ETSI TR 104 029 V0.0.6 (2024-12)

Securing Artificial Intelligence (SAI_TC);
Global Ecosystem

Reference
DTR/SAI-001

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 Web server (https://ipr.etsi.org).

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™** and **LTE™** 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.

Executive summary

The emergence and expansion of organisations and activities related to AI security is expanding expansively and rapidly. This phenomenon parallels the development and distribution of AI code and product offerings by millions of providers among billions of users everywhere, all the time. The present document provides insight into the breadth and diversity of the organisations and activities globally.

Introduction

Although Artificial Intelligence activities and related standards work in bodies such as ETSI have existed for more than two decades, the recent significant scaling in technologies and implementations has resulted in a new continuously expanding AI ecosystem that overlays existing bodies and adds new ones. Most are relevant in part or in whole to AI security. Insularity among these bodies is common. In the spirit of spanning barriers and producing a holistic portrayal of this new ecosystem, the present Technical Report aims to discover and identify all these bodies and activities, including relationships among them.

1 Scope

The present document provides a structured enumeration and description of organisations and activities globally relevant to AI security

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 are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

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

[i.2] Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) No 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act) (Text with EEA relevance). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202402847

[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) (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[i.4] Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive). https://commission.europa.eu/document/f9ac0daf-baa3-4371-a760-810414ce4823_en

[i.5] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC (Text with EEA relevance). https://eur-lex.europa.eu/eli/dir/2024/2853/oj

[i.6] Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011 (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2022/2554/oj

[i.7] Council of Europe Treaty Series – No. 225, "Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law". https://rm.coe.int/1680afae3c

…

[i.10] "European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)),"

https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html

[i.20] UK DSIT, "AI Cyber Security Code of Practice."

NOTE: Available at: https://www.gov.uk/government/calls-for-evidence/call-for-views-on-the-cyber-security-of-ai/call-for-views-on-the-cyber-security-of-ai#ai-cyber-security-code-of-practice

[i.21] UK NCSC, US CISA + 21 national security agencies, Guidelines for secure AI system development, 1.0

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

[i.22] NIST, AI 600-1, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile".

NOTE: Available at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

[i.23] NIST, Cybersecurity White Paper (CSWP) 31, "Proxy Validation and Verification for Critical AI Systems".

NOTE: Available at: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.31.pdf

[i.24] UK Office for Artificial Intelligence: AI Standards Hub.

NOTE: Available at: https://aistandardshub.org/ai-standards-search/

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the following terms apply:

Artificial intelligence: machine learning (ML) applications that involve software components (models) that allow computers to recognise and bring context to patterns in data without the rules having to be explicitly programmed by a human generate predictions, recommendations, or decisions based on statistical reasoning. [i.21]

Artificial intelligence: ability of a system to handle representations, both explicit and implicit, and procedures to perform tasks that would be considered intelligent if performed by a human. [GR SAI 004]

Artificial intelligence: a set of technologies that enable computers to perform a variety of advanced functions, including the ability to see, understand and translate spoken and written language, analyze data, make recommendations, and more. [Google Cloud]

Artificial intelligence: the field of computer science dedicated to solving cognitive problems commonly associated with human intelligence, such as learning, creation, and image recognition. [Amazon Web Services]

Artificial intelligence: the capability of a computer system to mimic human-like cognitive functions such as learning and problem-solving. [Microsoft Azure]

Artificial intelligence: systems and machines that can imitate human intelligence to perform tasks and iteratively improve themselves based on the collected information. [Alibaba Cloud]

Artificial intelligence: a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments, and vary in their levels of autonomy and adaptiveness after deployment. [OECD]

Artificial intelligence: a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. [i.10]

Artificial intelligence: an engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives. [ISO/IEC 22989, ITU-T Y.supp 72]

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

AI	Artificial Intelligence
CTI	Cyber Threat Intelligence
ML	Machine Learning

4 Global AI security ecosystem

4.1 Organization of the ecosystem forums and activities

This clause organizes the global cyber security ecosystem as six groups of forums and activities that are fundamental collaborative mechanisms for cyber security and its evolution:

1) forums that develop techniques, technical standards and operational practices;
2) major IT developer forums affecting cyber security;
3) activities for continuous information exchange;
4) centres of excellence;
5) reference libraries, continuing conferences; and
6) heritage sites and historical collections.

In some cases, the same parent organization hosts multiple forums and activities that are attributed to different groups. In other cases, the organization hosts numerous forums where several of them have fully or substantially dedicated cyber security functions - which are indented under the parent. Because of the very large numbers of forums, and in the interests of providing a useful understanding of the ecosystem, only very short descriptions are provided, and the reader is encouraged to use the URI links to fully appreciate the work being done.

This compilation attempts to be an inclusive as possible to expand the collective insight into the extent of the ecosystem. Toward this objective, it includes collaborative mechanisms that are frequently overlooked but enormously significant in the cyber security arena such as developer forums for the major IT platforms, centres of excellence that are rapidly growing in numbers worldwide, and continuing conferences - even hacker major global hacker events that regularly reveal AI security vulnerabilities that were previously unknown.

This material is augmented by annex A which contains national AI security ecosystems that have been published in national strategy or other publicly available material. Annex B contains depictions of relationships among these ecosystems.

4.2 Fora that develop techniques, technical standards and operational practices

The forums listed below are well known venues engaging in significant global collaboration to produce techniques, technical standards and operational practices for cyber security. Where the venues operate only at a national level, they are placed in Annex A.

3GPP - 3rd Generation Partnership Project. 3GPP is the largest and most active mobile telecommunication and ICT standardisation organisation worldwide. The advent of AI has led to the development of multiple work items across nearly all of its dozens of subgroups relating to the use of AI at the air interface, for network management and media, and for application AI based services and applications in 5G, 6G, and beyond. Rel-18 AI work items include: FS_AIMLsys, AIML_MT, FS_AIML_MGMT, FS_AI4Media, FS_NR_AIML_air, NR_AIML_NGRAN, FS_AI4Media. Rel-19 AI work items include: FS_AIML_MT_Ph2, FS_AIML_CN, AIML_CN, and AIML_CN.

Alan Turing Institute. Alan Turing first considered the question, "Can machines think?" in his seminal paper, Computing Machinery and Intelligence, published in 1950. The Institute's activities today shape the foundations of AI, safe and ethical principles, and inform legal and regulatory frameworks in the UK and internationally. https://www.turing.ac.uk/research/research-programmes/artificial-intelligence

AI Alliance. Focused on accelerating and disseminating open innovation across the AI technology landscape to improve foundational capabilities, safety, security and trust in AI, and to responsibly maximize benefits to people and society. Over 50 Founding Members and Collaborators globally. https://thealliance.ai/

Artificial Intelligence Security Center (AISC). As part of the NSA Cybersecurity Collaboration Center (CCC), it is a focal point for developing best practices, evaluation methodology and risk frameworks with the aim of promoting the secure adoption of new AI capabilities across the national security enterprise and the defense industrial base. https://www.nsa.gov/AISC/ https://www.nsa.gov/About/Cybersecurity-Collaboration-Center/

AI Safety Summit Coalition. The Declaration and other measures discussed at the summit provide an opportunity to build coordinated international AI regulation. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023. See DSIT, below.

BEREC – Body of European Regulators for Electronic Communications. BEREC assists the European Commission (EC) and the National Regulatory Authorities (NRAs) in implementing the EU regulatory framework for electronic communications. It provides advice on request and on its own initiative to the European institutions and complements, at European level, the regulatory tasks performed by the NRAs at national level. https://www.berec.europa.eu/en/berec/mission-strategy BEREC completed a public consultation on the impact of AI solutions in the telecommunications sector on regulation. https://www.berec.europa.eu/en/document-categories/berec/reports/berec-report-on-the-impact-of-artificial-intelligence-ai-solutions-in-the-telecommunications-sector-on-regulation

CableLabs. Serves as the principal global R&D and standards body for the cable industry. It is developing standards for next-generation networks with generative artificial intelligence. CableLabs maintains an AI archives page. https://www.cablelabs.com/blog/tag/ai

CEN - Comité Européen de Normalisation. Provides a platform for the development of European Standards and other technical documents in relation to various kinds of products, materials, services and processes. Notably it is a member of the CSCG (Cybersecurity Coordination Group) to the EC. https://www.cen.eu/.

CEN-CENELEC JTC 21 – Artificial Intelligence. JTC 21 shapes standardisation that ensures the development of trustworthy AI systems that respect fundamental values and human rights recognized in Europe. It brings together nominated experts from 12 EU countries through their national standardization organisation and has a delegation from the European Commission to transpose primarily ISO standards into EU normative versions. It presently consists of five working groups on strategy, and operational, engineering, foundational/societal aspects, plus a group on joint AI EN standardization. https://www.bdva.eu/cen-cenelec-jtc-21-artificial-intelligence

CENELEC - European Committee for Electrotechnical Standardization. CENELEC is responsible for standardization in the electrotechnical engineering field. Its cyber security activity relates to coordination on smart grid information security. Notably it is a member of the CSCG (Cybersecurity Coordination Group) to the EC. CEN/CLC/JTC13 on Cybersecurity and Data Protection develops standards for cybersecurity and data protection covering all aspects of the evolving information society. http://www.cenelec.eu/.

CEPT – European Conference of Postal and Telecommunications Administrations. CEPT's activities include co-operation on commercial, operational, regulatory and technical standardisation issues among 46 countries. https://www.cept.org/cept CEPT's multiple constituent bodies are engaged in an array of AI related activities for both telecommunication and radio use.

CISA – Cybersecurity & Infrastructure Security Agency. CISA is the operational lead for U.S. cybersecurity and the national coordinator for critical infrastructure security and resilience that serves as the lead agency for AI security and associated standards. Globally, it sets the SBOM standards together with NTIA and NSA, and notably AI BOM specifications. https://www.cisa.gov/ai and https://www.cisa.gov/sbom and https://www.cisa.gov/sites/default/files/2023-10/Shifting-the-Balance-of-Cybersecurity-Risk-Principles-and-Approaches-for-Secure-by-Design-Software.pdf

- **National Vulnerability Database (NVD)**. CISA has operational responsibility for the NVD and adding multiple enhancements including a new process for delineating security flaws in AI technology assisted by the NSA AI Security Center.

COE – Council of Europe. COE is comprised by its Member and Observer States worldwide that have prepared an array of ICT-related treaty instruments. https://www.coe.int/en/web/about-us/who-we-are COE is a global international organisation and not part of the European Union. In early 2022, the COE inaugurated a Committee on Artificial Intelligence to draft a global Convention on Artificial Intelligence, Human Rights and the Rule of public law. https://www.coe.int/en/web/artificial-intelligence/cai. The new global legally binding treaty instrument on AI and human rights was adopted at Vilnius 5 Sep 2024 and has been signed by eleven nations. https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

CSA – Cloud Security Alliance. CSA is the principal global organisation of cloud services providers and promotes the use of best practices for providing security assurance within cloud computing and provides education on the uses of cloud computing to help secure all other forms of computing. https://cloudsecurityalliance.org/ CSA presently has five working groups focussed on AI: AI Safety Initiative, AI Technology and Risk, AI Governance & Compliance, AI Controls, and AI Organizational Responsibilities. https://cloudsecurityalliance.org/research/working-groups/ https://cloudsecurityalliance.org/research/working-groups/ai-technology-and-risk

- CSA Large Language Model (LLM) Threats Taxonomy, https://cloudsecurityalliance.org/artifacts/csa-large-language-model-llm-threats-taxonomy
- Securing LLM Backed Systems: Essential Authorization Practices, https://cloudsecurityalliance.org/artifacts/securing-llm-backed-systems-essential-authorization-practices
- AI Model Risk Management Framework, https://cloudsecurityalliance.org/artifacts/ai-model-risk-management-framework
- Using AI for Offensive Security, https://cloudsecurityalliance.org/artifacts/using-ai-for-offensive-security

CTI AI Toolbox Project. In the evolving landscape of Cyber Threat Intelligence (CTI), the need for advanced, AI-assisted tooling is critical. CTI AI Tools is dedicated to empowering the CTI community with state-of-the-art AI capabilities to enhance their threat intelligence efforts. By leveraging Large Language Models (LLMs), a suite of tools designed to streamline and enrich the analysis, understanding, and management of CTI data are provided. https://test.cti.tools/

DARPA – Defense Advanced Research Projects Agency. DARPA and ARPA-H's Artificial Intelligence Cyber Challenge (AIxCC) will bring together the foremost experts in AI and cybersecurity to safeguard the software. https://aicyberchallenge.com/ AI Forward is DARPA's initiative to explore new directions for artificial intelligence (AI) research that will result in trustworthy systems for national security missions. A generic AI portal provides access to many DARPA AI projects.

DSIT – Department for Science, Innovation & Technology. DSIT is leading a global initiative for an AI Code of Practice developed by the Department and based on the National Cyber Security Centre's (NCSC) Guidelines for secure AI system development co-sealed by agencies from 18 countries introduced into ETSI TC SAI as a global specification. [i.20] [i.1] The UK Office for Artificial Intelligence became part of DSIT in 2024.

EC - European Commission. The European Commission is the EU's executive body. Multiple directorates have significant cyber security roles: CONNECT (Communications Networks, Content and Technology; DIGIT (Informatics); GROW (Internal Market, Industry, Entrepreneurship and SMEs) Enterprise and Industry); HR (Human Resources and Security), JRC (Joint Research Centre), JUST (Justice and Consumers); HOME (Migration and Home Affairs); RTD (Research and Innovation). http://ec.europa.eu/about/index_en.htm. EU AI activity is primarily focussed on two legislative enactments, the AI Act, Cyber Resilience Act, and AI Liability Directive, and issuance of

standardisation requests. In addition, pursuant to the Cyber Resilience Act (CRA), products with digital elements classified as high-risk AI system, should comply with the relevant provisions of the CRA [i.2].

- **ENISA - European Union Agency for Cybersecurity.** ENISA is the EU agency dedicated to achieving a high common level of cybersecurity across Europe. It contributes to EU cyber policy, enhances the trustworthiness of ICT products, services and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. https://www.enisa.europa.eu/about-enisa/regulatory-framework ENISA maintains an AI reference site, produced several AI reports, and maintains an Ad-Hoc AI cybersecurity working group. https://www.enisa.europa.eu/topics/iot-and-smart-infrastructures/artificial_intelligence
- **European AI Office**. The European AI Office will be the centre of AI expertise across the EU. It will play a key role in implementing the AI Act - especially for general-purpose AI - foster the development and use of trustworthy AI, and international cooperation. https://digital-strategy.ec.europa.eu/en/policies/ai-office
- **TEF-Health**. Program for large-scale reference testing and experimentation facilities will offer a combination of physical and virtual facilities, in which technology providers can get support to test their latest AI-based soft-/hardware technologies in real-world environments. https://tefhealth.eu/what_is_tef

EP – European Parliament. European Parliament committees also play a role as co-legislator, being mostly concerned with civil rights aspects of technology and oversight of EC work. https://www.europarl.europa.eu/committees/en/aida/home/highlights

EDRM – Electronic Discovery Reference Model. EDRM creates global resources to improve e-discovery, privacy, security, and information governance that include leadership, standards, tools, guides, and test datasets to strengthen best practices throughout the world. https://edrm.net/edrm-projects/artificial-intelligence/

EFTA – European Free Trade Association. EFTA is the intergovernmental organisation of Iceland, Liechtenstein, Norway and Switzerland set up in 1960 by its then seven Member States for the promotion of free trade and economic integration between its members. https://www.efta.int/about-efta EFTA maintains an array of policy and legislative activities related to AI.

ETSI - European Telecommunications Standards Institute. ETSI produces globally-applicable standards for Information and Communications Technologies (ICT), including fixed, mobile, radio, converged, broadcast and internet technologies. Notably, it hosts the Technical Committee for Cybersecurity and is a member of the CSCG (Cybersecurity Coordination Group) to the EC. https://www.etsi.org/. Most of ETSI's Technical Committees and Industry Standards Groups have had AI related work for many years. Today, two serve as principal forums for AI work.

- **SAI - Securing Artificial Intelligence**. SAI develops 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. https://portal.etsi.org/tb.aspx?tbid=913&SubTB=913#/ SAI also facilitates AI workshops and conferences. https://www.etsi.org/events/2277-etsi-artificial-intelligence-conference

- **OCG SAI - Securing Artificial Intelligence**. OCG SAI exists as a subgroup within ETSI's Operational Co-ordination Group and facilitates the exchange of views and information on AI activities within ETSI. https://portal.etsi.org/tb.aspx?tbid=889&SubTB=889#/

Thirty-nine additional ETSI groups with 176 AI work items include: AFI, ARF, ATTM, CDM, CIM, CYBER, E4P, EE, eHealth, EMTEL, ENI, ERM, ETI, F5G, GRID, HF, INT, IPE, ISI, ITS, LI, MEC, NFV-IFA, NFV-REL, NFV-SEC, NFV-TST, NGP, NTECH, OEU, OneM2M, PDL, ISG-SAI, SES, SmartBAN, SmartM2M, STQ, TISPAN, UCI, USER, ZSM.

FIRST - Forum of Incident Response and Security Teams. FIRST is the international organization of CERTs/CSIRTs who cooperatively handle computer security incidents and promote incident prevention programs. FIRST members develop and share technical information, tools, methodologies, processes and best practices. It also promotes the creation and expansion of Incident Response teams globally though global, regional, and national workshops and conferences. http://www.first.org. Through FIRST's Special Interest Groups (SIGs) and BOFs, it develops an array of significant cyber security techniques and standards and maintains means for exchanging threat information globally. Over the past decade, FIRST has led may of the principal efforts to both advance the use of AI for cybersecurity risk reduction as well as potential threats. It treats the subject at most of its many conferences globally and maintains a Special Interest Group (SIG) devoted to AI Security. https://www.first.org/global/sigs/ai-security/

GPAI - Global Partnership on Artificial Intelligence. GPAI is a multi-stakeholder initiative which aims to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities. Its secretariat is hosted at the OECD, and consists of an array of bodies and working groups that include responsible AI, data governance, and specialised topic area. https://gpai.ai/projects/

Massachusetts Institute of Technology (MIT) AI Risk Repository. The AI Risk Repository is a comprehensive meta-review, database, and taxonomy of risks from Artificial Intelligence that provides a shared understanding of AI risks can impede our ability to comprehensively discuss, research, and react to them. This gap is addressed by creating an AI Risk Repository to serve as a common frame of reference. It is comprised of a living database of 777 risks extracted from 43 taxonomies, which can be filtered based on two overarching taxonomies and easily accessed, modified, and updated via the freely accessible website and online spreadsheets. https://airisk.mit.edu/

GSMA™ - GSM Association. GSMA is the global *organization* of GSM and related mobile providers and vendors, and today the largest telecommunication industry entity. GSMA's Fraud and Security Working Group is the global mechanism for exchanging information, developing standards and techniques, and collaborating on mobile cyber security in many other forums. It works closely with 3GPP groups, especially SA3 (Security) - providing support for cyber security information assurance initiatives. http://www.gsma.com/. As the global leader of the mobile industry, GSMA has initiated several applied initiatives and sharing the insights gained.
https://www.gsma.com/artificialintelligence/

IEEE – Institute of Electrical and Electronics Engineers. The IEEE is a global association of technical professionals. https://www.ieee.org/ Its AI activities parallel the three major subdivisions – published papers, education, conferences, and standards – principally relating to solid state and radio technologies. IEEE Transactions on AI. https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=9078688. IEEE Academy on AI. IEEE https://www.ieee.org/education/academy-index/artificial-intelligence1.html. Conference on AI (CAI), https://cai.ieee.org/2023/, AI Standards Committee, https://sagroups.ieee.org/ai-sc/standards/.

IETF - Internet Engineering Task Force. IETF is a standards organization for the Internet and is responsible for the technical standards that make up the Internet protocol suite. Many of these activities are cyber security related. Its Internet Architecture Board (IAB) also oversees development of cyber security capabilities. IETF groups change frequently and its website should be consulted for the latest activities. https://www.ietf.org. Recent AI related work is focussing on AI applied to IETF protocols such as a Framework for Network Management.
https://datatracker.ietf.org/doc/draft-pedro-nmrg-ai-framework/

ISO - International Organization for Standardization. The ISO is a Swiss based private international standards development and publishing body composed of representatives from various national standards organizations with multiple committees - several of which have significant cyber security related activity. http://www.iso.org.

- **ISO/IEC JTC 1/SC 42 – Artificial Intelligence**. SC42 Serve as the focus and proponent for JTC 1's standardization program on Artificial Intelligence and provides guidance to JTC 1, IEC, and ISO committees developing Artificial Intelligence applications. It presently has five working groups dealing with foundational standards, data, trustworthiness, use cases and applications, and computational approaches and characteristics of AI systems. It also maintains several ad hoc working groups and joint groups with SC7 (testing), 215 (health informatics), SC65 (functional safety), and 37 (natural language processing)
https://www.iso.org/committee/6794475.html

ITU - International Telecommunication Union. The ITU exists as the principal global intergovernmental body for ICT with three sectors dealing with the development and publication of Recommendations for radio systems (ITU-R), telecommunications (ITU-T), and development assistance (ITU-D). https://www.itu.int. AI work is spread across all the ITU organs. The ITU General Secretariat also publishes the annual compendium of UN Activities on AI.
https://www.itu.int/pub/S-GEN-UNACT

- **Plenipotentiary Conference**. Recent global treaty conferences have adopted AI related treaty resolutions.

- **ITU-R - Telecommunication Radiocommunication Sector**. The ITU-R consists of an Assembly that meets every four years to approve its structure and general work areas, six Study Groups that meet annually, and a Secretariat that publishes the materials and maintains several radiocommunication databases. The ITU-R AI activity is increasing significantly for radio services, devices, spectrum management, and mass media program. There are presently 8 adopted ITU-T specifications. https://www.itu.int/en/ITU-R/Pages/default.aspx.

- **ITU-T - Telecommunication Standardization Sector**. The ITU-T consists of an Assembly that meets every four years to approve its structure and general work areas, eleven Study Groups that meet annually, and a Secretariat that publishes the materials and maintains several telecommunications databases. The ITU-T AI

activity across all of the ITU's study groups and ICT services. There are presently 17 adopted specifications and 30 ongoing work items. https://www.itu.int/en/ITU-T/about/Pages/default.aspx There is a concentration in SG16 dealing with AI for multimedia services and image standards. ITU-T has also hosted AI Security Focus Groups and Workshops, as well as annual AI global summits for the past decade. https://en.wikipedia.org/wiki/ITU_AI_for_Good

- **ITU-D - Development Sector**. Provides technical assistance and in the creation, development and improvement of telecommunications in developing countries. http://www.itu.int/en/ITU-D/Pages/default.aspx. ITU-D. The Development Sector working groups have several new AI related work in ITU-D SG2.

MITRE - MITRE is a globally active non-profit research and development centre that is responsible for multiple significant global cyber security techniques, standards making and related secretariat activities. The activity occurs through multiple individual on-line activities, frequent workshops, and significant involvement in other global forums listed below. http://www.mitre.org/capabilities/cybersecurity/overview/cybersecurity-resources/standards.

- **ATLAS™ (Adversarial Threat Landscape for Artificial-Intelligence Systems)**: A globally accessible, living knowledge base of adversary tactics and techniques based on real-world attack observations and realistic demonstrations from AI red teams and security groups. https://atlas.mitre.org/

- **MITRE Engenuity Center for Threat-Informed Defense Project Secure AI**: AI-enabled systems are susceptible to traditional cybersecurity vulnerabilities, as well as novel attacks based on the unique vulnerabilities of AI-enabled systems. Cyber practitioners must take a holistic view of AI threats and vulnerabilities within the context of the larger system. The Secure AI project will: 1) expand the database of adversary tactics, techniques, and case studies for AI-enabled systems through incident sharing metrics and mechanisms, 2) document new case studies within ATLAS that address vulnerabilities in industry-relevant systems, including generative AI, 3) describe new relevant mitigations based on documented AI incidents, 4) align ATLAS tactics, techniques, and procedures (TTPs) with the current version of MITRE ATT&CK TTPs.

- **EMB3D™**: Threat model and resources that expands on three existing models – CWE, ATT&CK and CVE. The platform provides a cultivated knowledge base of cyber threats to embedded devices, offering a common understanding of these threats with the security mechanisms required to mitigate them. The treat model can be applied to AI platforms. https://emb3d.mitre.org/

NATO - North Atlantic Treaty Organization. Against the background of increasing dependence on technology and on the Internet, the Alliance is advancing its efforts to confront the wide range of cyber threats targeting NATO's networks on a daily basis. NATO has moved forward with five cyber security actions: developing NATO Policy on Cyber Defence, assisting individual Allies, increasing NATO cyber defence capacity, cooperating with partners, and cooperating with industry. The Allies have also committed to enhancing information sharing and mutual assistance in preventing, mitigating and recovering from cyber-attacks. http://www.nato.int/cps/en/natohq/topics_78170.htm. NATO has established an AI Strategy that focuses on responsible use and the creation of a Data and Artificial Intelligence Review Board (DARBB0 that is developing toolkits and certification standards. https://www.nato.int/cps/en/natohq/official_texts_208374.htm

NIST – Artificial Intelligence. NIST aims to cultivate trust in the design, development, use and governance of Artificial Intelligence (AI) technologies and systems in ways that enhance safety and security and improve quality of life. NIST focuses on improving measurement science, technology, standards and related tools — including evaluation and data. Notably, the work encompasses an AI Risk Management Framework, and Test Methods for critical systems [i.22] and [i.23] https://www.nist.gov/artificial-intelligence

OASIS - Organization for the Advancement of Structured Information Standards. OASIS is a major global body for developing and publishing worldwide standards for security, Internet of Things, cloud computing, energy, content technologies, emergency management, and other areas requiring structured information exchange. https://www.oasis-open.org/org/ In 2001, it released the standard for Artificial Intelligence Markup Language (AIML). https://xml.coverpages.org/aiml-ALICE.html Today, nearly all of its numerous standards committees are pursuing AI related work items for product and professional sectors and cybersecurity. https://www.oasis-open.org/?s=artificial+intelligence.

OECD - Organisation for Economic Co-operation and Development. OECD is an intergovernmental organisation with 38 member countries, founded in 1961 to stimulate economic progress and world trade. https://www.oecd.org/ The OECD maintains an array of activities and resources related to AI implementations and policies. https://www.oecd.org/digital/artificial-intelligence/

OpenSSF – Open Source Security Foundation. The OpenSSF maintains a Working Group on Artificial Intelligence/Machine Learning (AI/ML) Security. The WG was formed in September 2023 after the growing problem

of AI/ML Security in open source. It focuses on the possible security impacts of AI / ML technologies on open source software, maintainers, communities, and their adopters, along with how OSS projects could safely or effectively leverage LLMs to improve their security posture. https://github.com/ossf/ai-ml-security

Open Worldwide Application Security Project Foundation (OWASP) AI Exchange. An open source collaborative document to advance the development of global AI security standards and regulations. It provides a comprehensive overview of AI threats, vulnerabilities, and controls to foster alignment among different standardization initiatives**.** https://owaspai.org/

RAND. The first successful Artificial Intelligence program that used Information Processing Languages (IPLs) was developed in RAND's Systems Research Laboratory in 1957. It developed the concept of packet switched networks, and produced the first comprehensive studies on cybersecurity. RAND contributes to a wide array of AI developments today. https://www.rand.org/topics/artificial-intelligence.html

TCG - Trusted Computing Group®. TCG develops, defines and promotes open, vendor-neutral, global industry standards, supportive of a hardware-based root of trust, for interoperable trusted computing platforms. Its platforms provide for authentication, cloud security, data protection, mobile security, and network access & identity. TCG presently has twelve working groups. http://www.trustedcomputinggroup.org/ TCG is leveraging several of its platforms, including TPM, DICE, MARS, FIM, and RIM to enhance security of AI systems. https://trustedcomputinggroup.org/safeguarding-the-future-of-ai-the-imperative-for-responsible-development/

UNCITRAL - United Nations Commission on International Trade Law. UNCITRAL is the core legal body of the United Nations system in the field of international trade law. https://uncitral.un.org/en/about

UNCTAD – United Nations Conference on Trade and Development. UNCTAD is an intergovernmental organization within UN that promotes the interests of developing countries in world trade established in 1964 whose primary objective of UNCTAD is to formulate policies relating to all aspects of development, including trade, aid, transport, finance and technology. https://unctad.org/

UNESCO – United Nations Educational, Scientific and Cultural Organization. UNESCO is UN specialised agency with the aim of promoting world peace and security through international cooperation in education, arts, sciences and culture; and maintains one of the principal global treaties relating to Intellectual Property Rights. https://www.unesco.org/en/faq Many of its constituent bodies and groups are treating AI, including the impact on intellectual property.

UNIDO – United Nations Industrial Development Organization. UNIDO in July 2024 inaugurated the new Global Alliance on AI for Industry and Manufacturing Centre of Excellence (AIM Global CoE) in Shanghai.

United Nations AI Advisory Body. A multi-stakeholder High-level Advisory Body on AI to undertake analysis and advance recommendations for the international governance of AI. https://www.un.org/en/ai-advisory-body/about

WIPO – World Intellectual Property Organisation. The availability of large amounts of training data and advances in affordable high computing power are fueling AI's growth. AI intersects with intellectual property (IP) in a number of ways. https://www.wipo.int/about-ip/en/frontier_technologies/ai_and_ip.html

WHO – World Health Organization. Big Data and artificial intelligence. https://www.who.int/teams/health-ethics-governance/emerging-technologies/big-data-and-artificial-intelligence

World Bank. AI, big data diffusion, and policy. https://www.worldbank.org/en/events/2023/05/15/artificial-intelligence-big-data-and-policy

W3C® - World Wide Web Consortium. W3C develops protocols and guidelines for WWW services. http://www.w3c.org/ W3C AI activity has been focussed on accessibility research. https://www.w3.org/WAI/research/ai2023/

4.3 Major IT developer forums affecting AI security

Alibaba Cloud AI. Provides cloud infrastructure, data intelligence processing and AI engineering capabilities, algorithms that solve problems across a range of scenarios, and refined the solution for years in customer and Alibaba environments. Alibaba provides an all-in-one cloud-native group of AI solutions for enterprises and developers. https://www.alibabacloud.com/solutions/ai/data-intelligence

Apple Developer. Apple has released MLX, an AI framework designed to run on its chips and bring generative AI apps to Apple devices. https://developer.apple.com/machine-learning/

AWS - Amazon Web Services® Forum. A developer forum for services hosted on the Amazon data centre platforms. https://forums.aws.amazon.com/forum.jspa?forumID=30 AWS hosts Amazon AI Conclave – a developer event in collaboration with intel to discuss techniques, and share common use cases and ideas for implementing real-world projects using generative AI that brings together technology leaders, startup founders, and AWS AI & ML. https://aws.amazon.com/events/ai_conclave/

Baidu. Baidu introduced ERNIE Bot, its latest generative AI product and knowledge-enhanced large language model (LLM), on March 16, 2023. This advanced technology can comprehend human intentions and deliver accurate, logical, and fluent responses approaching human level. http://research.baidu.com/Blog/index-view?id=183

BlackBerry. A provider of intelligent security software and services to enterprises and governments. https://blogs.blackberry.com/en/category/security/artificial-intelligence

Broadcom. A provider of a broad array of ICT infrastructure technology, it is making available new AI platforms embedded in its products. https://www.broadcom.com/blog/generative-ai-and-the-transformation-of-everything

Cisco. A provider of a broad array of ICT infrastructure technology, it is making new AI solutions available as part of its portfolio. https://www.cisco.com/c/en_uk/solutions/artificial-intelligence.html#~benefits

GitHub Universe. A developer software exchange forum for AI software among over 100 million users. https://github.com/. GitHub Universe provides an annual developers forum. https://githubuniverse.com/

Google. As one of the early AI leaders, Google now has multiple products and services for both developers and customers. Google AI is a division of Google dedicated to artificial intelligence with multiple facilities worldwide and products at both the hardware and application levels based on its Gemini platform model and Bard AI tool. https://www.google.com/chrome/dev/ Other Google AI venues include the Google I/O developer conference, and Google Cloud – which includes a Security AI Workbench. https://cloud.google.com/products/ai https://cloud.google.com/security/ai

HP Enterprise AI. A developer forum to train and tune AI models faster. https://www.hpe.com/us/en/solutions/ai-artificial-intelligence.html

Huawei. Huawei is a leader in the development of AI software and hardware products, and participant in numerous technical and standards bodies to facilitate implementation. https://forum.huawei.com/enterprise/en/how-has-huawei-contributed-to-the-advancement-of-artificial-intelligence-ai-and-machine-learning-ml-technologies/

IBM. As one of the early AI leaders, IBM offers an array of AI platforms and shapes the governance platforms. https://www.ibm.com/artificial-intelligence. IBM offers an array of AI services, including a developer forum of next generation resources for AI builders to create trusted solutions. https://developer.ibm.com/technologies/artificial-intelligence/

Intel. A developer of an array of AI services, it provides developer forum for bringing AI everywhere with Intel AI solutions. https://www.intel.com/content/www/us/en/artificial-intelligence/overview.html

Microsoft. A developer of a broad array of solutions empowered by AI, including Copilot for everyday productivity and creativity. https://www.microsoft.com/en-us/ai

Nvidia GTC AI Conference. An annual conferences with workshops oriented around Nvidia products. https://www.nvidia.com/gtc/training/

Open AI Developer Conference. Open AI DevDay – focuses on the implementations of ChatGTP which has more than 100 million users - hosts two million developers and significantly shapes AI security and is hosting a global AI app store. https://devday.openai.com/

RSA Conference. RSA's annual event of tens of thousands cybersecurity professionals and hundreds of exhibitors has become its own global community and increasingly AI focussed. The 2024 event hosted 136 sessions on AI Security. https://www.rsaconference.com/usa

RedHat. A developer forum for applications running on the OpenShift Cloud OS. It has created Red Hat® OpenShift® AI as an open source platform for building, training, testing and serving models for AI-enabled applications. https://www.redhat.com/en/products/ai

Oracle® AI. Provides a comprehensive AI portfolio integrated in its cloud applications on a best-in-class AI infrastructure and with state-of-the-art generative AI innovations. https://www.oracle.com/artificial-intelligence/

Protect AI. Provides the first AI/ML Supply Chain Vulnerability Database as part of its Sightline platform that enable detection, assessment, and remediation of vulnerabilities in the AI/ML supply chain, providing early warning, contextualized insights, and remediation advice for risks. https://protectai.com/newsroom/announcing-sightline-the-first-ai/ml-vulnerability-database

SourceForge. A developer software exchange forum which has expanded to include Open Source AI Software development. https://sourceforge.net/directory/artificial-intelligence/

Splunk AI. Splunk is a leading developer of resilient solutions for complex networked systems that has adapted AI capabilities to further customer capabilities. https://www.splunk.com/en_us/solutions/splunk-artificial-intelligence.html

Tesla. Tesla is a leading develop of AI hardware, software, and services for transportation vehicles and autonomous objects. https://www.tesla.com/AI

VMware® Community. A developer forum for applications running the VMware OS which has expanded to include AI Solutions. https://www.vmware.com/artificial-intelligence.html

XDA Developers Forum. A developer software exchange forum which has expanded to include AI. https://xdaforums.com/c/artificial-intelligence-ai-machine-learning-ml.12755/

4.4 Activities for continuous information exchange

.AI domain, 2,160 registered domains as of 20 Nov 2023

…..

4.5 Centres of excellence

AICOE - AI Center of Excellence. https://aicoe.ai/

CIS - Le Centre Internet et Société. CIS is a research centre in Centre national de la recherche scientifique (CNRS) aims to foster expertise and critical reflection on the emerging issues of artificial intelligence (AI) and maintains the working group GDR2019 on Artificial intelligence and humanities. https://cis.cnrs.fr/en/working-groups-gdr/

DFKI AI - Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI). https://www.dfki.de/en/web

European Association for Artificial Intelligence (EurAi). EurAi was established in July 1982 (formerly ECCAI) as a representative body for the European Artificial Intelligence community. Its aim is to promote the study, research and application of Artificial Intelligence in Europe. EurAI sponsors invited talks at international conferences. Since 2015, EurAi has sponsored more than 70 conferences. https://www.eurai.org/about

Johns Hopkins Data Science and AI Institute. https://ep.jhu.edu/programs/artificial-intelligence/

Ministère de l'Enseignement Supérieur et de la Recherche Intelligence artificielle. https://www.enseignementsup-recherche.gouv.fr/fr/intelligence-artificielle-de-quoi-parle-t-91190

NUS AII – National University of Singapore Artificial Intelligence Institute. https://ai.nus.edu.sg/

Open_Future. AI and the Commons explores the intersections between AI and openness. The release of powerful AI models under open licenses was a breakthrough moment for the world of open, indicating the emergence of a new field in which the principles of open are applied. This is a nascent field in which there are still no established norms for openly sharing different elements of the machine learning stack: data, model, and code. https://openfuture.eu/our-work/ai-and-the-commons/

QUB AI – Queen's University Belfast AI Hub. Artificial Intelligence Hub - Queen's DigiHub

SAFE – Stanford Center for AI Safety. https://aisafety.stanford.edu/

SAIRI – Shanghai Artificial Intelligence Research Institute. https://www.sairi.com.cn

SEI – Software Engineering Institute. https://www.sei.cmu.edu/our-work/artificial-intelligence-engineering/

SGAICO - Swiss Group for Artificial Intelligence and Cognitive Science - https://www.swissinformatics.org/sgaico.php?Lang=en

University of Delaware AICOE. https://sites.udel.edu/ai/

USDOD COE. https://www.dod-coe4ai-ml.org/

University of Edinburgh Quantum & AI Centre of Excellence. https://www.ed.ac.uk/c/60-years-computer-science-ai

…..

4.6 Reference libraries, continuing conferences, and publications

Ai4 2025, Las Vegas. https://ai4.io/vegas/

The AI Conference, San Francisco. https://aiconference.com/

Dreamforce, San Francisco. https://www.salesforce.com/dreamforce/

PYTORCH, San Francisco, Tokyo. https://events.linuxfoundation.org/about/calendar/?_sft_lfevent-category=ai-events

Ray Summit, San Francisco. https://raysummit.anyscale.com/flow/anyscale/raysummit2024/landing/page/eventsite

AI &Big Data Expo, Amsterdam. https://www.ai-expo.net/

World Summit AI, Amsterdam. https://worldsummit.ai/

Big Data Conference Europe, Lithuania. https://bigdataconference.eu/

AI World Congress, London. https://aiconference.london/

4.7 Heritage sites and historical collections

Museum of AI, https://museumof.ai/

The History of Artificial Intelligence, https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

4.8 Additional exchange sources and methods

Annex A:
National AI security ecosystems

The OECD maintains a National AI Policies and Strategies site. OECD's live repository of AI strategies & policies - OECD.AI

United Kingdom. National AI Strategy. https://assets.publishing.service.gov.uk/media/614db4ecd3bf7f7187208500/National_AI_Strategy__mobile_version_.pdf

United States. Department of Homeland Security, "Mitigating Artificial Intelligence (AI) Risk: Safety and Security Guidelines for Critical Infrastructure Owners and Operators", Apr 2024. https://www.dhs.gov/sites/default/files/2024-04/24_0426_dhs_ai-ci-safety-security-guidelines-508c.pdf

United States, Executive Office of the President, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Annex B:
AI security relationship diagrams

Annex C:
Societal constraints on AI

The societal impact of AI systems is increasingly discussed in relation to developments of the EU AI Act, technology development of generative AI and its potential copyright infringements, deep fake media, misinformation, job losses etc. The types of harms from AI can be separated into individual, collective and societal.

Individual harm occurs when one or more interests of an individual are wrongfully thwarted. This is the case, for instance, when the use of a biased facial recognition system—whether in the context of law enforcement or in other domains—leads to wrongful discrimination.

Collective harm occurs when one or more interests of a collective or group of individuals are wrongfully thwarted. Just as a collective consists of the sum of individuals, so does this harm consist of the sum of harms suffered by individual members of the collective. The use of the abovementioned biased facial recognition system, for instance, can give rise to collective harm, when a specific collective of people are discriminated against.

Societal harm occurs when one or more interests of society are wrongfully thwarted. In contrast with the above, societal harm is thus not concerned with the interests of a particular individual or the interests shared by a collective of individuals. Instead, it concerns harm to an interest held by society at large, going over and above the sum of individual interests.

In the broader media landscape and public sphere, the application of deep technology is what often comes to mind when considered societal harm of AI. A deepfake refers to a technique that uses AI to create or alter videos, images, or audio recordings to make them appear authentic but are manipulated or synthesized. The term "deepfake" is a combination of "deep learning" and "fake."

Once trained, deepfake algorithms can generate new content by synthesizing realistic-looking faces or altering existing videos by swapping faces or modifying expressions, gestures, and lip movements. This deepfake technology can create highly convincing reproduced videos or images, often involving celebrities or public figures, and can be used for various purposes, including entertainment, political satire, or malicious activities such as spreading disinformation or defamation.

While deepfakes have gained attention for their potential negative implications, including their role in deceptive campaigns or privacy violations, researchers and technology companies are also actively developing countermeasures to detect and mitigate the impact of deepfake content.

The application of deepfake technology raises several ethical dilemmas and concerns. Deepfakes can use images or videos of individuals without their consent, potentially violating their privacy and autonomy. Non-consensual deepfake videos can cause significant harm to individuals by exploiting and manipulating their likeness for explicit or damaging content. This can lead to severe harm especially when minors are targeted, including loss of employment opportunities, public humiliation, or damage to personal relationships. There is risk of victims if feeling helpless and overwhelmed of committing self-harm or suicide.

Questions surrounding the boundaries of creativity, intellectual property, and artistic expression come up when determining when and where to use deepfake technology. It can challenge the authenticity and integrity of art, entertainment, and journalism, as well as impact the livelihoods of artists and content creators.

It is worth noting that deepfake technology is not all bad. There are positive potentials in deepfake technology. It opens doors for use cases that can bring enhancements to the world, such as improving accessibility for individuals with disabilities, educational tools to simulate scenarios and events that are otherwise unreachable or inventing personalized virtual assistants capable of human-like interactions and virtual companionship.

Annex:
Bibliography

-

History

Document history		
V0.0.1	January 2024	Initial draft
V0.0.2	February 2024	Early draft
V0.0.3	February 2024	Early draft
V0.0.4	May 2024	Early draft
V0.0.5	September 2024	Stable draft
V0.0.6	Dece,ber 2024	Stable draft