
WG 2	Operational aspects	WG 4	Foundational and societal aspects

MAIN PUBLISHED STANDARDS ON TRUSTWORTHY AI

ISO/IEC JTC 1/SC 42	
ISO/IEC TR 24028:2020	Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence
ISO/IEC 23894:2023	Information technology — Artificial intelligence — Guidance on risk management
ISO/IEC TR 24027:2021	Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making
ISO/IEC TR 24029-1:2021	Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview
ISO/IEC 25059:2023	Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

MAIN ONGOING PROJECTS ON TRUSTWORTHY AI

ISO/IEC JTC 1/SC 42	
ISO/IEC CD TR 5469	Artificial intelligence — Functional safety and AI systems
ISO/IEC CD TS 6254	Information technology — Artificial intelligence — Objectives and approaches for explainability of ML models and AI systems
ISO/IEC CD TS 8200	Information technology — Artificial intelligence — Controllability of automated artificial intelligence systems
ISO/IEC CD TS 12791	Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks
ISO/IEC CD 12792	Information technology — Artificial intelligence — Transparency taxonomy of AI systems
ISO/IEC AWI TS 17847	Information technology — Artificial intelligence — Verification and validation analysis of AI systems
ISO/IEC AWI TR 21221	Information technology – Artificial intelligence – Beneficial AI systems
ISO/IEC 24029-2	Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods
ISO/IEC DTS 25058	Software and systems engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of AI systems
ISO/IEC CD TS 29119-11	Software and systems engineering — Software testing — Part 11: Testing of AI systems
ISO/IEC AWI TR 42106	Information technology — Artificial intelligence — Overview of differentiated benchmarking of AI system quality characteristics
CEN/CLC/JTC 21	
prCEN/CLC/TR	Artificial Intelligence - Overview of AI tasks and functionalities related to natural language processing
prCEN/TR	AI Risks - Check List for AI Risks Management
prEN	Accuracy of natural language processing systems
prEN	Artificial Intelligence trustworthiness characterisation
prEN	Competence Requirements for AI ethicists professionals
prEN	AI-enhanced nudging