Workshop “AI Security Standards: A Step Forward in the Evaluation and Management of AI”

A view from ETSI TC SAI

Presented by: Scott CADZOW, ETSI TC SAI Chair

8th May 2025

The genAI proposal

Pose a dilemma and work from there.

The dilemma: “Imagine a world where AI systems make life-changing decisions — from approving loans to diagnosing illnesses — but without any clear rules on how secure or fair they are”

The comeback: “Can we truly trust AI without robust security standards?”

The proposition:

AI Security Standards = Guidelines and protocols to ensure AI systems are protected against misuse, manipulation, and failure.

Evaluation = How we test and verify AI system performance and resilience.

Management = How we govern, monitor, and update AI systems over time.

To close:

Security standards aren’t a limitation — they’re a launchpad for innovation. The future of trustworthy AI starts with how we evaluate and manage it today.

Timeline of TC SAI

- Kick off meeting of TC SAI#01 was December 3rd and 4th 2023 (18+ months ago)
- Adopted the work programme of ISG SAI and added new work items too
- Publications so far (18 months of activity): 16
- Active work items as of now (May 2025): 6 -- more expected to be proposed in the June meeting

SAI publications in period to January 2025

ETSI deliverable	title
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 104 031 V1.1.1 (2024-02)	Securing Artificial Intelligence (SAI); Collaborative Artificial Intelligence
ETSI TR 104 032 V1.1.1 (2024-02)	Securing Artificial Intelligence (SAI); Traceability of AI Models
ETSI TR 104 048 V1.1.1 (2025-01)	Securing Artificial Intelligence (SAI); Data Supply Chain Security
ETSI TR 104 222 V1.2.1 (2024-07)	Securing Artificial Intelligence; Mitigation Strategy Report
ETSI TR 104 062 V1.2.1 (2024-07)	Securing Artificial Intelligence; Automated Manipulation of Multimedia Identity Representations
ETSI TR 104 066 V1.1.1 (2024-07)	Securing Artificial Intelligence; Security Testing of AI
ETSI TR 102 221 v1.1.1 (2025-01)	Securing Artificial Intelligence; Problem statement

SAI most recent publications

ETSI deliverable	title	Scope
ETSI TR 104 051	Security aspects of using AI/ML techniques in telecom sector	The use of AI to facilitate the use cases may cause AI security and privacy issues specific to the telecom industry. The scope of this proposed work item will be to investigate security and privacy issues related to the use of AI in the telecom industry sector.
ETSI TR 104 065	AI Act mapping and gap analysis	An analysis of the standardisation requirements of the AI Act against the workplan of ETSI (across all TBs) in order to identify gaps and the means to fill them
ETSI TS 104 223	Cyber Security Requirements for AI	based on public consultation on a Code of Practice in the UK, and will establish baseline cyber security requirements for AI models and systems that enable them to embed cyber security and resilience across the AI lifecycle.
ETSI TS 104 224	Explicability and transparency of AI processing	Address the issues of design of AI platforms (data, algorithms, frameworks) that are able to give support to claims of explainability and transparency of decisions
ETSI TR 104 029	Global AI Security Ecosystem	Structured enumeration and description of organisations and activities globally relevant to AI security
ETSI TR 104 030	AI Critical Security Controls	Applies the latest version of the Critical Security Controls and facilitation mechanisms for effective risk control and enhanced resilience of AI sector products and services
ETSI TS 104 050	AI Ontology and definitions	Defines a common terminology for AI (aligned to CEN/ISO). Defines what is meant by these terms in the context of cyber and physical security and with an accompanying narrative that should be readily accessible by both experts and less informed audiences across multiple industries.
ETSI TR 102 123	Implementation guide for organizations implementing baseline cyber security requirements for AI	Guidance to help stakeholders in the AI supply chain in meeting the cyber security provisions defined for AI models and systems in ETSI TS 104 223

SAI active work items

ETSI deliverable	title	Scope
ETSI TR 104 029	Global AI Security Ecosystem	Structured enumeration and description of organisations and activities globally relevant to AI security
ETSI TS 104 033	AI Computing Platform Security Framework	Defines the essential requirements for the execution environment and related resources for supporting/hosting AI applications in an AI system,
ETSI TR … …	Understanding and Preventing Harm from Generative AI	Intended to provide an understanding of the harm from Generative AI along with presenting the different ways to prevent that harm. This includes but is not limited to malicious code generation, deepfakes, spam messages, disinformation etc. The areas also covered are the issues of AI hallucinations, loss of confidentially and IPR infringements. The types of methods to counter the harm from Generative AI to be included are detection, enforcement, reporting and removal
ETSI TS 104 158	AI Common Incident Expression (AICIE)	AI Common Incident Expression together with a Common Incident Scoring System that is compatible with lightweight information exchange standards. NOTE: This addresses the many requirements for common vulnerability reporting and is aligned with work in ETSI CYBER and with global initiatives in this domain
ETSI TS 104 225 **PROPOSED**	Privacy aspects of AI/ML systems	The intent of this work is to take the text of (the previously published) TR 104 225 and to extend it to make normative provisions addressing privacy and data protection. This will address aspects including "the right to be forgotten" and how to ensure if data is removed from an AI system that it is not recoverable.
New TR **PROPOSED**	Application of Functional Safety consideration in AI systems development	The intent of this work item is to identify the actions of AI developers to ensure that their products and systems remain functionally safe when deployed. This work item will report on the core FS measures and how they work alongside the core principles of SAI

Some points to note that impact TC SAI's platform

ETSI has **not** been invited to address the AI Act Standardisation Request

Impacts the content of ETSI TR 104 065 to some extent -- ETSI's output is still valid to the AI Act

AI Act will seek to ensure AI acts in support of, and protect, human rights

- This works today for Narrow Intelligence but breaks down when the AI users do not support the same human rights

AI Act will seek to ensure AI acts ethically

- Ethics cannot be codified to give simple binary right/wrong answers

The AI Standardisation actors are explicitly recognised in TR 104 029

- The big SDOs: CEN, ETSI, ITU-T, IEEE, *IETF*. The big actors: Meta, X, Google/Alphabet, Tesla, Microsoft, Apple. The governments: EU, Nation States, UN … .
- TR 104 029 is "published" at each meeting of SAI showing the latest ecosystem and actors in AI

ETSI and AI and Security

ETSI has many groups addressing each of AI and cybersecurity, with SAI bridging between the domains

- TC CYBER - baseline cybersecurity addressing horizontals (controls, primitives, methods)
- TC ESI - signature and trust frameworks, governance, eIDAS, Digital Wallet, Smart Contracts and so on
- ETSI TC DATA - incorporating former TBs SmartM2M (semantics of data and how semantic knowledge is exchanged), CDM (marine data for ship/harbour movements), CIM (context aware data (NGSI-LD))
- ISG ENI - improving performance of networks from the edge
- ISG ZSM - Zero touch management of networks
- MTS - working out how to test AI and AI solutions

Nothing is really new in AI

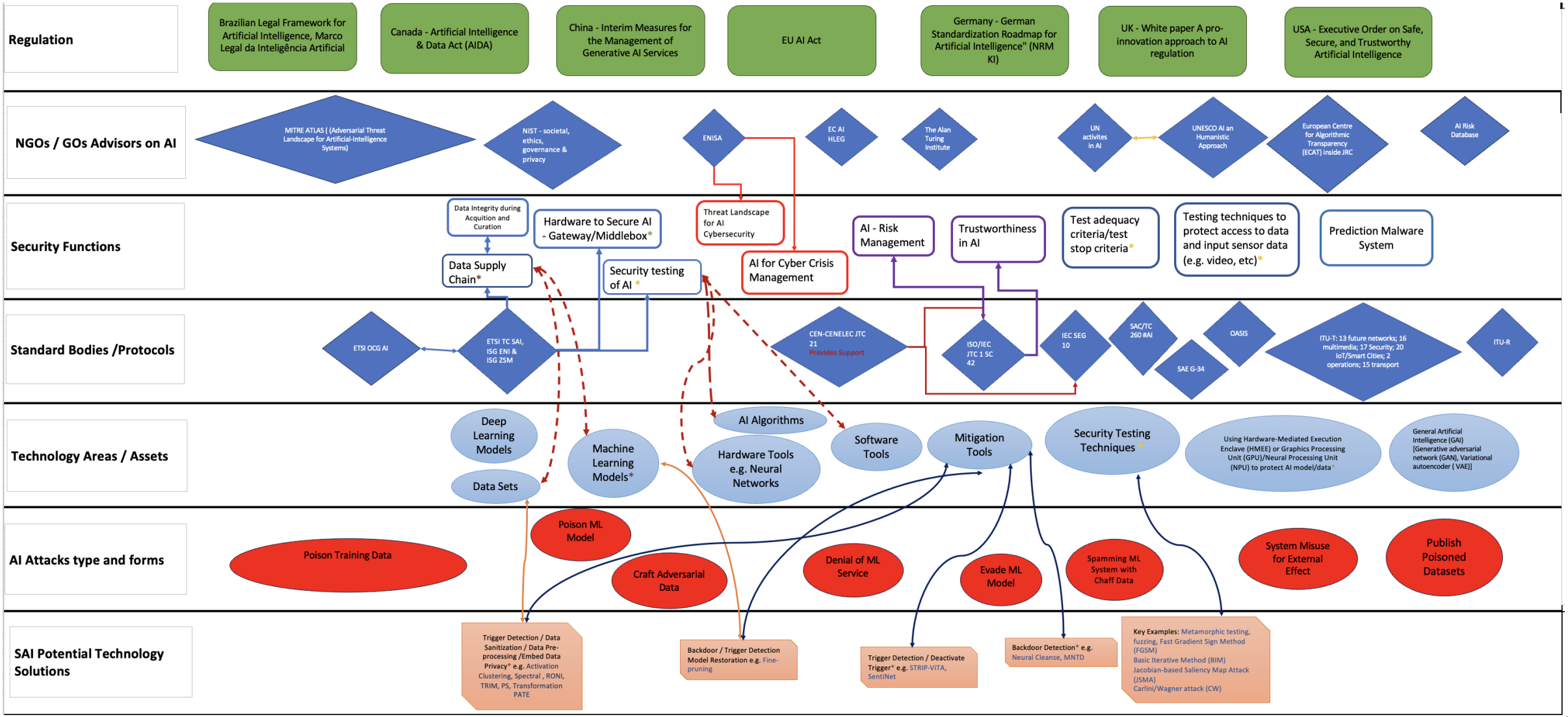
In ML mode most of the approaches are based on statistical analysis

- Identify correlations from data - learn the correlations, expand to identify supporting correlations, suggest causations
- Recognise patterns - expand the recognition base by reinforcement learning
- Generative tools - from a ruleset, a data source, and a prompt, identify reasonable text to answer the prompt. Uses semantic analysis of the request to find data sources and build from those, using the ruleset to generate new text (that gets added to the data source)

The hype is about General AI (AGI) whereas most AI is very narrow AI (ANI)

- Recommender systems could fall into AGI or ANI and into any of the risk classes - this is why the type of AI system is insufficient to determine what the AI is

TC SAI - a view on our world



TC SAI – capturing a global vision for Europe

ETSI's new global standard (TS 104 223) and accompanying implementation guide (TR 104 128) is based on activity led by the UK to conduct a global consultation on how to develop secure AI systems and to translate the results of the consultation into a globally recognised code of practice.

In addition TS 104 223, alongside other work in ETSI has been mapped to the EU's AI Act in TR 104 065 and maps to most of the requirements found there.

https://www.gov.uk/government/publications/ai-cyber-security-code-of-practice

https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf

Mitigating the AI security problem

Knowledge, proof, verification are at the heart of making AI secure

- Knowing what the data should be and where it comes from -- Supply chain integrity and provenance
- Knowing what the algorithm should be doing and is doing it

The core conventions of the CIA paradigm still apply → Reinforced by wider adoption of the zero-trust model answering questions of What — Why — When — How — Where — Who

- Authenticate and verify authorisation → Least persistence and least privilege models
- Verifying integrity is not going to be as straightforward as the data is always changing

Notably AI will need to police itself - without interfering with the core functions

- This is not new - we already use software to protect software but it’s a new level of protection we’re seeking

Some of the problems we're trying to solve or resolve

… to define what would be considered an AI threat and how it might differ from threats to traditional systems.

There was no **common** understanding of what constitutes an attack on AI and how it might be created, hosted and propagated when ETSI started in ISG SAI.

- Lots of scare stories inspired by novels, TV, film (the Skynet or HAL-9000 scenario)

As a standards body we need to provide a narrative that should be readily accessible by both experts and less informed audiences across the multiple industries and stakeholders in AI to allow for understanding and debate around the problem.

- Debunking the fictional narratives and promoting a rational discussion

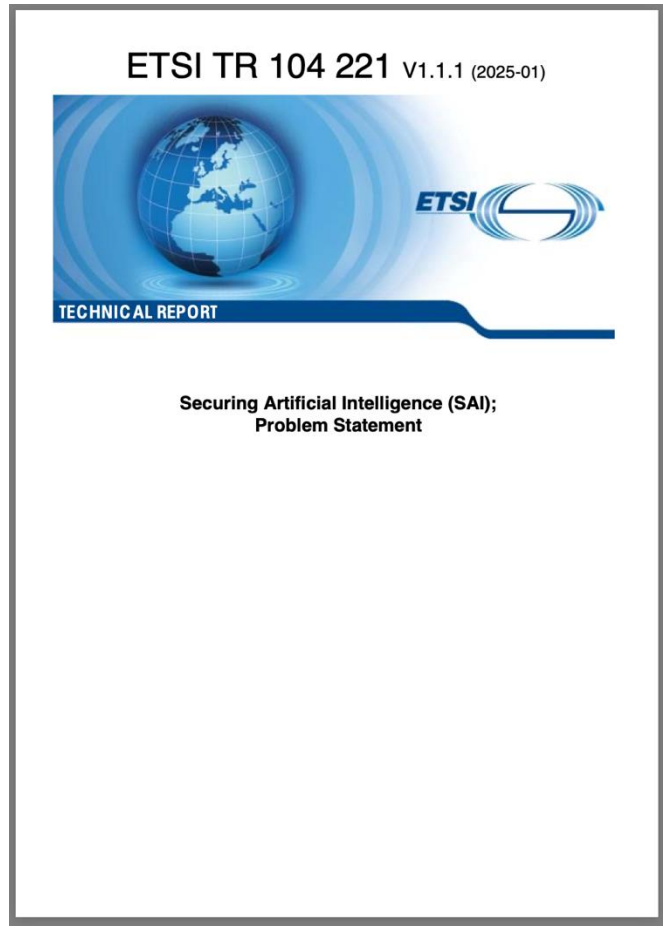
It is essential to address AI in as many forms as it can take: as a system in its own right (rare); as a component of a system; as an adversarial attacker; and, as a system defender

The problem statement - a precis

ETSI TR 104 221 “Problem Statement”

Describes the problem of securing AI-based systems and solutions, with a focus on machine learning, and the challenges relating to confidentiality, integrity and availability at each stage of the machine learning lifecycle. It also describes some of the broader challenges of AI systems including bias, ethics and explainability. A number of different attack vectors are described, as well as several real-world use cases and attacks

Lifecycle Phase	Issues
Data Acquisition	Integrity
Data Curation	Integrity
Model Design	Generic issues only
Software Build	Generic issues only
Train	Confidentiality, Integrity, Availability
Test	Availability
Deployment	Confidentiality, Integrity, Availability
Upgrades	Integrity, Availability



The result of lots of experts coming together to agree on what we need to work on

In closing - some remarks

Without standards the key points from slide 2 will not be addressed, or addressed too late to matter.

AI Security Standards = Guidelines and protocols to ensure AI systems are protected against misuse, manipulation, and failure. (TS 103 224 for example)

Evaluation = How we test and verify AI system performance and resilience. (TR 104 066, TR 104 048 as examples)

Management = How we govern, monitor, and update AI systems over time. (AICIE for example)

ETSI is active in all of these areas to a greater or lesser degree.

Thanks for listening

Scott CADZOW, scott at cadzow dot com, somewhere in England