ETSI TR 104 128 V1.1.1 (2025-05)

Securing Artificial Intelligence (SAI);
Guide to Cyber Security for AI Models and Systems

Reference
DTR/SAI-0012

Keywords
artificial intelligence, security

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI IPR online database.

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™**, **LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM®** and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Securing Artificial Intelligence (SAI).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Introduction

The growing deployment and technological advancements of Artificial Intelligence (AI) has further reiterated the need for tailored security requirements for AI systems. The present document will guide stakeholders across the AI supply chain on implementations of ETSI TS 104 223 [i.1] by providing non-exhaustive scenarios as well as examples of practical solutions to meet these provisions.

1 Scope

The present document gives 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 [i.1]. These stakeholders could include a diverse range of entities, including large enterprises and government departments, independent developers, Small and Medium Enterprises (SMEs), charities, local authorities and other non-profit organizations. The present document will also be useful for stakeholders planning to purchase AI services. Additionally, the present document has been designed to support the future development of AI cyber security standards, including specifications that could inform future assurance and certification programmes. Where relevant, the present document signposts supporting specifications and international frameworks.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long-term validity.

The following referenced documents may be useful in implementing an ETSI deliverable or add to the reader's understanding, but are not required for conformance to the present document.

[i.1] ETSI TS 104 223: "Securing Artificial Intelligence (SAI); Baseline Cyber Security Requirements for AI Models and Systems".

[i.2] ETSI TR 104 222: "Securing Artificial Intelligence; Mitigation Strategy Report".

[i.3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[i.4] ISO/IEC 22989:2022: "Information technology — Artificial intelligence — Artificial intelligence concepts and terminology".

[i.5] NCSC: "Machine learning principles".

[i.6] NCSC: "Guidelines for Secure AI system development".

[i.7] CISA: "Joint Cybersecurity Information - Deploying AI Systems Securely".

[i.8] MITRE: "ATLAS Framework".

[i.9] NIST AI 100-2 E2023: "Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations".

[i.10] OWASP®: "AI Exchange".

[i.11] OWASP®: "Top 10 for LLM Applications".

[i.12] ETSI TR 104 032: "Securing Artificial Intelligence (SAI); Traceability of AI Models".

[i.13] ETSI TR 104 225: "Securing Artificial Intelligence TC (SAI); Privacy aspects of AI/ML systems".

[i.14] ETSI TR 104 066: "Securing Artificial Intelligence (SAI); Security Testing of AI".

[i.15] ISO/IEC 42001:2023: "Information technology — Artificial intelligence — Management system".

[i.16] ISO/IEC 25059:2023: "Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI system)".

[i.17] NIST: "AI Risk Management Framework".

[i.18] International Scientific Report on the Safety of Advanced AI: "Interim Report".

[i.19] ICO: "Data Protection Audit Framework".

[i.20] ICO: "Generative AI: eight questions that developers and users need to ask".

[i.21] OWASP: "LLM Applications Cybersecurity and Governance Checklist".

[i.22] ICO: "How should we assess security and data minimisation in AI?".

[i.23] AI Village Defcon 2024 Report: "Generative AI Red Teaming Challenge: Transparency Report".

[i.24] OWASP®: "Threat Modeling Cheat Sheet".

[i.25] ICO: "Newsletters".

[i.26] NCSC: "News on AI".

[i.27] MITRE: "Slack Channel".

[i.28] OWASP®: "Slack Invite".

[i.29] NIST: "NIST Secure Software Development Framework for Generative AI and for Dual Use Foundation Models Virtual Workshop".

[i.30] OWASP®: "OWASP Secure Coding Practices-Quick Reference Guide".

[i.31] NCSC: "Secure Development and Deployment Guidance".

[i.32] ICO: "Do we need to consult the ICO?".

[i.33] NCSC: "Risk management".

[i.34] ICO: "What is the impact of Article 22 of the UK GDPR on fairness?".

[i.35] ICO: "AI and data protection risk toolkit".

[i.36] PCI® DSS.

[i.37] FCA Standards.

[i.38] EU Artificial Intelligence Act: "The AI Act Explorer".

[i.39] NCSC: "Risk management - Threat Modelling".

[i.40] NIST: "NIST AI RMF Playbook".

[i.41] NIST: "NIST AI RMF Crosswalk Documents".

[i.42] ICO: "What are the accountability and governance implications of AI?".

[i.43] European AI Alliance: "Implementing AI Governance: from Framework to Practice".

[i.44] OWASP®: "New OWASP AI Security Center of Excellence (CoE) Guide".

[i.45] MITRE ATT&CK®.

[i.46] CSA: "Guidelines and Companion Guide on Securing AI Systems".

[i.47] NCSC: "Secure Design Principles".