ETSI TR 104 119 V1.1.1 (2025-09)

TECHNICAL REPORT

Methods for Testing & Specification (MTS); AI Testing; Guidelines for Documentation of AI-enabled Systems

Reference
DTR/MTS-104119

Keywords
AI, testing

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 Methods for Testing and Specification (MTS).

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.

1 Scope

The present document provides guidelines and recommendations for documentation schemes that support the continuous and consistent documentation of quality and quality related attributes for AI-enabled systems.

This includes an analysis of current documentation schemes and Use case examples. It also defines a process how to document AI-enabled systems.

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] ISO/IEC TR 24028:2020: "Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence".

[i.2] 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, i.92i.2 (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828.

[i.3] OECD, adopted on 2019-05-22, amended on 2023-11-08: "Recommendation of the Council on Artificial Intelligence".

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

[i.5] High-Level Expert Group on Artificial Intelligence (HLEG AI), published 2019-04-08: "Ethics guidelines for trustworthy AI".

[i.6] ISO/IEC 25059:2023: "Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems".

[i.7] ISO/IEC 25010:2023: "Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Product quality model".

[i.8] ISO/IEC 25019:2023: "Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality-in-use model".

[i.9] ISO/IEC TR 5469:2024: "Artificial intelligence — Functional safety and AI systems".

[i.10] ISO/IEC 23894:2023: "Information technology — Artificial intelligence — Guidance on risk management".

[i.11] ISO/IEC 42001:2023: "Artificial intelligence - Management system".

[i.12] ISO/IEC 27001:2022: "Information security, cybersecurity and privacy protection — Information security management systems — Requirements".

[i.13] ISO 31000:2018: "Risk management — Guidelines".

[i.14] ISO 9241-210:2019: "Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems".

[i.15] ISO 14971:2019: "Medical devices — Application of risk management to medical devices".

[i.16] ISO 13849-1:2023: "Safety of machinery — Safety-related parts of control systems. Part 1: General principles for design".

[i.17] ISO 21815-1:2022: "Earth-moving machinery — Collision warning and avoidance. Part 1: General requirements".

[i.18] EN ISO 16001:2017 "Earth-moving machinery — Object detection systems and visibility aids — Performance requirements and tests".

[i.19] T. Gebru, J. Morgenstern, B. Vecchione, et al.: "Datasheets for datasets", Communications of the ACM, volume 64, issue 12, pp. 86-92.

[i.20] J. Giner-Miguelez, A. Gómez, and J. Cabot: "DescribeML: a tool for describing machine learning datasets", MODELS '22 Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems, pp. 22-26, published 2022-11-09.

[i.21] D. Adkins, B. Alsallakh, A. Cheema, et al.: "Method cards for prescriptive machine-learning transparency", CAIN '22 Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI, pp. 90-100, published 2022-10-17.

[i.22] M. Arnold, R. K. E. Bellamy, M. Hind, et al.: "FactSheets: Increasing trust in AI services through supplier's declarations of conformity", IBM Journal of Research and Development, volume 63, issue 4/5, pp. 6:1-13, published 2019-09-18.

[i.23] ISO/IEC TR 29119-11:2020: "Software and systems engineering — Software testing — Part 11: Guidelines on the testing of AI-based systems".

[i.24] ISO/IEC/IEEE 26514:2022: "Systems and software engineering — Design and development of information for users".

[i.25] ETSI EG 204 061: "Human Factors (HF); ETSI Accessibility Strategy; Accessibility of ETSI Deliverables and Improvement of the Development Process of Deliverables".

[i.26] EN 301 549 (V3.2.1) (2021-03): "Accessibility requirements for ICT products and services".

[i.27] ISO/IEC/IEEE 26511:2018: "Systems and software engineering — Requirements for managers of information for users of systems, software, and services".

[i.28] ISO/IEC/IEEE 26512:2018: "Systems and software engineering — Requirements for acquirers and suppliers of information for users".

[i.29] ISO/IEC/IEEE 26513:2017: "Systems and software engineering — Requirements for testers and reviewers of information for users".

[i.30] ISO/IEC/IEEE 15026-2:2022: "Systems and software engineering — Systems and software assurance — Part 2: Assurance case".

[i.31] Center for Democracy & Technology (CDT): "Best Practices in AI Documentation: The Imperative of Evidence from Practice".

[i.32] Chmielinski, K. S., Newman, S., Taylor, M., Joseph, J., Thomas, K., Yurkofsky, J., & Qiu, Y. C.: "The dataset nutrition label (2nd gen): Leveraging context to mitigate harms in artificial intelligence", arXiv preprint arXiv:2201.03954.

[i.33] Procope C., Cheema A., Adkins D., Alsallakh B., Green N., McReynolds E., Pehl G., Wang E., & Zvyagina P.: "System-Level Transparency of Machine Learning", Technical Report, Meta AI.

[i.34] Hutchinson, B., Smart, A., Hanna, A., Denton, E., Greer, C., Kjartansson, O., & Mitchell, M.: "Towards accountability for machine learning datasets: Practices from software engineering and infrastructure", In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 560-575).

[i.35] Pushkarna, M., Zaldivar, A., & Kjartansson, O.: "Data cards: Purposeful and transparent dataset documentation for responsible ai", In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 1776-1826).

[i.36] L. Derczynski, H. Rose Kirk, V. Balachandran, et al.: "Assessing Language Model Deployment with Risk Cards", v1, published 2023-03-31.

[i.37] N. Barman, Y. Reznik, and M. Martini: "Datasheet for Subjective and Objective Quality Assessment Datasets", IEEE, 15th International Conference on Quality of Multimedia Experience (QoMEX) 2023, published 2023-07-18.

[i.38] M. P. Sendak, M. Gao, N. Brajer, and S. Balu: "Presenting machine learning model information to clinical end users with model facts labels", npj Digit. Med. 3, 41, published 2020-03-23.

[i.39] M. P. Hauer, T. D. Krafft, K. Zweig: "Overview of transparency and inspectability mechanisms to achieve accountability of artificial intelligence systems", Cambridge University Press, published 2023-11-24.

[i.40] M. Mitchell, S. Wu, A. Zaldivar, et al.: "Model Cards for Model Reporting", Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 220-229, published 2019-01-29.

[i.41] T. K. Gilbert, N. Lambert, S. Dean, et al.: "Reward Reports for Reinforcement Learning", AIES '23 Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, pp. 84-130, published 2023-08-29.

[i.42] Zweig, K. A., Wenzelburger, G., & Krafft, T. D.: "On chances and risks of security related algorithmic decision making systems", European Journal for Security Research, 3, pp. 181-203.

[i.43] Wirth, R., & Hipp, J.: "CRISP-DM: Towards a standard process model for data mining", In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1, pp. 29-39).

[i.44] Angelina McMillan-Major, Emily M. Bender: "A Guide for Creating and Documenting Language Datasets with Data Statements Schema Version 3".

[i.45] Angelina McMillan-Major, Salomey Osei, et al.: "Reusable Templates and Guides For Documenting Datasets and Models for Natural Language Processing and Generation: A Case Study of the HuggingFace and GEM Data and Model Cards".

[i.46] Diaz, M., Kivlichan, I. D., Rosen, R., Baker, D., Amironesei, R., Prabhakaran, V., & Denton, E. (2022, June): "CrowdWorkSheets: Accounting for individual and collective identities underlying crowdsourced dataset annotation".

[i.47] Goel, K., Rajani, N. F., Vig, J., Tan, S., Wu, J., Zheng, S., Xiong, C., Bansal, M., & Ré, C. (2021, January): "Robustness Gym: Unifying the NLP evaluation landscape".

[i.48] Raji, I. D., & Yang, J. (2020, January): "ABOUT ML: Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles".

[i.49] Shen, H., Deng, W. H., et al.: "Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation", Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.

[i.50] C. Seifert, S. Scherzinger, et al.: "Towards Generating Consumer Labels for Machine Learning Models", Invited Paper 2019.

[i.51] Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies.

[i.52] Floridi, L., Cowls, J., Beltrametti, M. et al.: "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", Minds & Machines 28, pp. 689-707 (2018).

[i.53] Adewole Adamson and Avery Smith: "Machine Learning and Health Care Disparities in Dermatology". JAMA Dermatology 154 (08/2018).

[i.54] S. Alder: "AI Company Exposed 2.5 Million Patient Records Over the Internet". HIPPA Journal (2020).

[i.55] Anmol Arora: "Conceptualising Artificial Intelligence as a Digital Healthcare Innovation: An Introductory Review". Medical Devices: Evidence and Research Volume 13 (08/2020), pp. 223-230.

[i.56] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.

[i.57] K. Ferryman and M. Pitcan: "Fairness in precision medicine". Data & Society, 2018.

[i.58] S. D. Fihn, S. Saria, E. Mendonça, S. Hain, M. Matheny, N. Shah, H. Liu, and A. Auerbach: "Deploying AI in clinical settings". In Artificial intelligence in health care: The hope, the hype, the promise, the peril, Matheny M, Israni ST, Ahmed M, and Whicher D (Eds.). National Academy of Medicine, Washington, DC, 2019.

[i.59] U.S. Food and Drug Administration (FDA): "Proposed Regulatory Framework for Modifications to Artificial Intelligence / Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)", 2019.

[i.60] Sara Gerke, Timo Minssen, and Glenn Cohen: "Ethical and legal challenges of artificial intelligence-driven healthcare", 2020, pp. 295-336.

[i.61] M. Ghassemi: "Exploring Healthy Models in ML for Health. AI for Healthcare Equity Conference", AI & Health at MIT, 2021.

[i.62] WHO TEAM: Health Ethics & Governance: "Ethics and governance of artificial intelligence for health: WHO guidance". Geneva: World Health Organization (WHO), 2021.

[i.63] Lucy Hocking, Sarah Parks, Marlene Altenhofer, and Salil Gunashekar: "Reuse of health data by the European pharmaceutical industry: Current practice and implications for the future", 2019.

[i.64] Kelly M. Hoffman, Sophie Trawalter, Jordan R. Axt, and M. Norman Oliver: "Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites". Proceedings of the National Academy of Sciences 113, 16 (2016), pp. 4296-4301, 2016.

[i.65] Bert-Jaap Koops: "The concept of function creep". Law, Innovation and Technology 13, 1 (2021), pp. 29-56, 2021.

[i.66] Zachary Chase Lipton: "The Doctor Just Won't Accept That!", arXiv: Machine Learning (2017).

[i.67] Alex McKeown, Miranda Mourby, Paul Harrison, Sophie Walker, Mark Sheehan, and Ilina Singh: "Ethical Issues in Consent for the Reuse of Data in Health Data Platforms". Science and Engineering Ethics 27 (02/2021).

[i.68] Marçal Mora-Cantallops, Salvador Sánchez-Alonso, Elena García-Barriocanal, and Miguel-Angel Sicilia: "Traceability for Trustworthy AI: A Review of Models and Tools". Big Data and Cognitive Computing 5, 2 (2021).

[i.69] Myura Nagendran, Yang Chen, Christopher A Lovejoy, Anthony C Gordon, Matthieu Komorowski, Hugh Harvey, Eric J Topol, John P A Ioannidis, Gary S Collins, and Mahiben Maruthappu: "Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies". BMJ 368 (03/2020), m689.

[i.70] BBC News (2017): "Google DeepMind NHS app test broke UK privacy law".

[i.71] J. Brian Pickering: "Trust, but Verify: Informed Consent, AI Technologies, and Public Health Emergencies". Future Internet 13 (2021), p. 132.

[i.72] Thomas Ploug and Soren Holm.: "Meta Consent - A Flexible Solution to the Problem of Secondary Use of Health Data". Bioethics 30, 9 (2016), pp. 721-732.

[i.73] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes: "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing", 2020.

[i.74] Andrew D. Selbst and Julia E. Powles: "Meaningful Information" and the Right to Explanation. In FAT, 2017.

[i.75] Ameneh Shamekhi, Ha Trinh, Timothy W. Bickmore, Tamara R. DeAngelis, Theresa Ellis, Bethlyn V. Houlihan, and Nancy K. Latham: "A Virtual Self-Care Coach for Individuals with Spinal Cord Injury". In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (Reno, Nevada, USA) (ASSETS '16). Association for Computing Machinery, New York, NY, USA, pp. 327-328.

[i.76] Helen Smith: "Clinical AI: opacity, accountability, responsibility and liability". AI & SOCIETY 36 (06/2021).

[i.77] Darshali Vyas, Leo Eisenstein, and David Jones: "Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms". New England Journal of Medicine 383 (06 2020).

[i.78] Blay Whitby: "Automating Medicine the Ethical Way", 2015.

[i.79] Guang Yang, Qinghao Ye, and Jun Xia: "Unbox the Black-box for the Medical Explainable AI via Multi-modal and Multi-centre Data Fusion: A Mini-Review", Two Showcases and Beyond, 2021.

[i.80] Robert Challen, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, and Krasimira Tsaneva-Atanasova: "Artificial intelligence, bias and clinical safety". BMJ Quality & Safety 28 (01/2019), bmjqs-2018.

[i.81] Samer Ellahham, Nour Ellahham, and Mecit Can Emre Simsekler: "Application of Artificial Intelligence in the Health Care Safety Context: Opportunities and Challenges". American Journal of Medical Quality 35, 4 (2020), pp. 341-348.

[i.82] Ravi Manne and Sneha Kantheti: "Application of Artificial Intelligence in Healthcare: Chances and Challenges". Current Journal of Applied Science and Technology 40 (05/2021), pp. 78-89.

[i.83] Jessica Morley and Luciano Floridi: "An Ethically Mindful Approach to AI for Health Care". The Lancet 395 (01/2020), pp. 254-255.

[i.84] Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023): "Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation".

[i.85] Article 22 EU GDPR: "Automated individual decision-making, including profiling".

[i.86] "FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare".

[i.87] Annemarie Hamlin, Chris Rubio, Michele DeSilva: "Technical Writing".

[i.88] Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC.

[i.89] ISO/IEC AWI 42102: "Information technology — Artificial intelligence — Taxonomy of AI system methods and capabilities" [Draft]. Retrieved June 20, 2025.

[i.90] Schmid, T. (2023): "A Systematic and Efficient Approach to the Design of Modular Hybrid AI Systems". In AAAI Spring Symposium: MAKE.

[i.91] Holoyad, T., Schmid, T., & Hildesheim, W.: "Managing and understanding artificial intelligence: From classical to generative AI - A practice guide for decision-makers, developers and regulators in the age of the EU AI Act". Springer.

3 Definition of terms, symbols and abbreviations

3.1 Terms

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

affected person: individuals with limited technical knowledge who may be impacted by AI systems

NOTE: They require protection from arbitrary decisions and risks, and may need to take legal measures if they feel unfairly treated. They can be represented by NGOs who possess greater capacities, knowledge, and power. See also: AI subject.

AI customer: organization or entity that uses an AI product or service either directly or by its provision to AI users

NOTE: Aligned with ISO/IEC 22989 [i.4]: AI customer.

AI partner: organization or entity that provides services in the context of AI

NOTE 1: AI partners can perform technical development of AI products or services, conduct testing and validation of AI products and services, audit AI usage, evaluate AI products or services and perform other tasks.

NOTE 2: This includes roles like AI system integrators, who incorporate AI components into broader systems, data providers, AI evaluators, and AI auditors.

NOTE 3: Aligned with ISO/IEC 22989 [i.4]: AI partner.

AI producer: organization or entity that designs, develops, tests and deploys products or services that use one or more AI system

NOTE 1: This includes AI developers, who focus on creating AI models, implementing computational processes, and verifying both the computation and model performance.

NOTE 2: According to the EU AI Act [i.2] an AI producer is also included in the term Deployer, as "a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity".

NOTE 3: Aligned with ISO/IEC 22989 [i.4]: AI producer.

AI provider: organization or entity that offers products or services utilizing one or more AI systems

NOTE 1: AI providers include AI platform providers, who enable other stakeholders to produce AI services or products, and AI service or product providers, who deliver AI solutions directly to customers.

NOTE 2: The EU AI Act [i.2] defines Provider as "a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge".

NOTE 3: Aligned with ISO/IEC 22989 [i.4]: AI provider.

AI subject: organization or entity that is impacted by an AI system, service or product

NOTE 1: This can also include individuals or communities affected by AI applications, such as users of social networks or drivers of AI-automated vehicles.

NOTE 2: The EU AI Act [i.2] uses the term affected person equivalently.

NOTE 3: Aligned with ISO/IEC 22989 [i.4]: AI subject.

AI user: organization or entity that uses AI products or services

NOTE 1: AI user is a sub-role of AI customer.

NOTE 2: Aligned with ISO/IEC 22989 [i.4]: AI users.

auditor: professionals with deep technical knowledge and understanding of standards and regulations

NOTE 1: They identify compliance issues and ensure AI systems meet ethical and operational benchmarks, facilitating successful certification and deployment.

NOTE 2: Aligned with ISO/IEC 22989 [i.4]: AI auditor.

bias: systematic difference in treatment of certain objects, people, or groups in comparison to others

NOTE: Aligned with ISO/IEC 22989 [i.4]: bias.

deployer: entities utilizing AI-based products provided by others, requiring sufficient information to effectively incorporate these systems into their own products or services

NOTE: The EU AI Act [i.2] uses the term 'deployer' as "a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity". See also: AI producer.

document stakeholder: individual, group, or organization that can affect, be affected by or perceive itself to be affected by the document

NOTE: Aligned with ISO/IEC 22989 [i.4]: stakeholder.

documentation approach: strategy or plan to create or maintain a document by well structured decisions and activities which base on fundamental goals

documentation item: subject of a documentation artifact

NOTE: The subject of a documentation (i.e. the documentation item) can be e.g. a software component, AI model, AI-enabled systems, training data, processes, or organizational structures.

documentation scheme: framework of methods, tools and templates that realizes a documentation approach

NOTE: Documentation schemes are often supported by successful empiric evidence.

documentation technique: specific format, structure and modality of how information is presented by a document

NOTE: Documentation techniques are chosen to most effectively address the intended audience (recipients of the information to be transported) of a document.

provider (documentation): entities responsible for documenting the technical details of AI products, from requirements to test results, ensuring high-quality and trustworthy AI systems

regulator: organizations and entities that have the authority to set, implement and enforce the legal requirements as intended in policies set forth by policy makers

NOTE 1: Those entities are e.g. governmental organizations or bodies like the European Unio. They are responsible for setting guidelines and standards to ensure the ethical development and deployment of AI systems. They focus e.g. on transparency, accountability, and fairness, and assess the applicability and effects of current regulations in AI.

NOTE 2: Aligned with ISO/IEC 22989 [i.4]: regulators.

relevant authorities: organizations or entities that can have an impact on an AI system, service or product

NOTE 1: This includes policy makers and regulators.

NOTE 2: Aligned with ISO/IEC 22989 [i.4]: relevant authorities.

trust: individual stakeholders confidence that an entity, organization or individual behaves or reacts as expected by the individual

NOTE: Subjective factors influence a person's trust in a system, including personal experiences, beliefs, needs, emotional and rational thinking based on the perceived information on an AI system's impact. Mostly the system behaviour expected by a person includes the absence of negative consequences to itself.

trustworthiness: ability to meet stakeholder expectations in a verifiable way

NOTE: Aligned with ISO/IEC 22989 [i.4]: trustworthiness.

3.2 Symbols

Void.

3.3 Abbreviations

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

CE	Conformité Européenne (European Conformity)
CRISP-DM	CRoss-Industry Standard Process for Data Mining
DSL	Domain-Specific Language
GDPR	General Data Protection Regulation
GPAI	General-Purpose AI (System)
HIPAA	Health Insurance Portability and Accountability Act
HLEG	High-Level Expert Group on Artificial Intelligence
KPI	Key Performance Indicator
LLM	Large Language Model
NLP	Natural Language Processing
NLU	Natural Language Understanding
QoE	Quality of Experience
SMEs	Small and Medium Enterprises

4 Purpose of Documentation

The motivation to create effective documentation for AI-enabled systems became very high, since those systems are used so widely and regulation has been set up to limit potential risks that may arise from those system developed and put on the market with inappropriate characteristics. One goal of mitigating those risks is to gain trust of involved and affected persons by such AI-enabled systems.

Trust and trustworthiness

The deeper technical systems take effect on human life, the stronger the needs of humans are to trust those systems. The term *trust*, however, is not an objective characteristic of an entity all humans can believe in but rather a statement of an individual person who trusts the entity. Typically, one can trust a manifold of entities, such as other persons, groups of persons, organizations, technical systems, processes and so on. Trust itself can be described as the expectation of a person into an entity that it will behave "as expected" without having negative consequences on this person.

While the term trust can be seen as a statement of an individual person or stakeholder, *trustworthiness* is a characteristic of the entity, e.g. defined by ISO/IEC 22989 [i.4] as the "ability to meet stakeholders' expectations in a verifiable way". However, verifying trustworthiness cannot easily be measured objectively, as the corresponding stakeholder's expectations may depends on individuals. Thus, the term trustworthiness can only be verified, when it is broken down into a set of characteristics that can be measured more objectively. Currently, there exist several proposals of how trustworthiness can be supported. Annex B lists and discusses these definitions.

But even if all the broken-down characteristics have been assessed, trust can only arise in a stakeholder's mind, if the stakeholder is actually aware of all these characteristics and information. Thus, it is important to have appropriate means, i.e. documentations, to provide this information.

From regulatory perspective, the basis for trustworthiness is formed by legal requirements. The European AI Act requires, along with other standardization documents, compliance with quality characteristics such as transparency, accuracy, robustness, and fairness, as well as compliance with legal obligations to protect fundamental rights such as privacy, non-discrimination, and human control.

In practice, however, trust in AI systems is influenced by numerous factors that go beyond regulatory compliance. In this context, humans develop trust based on their previous experiences with AI, their personal beliefs and needs, and how well certain effects of the system match their expectations. The perception of trust is shaped by individual factors, such as the user's understanding of how the AI system works, their interaction with the system and their general attitude towards the technology. A catalyst is therefore needed to bridge the gap between the definition of trustworthiness in the AI law and the actual trust perceived by users. This catalyst is based on transparency in a decisive manner, enabling clarity to make AI systems assessable to demonstrate the system's alignment with the human dimensions of trust. To create a trustworthy AI systems, humans trust in, it is essential to consider both the formal requirements as well as the human dimension of trust.

Purposes of documenting AI based systems

AI documentation serves multiple critical purposes, each tailored to different stakeholders and regulatory environments. At its core, documentation is essential for ensuring transparency, accountability, and trustworthiness in AI systems, a deeper analysis of the risks associated with a not well documented AI systems is reported in Annex C.
AI documentation can vary widely, from fulfilling regulatory requirements to providing insights into the technical foundations of the system, depending on its intended audience and purpose:

- **Compliance with regulation and standards.**

  One of the primary reasons for AI documentation is to comply with regulatory standards. Governments and industry bodies impose strict guidelines to ensure AI systems operate fairly, safely, and without undue bias. Proper documentation helps organizations demonstrate compliance with laws such as the EU AI Act and GDPR. It details how data is handled, how models are tested for bias and fairness, and how decisions made by AI systems align with ethical and legal standards. By maintaining comprehensive records, organizations can navigate audits more effectively and mitigate legal risks.

- **Information source for benchmarking.**

  Beyond regulatory compliance, AI documentation also plays a crucial role in benchmarking. AI systems are often evaluated based on their accuracy, robustness, and efficiency. Benchmarking documentation provides details on the datasets used for evaluation, the methodologies employed, and comparative results against industry standards or previous models. This transparency allows researchers and developers to assess the strengths and weaknesses of an AI model, ensuring continuous improvement and fostering innovation.

- **Data transparency.**

  Another key aspect of AI documentation is providing insight into the data sources used for training and testing. Data is the foundation of any AI system, and understanding its origins, composition, and potential biases is vital. Documentation should describe how datasets were collected, cleaned, and processed. It should highlight any limitations or biases inherent in the data and outline steps taken to mitigate these issues. This level of transparency helps users and regulators assess the reliability and fairness of the AI system.

- **System design and development process transparency.**

  The development process and internal architecture of an AI system also require thorough documentation. This includes explanations of model design choices, training procedures, hyperparameter tuning, and algorithmic modifications. For engineers and researchers, such documentation serves as a roadmap, facilitating collaboration, debugging, and future iterations. Understanding how an AI system was built and the reasoning behind its design choices allows teams to refine their approaches and improve performance over time.

All these documentation efforts contribute to the broader goal of increasing trust in AI systems. Whether it's regulators ensuring compliance, researchers benchmarking performance, or end-users seeking reassurance about fairness and reliability, well-documented AI systems foster confidence. Transparency in data, design, and decision-making processes reassures stakeholders that the AI operates as intended and aligns with ethical and legal expectations. AI documentation is not just an administrative requirement, it is a fundamental pillar of responsible AI development. It bridges the gap between technical innovation and public trust, ensuring that AI systems are not only effective but also accountable, fair, and reliable.

5 Motivation for a Harmonized Documentation Scheme

5.1 Current AI Documentation Schemes

In the field of AI documentation, several approaches have been developed to support transparency, ethics, quality, reproducibility, discoverability, trust and accountability throughout the various stages of data and AI model life cycles. The present clause gives an overview of existing works grouped by the focus the documentation schemes available in the literature. Details are provided in Annex D giving information of what they document, the intended audience for the documentation, the stage of the development life cycle at which the documentation is created and the techniques employed for documenting. Also, their respective strengths, weaknesses, and existing gaps in the context of the EU AI Act are highlighted. The approaches are grouped into:

- **Data-focused documentation approaches (see clause D.1)** which primarily concentrate on the documentation of the datasets used in training, validation and test of AI systems or models. Approaches include strategies to document the creation and use of datasets [i.19], to document the structure, data provenance and social concerns of ML datasets [i.20], to enhance data quality standards by providing a clear and standardized way to describe datasets [i.32], to promote transparency and responsibility in AI dataset usage [i.35], and, to improve accountability in Machine Learning (ML) datasets [i.34].

- **Model-and-method-focused documentation approaches (see clause D.2)** that primarily focus on documenting ML models and methods used within AI systems. Approaches include strategies for documenting the characteristics of trained models, including their performance, intended use cases, and any relevant attributes for which performance may vary [i.40]. Others focus on supporting the robust auditing and evaluation of ML systems through the documentation of both ML models and non-ML components like data acquisition and human-in-the-loop interfaces [i.21].

- **System-focused documentation approaches (see clause D.3)** that focus on documenting the entirety of an AI system, including datasets, models and methods, APIs, and non-AI/ML components that interact as part of the overall AI system. Available solutions offer a broader perspective by providing documentation coverage for both, models and datasets [i.22], and documenting and communicating various aspects of ML systems, including data, models, and decision-making processes [i.33]. The aim of these strategies is to enhance user trust and understanding by providing clear and accessible information about how ML systems work and their potential impacts.

- **Domain specific documentation approaches (see clause D.4)** which primarily target the documentation of datasets, models, methods, and AI systems within a specific domain. Approaches provide strategies to ensure that critical model information is accurately conveyed to the end-users in the healthcare domain [i.38], to focus on structure assessment and the documentation of risks associated with language model applications [i.36], and, to document the Quality of Experience (QoE) of users [i.37].

5.2 Gaps to EU AI Act Requirements

The analysis in clause D.5 demonstrates that while current state-of-the-art AI documentation approaches generally fulfil many of the data-related documentation requirements outlined in the EU AI Act, significant gaps remain. For instance, while Data Cards, DescribeML, Factsheets, and the Dataset Development Life Cycle Documentation Framework provide the most comprehensive coverage of data-related requirements, they often overlook key elements. Specifically, validation procedures and impact assessments (which are critical when personal data is involved) are insufficiently documented. These components are essential to ensure transparency around how data has been validated for accuracy and fairness, and to understand the potential privacy implications of AI systems. Out of the evaluated approaches, only five sufficiently address these crucial data validation and impact assessment requirements, leaving a considerable gap in compliance with the EU AI Act.

Furthermore, a more pronounced shortcoming arises in the documentation of AI system-related requirements mandated by the EU AI Act. These requirements include documenting technical specifications, operational constraints, and system-level risk assessments, which are vital for ensuring the safe and responsible deployment of AI systems. While Factsheets, System Cards, and Model Facts Labels attempt to address most of these system-level needs, the remaining approaches offer minimal coverage. This lack of documentation for system-related aspects severely limits stakeholders' ability to assess the overall safety, performance, and accountability of AI systems. These gaps suggest that existing documentation approaches are predominantly data-centric and fail to provide a comprehensive view of the AI system as a whole, which is crucial for regulatory compliance and trustworthiness.

A similar shortcoming is evident when it comes to control-related documentation requirements. These controls include processes for continuous monitoring, human oversight mechanisms, and safeguards for mitigating risks during operation. The majority of the reviewed AI documentation approaches (see Annex D) either do not address control-related elements or do so only superficially. This omission poses a serious challenge, as these controls are necessary to ensure that AI systems remain compliant throughout their life cycle, particularly in high-risk applications. Without robust documentation on control mechanisms, it becomes difficult to ensure ongoing compliance, manage risks, and facilitate accountability as required by the EU AI Act.

The Assurance Case framework [i.30] (see Annex D) is theoretically suitable to address any documentation requirements, by demanding the respective evidences by any means necessary. Thereby, the provision of evidences relies on any other suitable documentation approaches. Additionally, the argumentation that states why the sub-claims sufficiently imply the main-claim may be highly subjective.

These shortcomings highlight a critical issue: existing AI documentation approaches are fragmented and lack consistency in addressing the full spectrum of requirements outlined in the EU AI Act. While certain approaches focus heavily on data transparency, they fall short in areas related to system architecture, performance monitoring, and control mechanisms, which are equally important for ensuring that AI systems are safe, transparent, and ethical.

The fragmented nature of these approaches points to the urgent need for a more holistic and integrated documentation framework. A unified approach would ensure comprehensive coverage of the Act's requirements, addressing not just the data-related aspects but also the system-level specifications, operational constraints, and control mechanisms necessary for regulatory compliance. Such a framework would enable developers, regulators, and users alike to have a clear, consistent, and complete understanding of AI systems, thus improving transparency, accountability, and trust in AI technologies.

5.3 Summary

Based on the inconsistencies and gaps identified, the development a unified AI documentation scheme is needed that ensures thorough coverage of all documentation requirements mandated by the EU AI Act. Such a unified scheme would streamline documentation processes, provide clarity, and facilitate compliance, ultimately fostering a safer and more accountable AI ecosystem. Taking the European single market as an example - proceeding from the current standardization landscape, operators seeking access to the European single market with AI-related products and services can encounter documentation shortcomings related to conformity assessment. Such shortcomings reflect a lack of guidance on how to properly declare conformity to fulfil the essential requirements for entering a market, e.g. the European single market. Such guidance is of decisive importance for ensuring consistency in fulfilling legislative obligations. For instance, in Europe, the fulfilment of such obligations results in the CE marking as the demonstration of the fulfilment of essential requirements from European legislations. Moreover, for AI systems, documentation is frequently lacking in life cycle-related information, such as updates, modifications, and performance monitoring, which are essential for market surveillance authorities. Especially the insufficient and inconsistent tracking of AI-related quality criteria makes it difficult to assess conformity before as well as after-market access. This provides uncertainty to operators heading to enter a market as well as the authorities' ability to perform thorough inspections and enforce legislative obligations.

6 Approach for Documenting AI-enabled Systems

6.1 Overview

Documentation, in general, supports a large variety of needs and is always to be tailored to specific situations: there is no one-fits-all format or method. This flexibility is particularly crucial for AI systems, as they often operate in dynamic environments and serve diverse stakeholders with varying technical expertise and regulatory requirements. The present document sets out an approach based on the following ideas and concepts.

What is being documented is considered as the *documentation item*. This could include an ML model, an algorithm, evaluation results, datasets, processes, or design decisions (see clause 6.2).

Documentation is created depending on who is creating it and for whom it is intended. This is determined by the *documentation stakeholder* and allows for different perspectives on the AI system or different abilities to understand the content of the documentation (see clause 6.3).

Moreover, documentation should be seen as an ongoing process, i.e. re-activated whenever the system is retrained, updated, or otherwise modified. The *documentation trigger* describes *when* within the system life cycle, documenting should start (see clause 6.4).

The *documentation method* addresses *how* the documentation will be created. This involves selecting suitable methods and formats that best represent the subject, such as textual descriptions, diagrams, or structured templates. Applying established templates and standards facilitates clear, accurate, and consistent documentation, ensuring it remains accessible and reliable for all stakeholders (see clause 6.5).

Finally, the *level and quality of documentation* encompass both the required level of detail, determined by the complexity and risk associated with the system and the quality characteristics needed to make the documentation accurate, complete, and fit for purpose (see clause 6.6).

6.2 Documentation items (what to document)

In the context of AI system documentation, a **documentation item** defines what is being documented to ensure transparency, accountability, and regulatory alignment across the AI system life cycle. Crucially, the documentation item is not the document itself but the subject of documentation, i.e. something that requires formal representation due to its relevance in process or system performance or compliance.

A documentation item should be a distinct workflow, artifact, or component that is part of the engineering, training or operation process of an AI-based system and warrants structured and traceable documentation.

EXAMPLE 1: A neural network model, the training workflow used to develop it, or the deployment pipeline supporting its operation are all documentation items. Each of these may require dedicated documentation that captures relevant information for stakeholders like developers, auditors, and regulators.

Each documentation item should be considered central to enabling understanding, transparency, traceability and demonstrating conformance to standards and regulations. Documentation items are diverse in nature and evolve over the AI system's life cycle. Proper documentation should ensure that stakeholders can understand how the system was developed, validated, deployed, or monitored.

EXAMPLE 2: Documentation items include datasets, data pipelines and workflows, AI models, user interfaces, regulatory compliance artifacts (e.g. audit logs, certifications, or risk assessments, which serve to document compliance-relevant items), as well as environmental conditions and use cases.

Each documentation item should serve stakeholders with a specific set of relevant information, whether they are end users, auditors, developers, or regulatory bodies.

Information elements are granular units of descriptive or operational detail. These elements should define the specific attributes, characteristics, and context necessary to fully understand the documentation item in question. Information elements are the "about" part of documentation - they detail the attributes, properties, and metadata that provide depth and clarity to the record.

EXAMPLE 3: In a model card, individual information elements might represent the following:

- *Intended Purpose:* The primary function and target use cases of the AI model.
- *Data Provenance:* Specifics on the origin of training data, such as the source or collection method.
- *Risk Management Details:* Descriptions of identified risks and the mitigation measures in place.
- *Performance Metrics:* Quantitative measures such as accuracy, F1 score, or robustness under stress conditions.

In high-risk AI systems, as outlined in regulatory frameworks like the European AI Act, information elements extend to cover critical details such as dataset scope, human oversight protocols, and cybersecurity measures. Each element represents a concrete requirement derived from legal texts, ensuring that every documentation item' documentation meets transparency and compliance standards.

Documentation items can be organized into several **high-level categories**, each addressing distinct facets of the AI system:

- **Process Documentation:** Records the life cycle processes including development, testing, and operational procedures.
- **Tools Documentation:** Focuses on the software, libraries, and platforms used throughout the AI life cycle.
- **Data Documentation:** Covers all aspects of the data used in the AI system.
- **Algorithms and Models Documentation:** Concentrates on the core AI properties like model architecture, hyperparameters, algorithms used for training etc.
- **Project and Regulatory Documentation:** Encompasses non-technical records such as requirements specifications, risk and test reports, and compliance files.
- **System Architecture and Environment Documentation:** Describes the technical environment, including hardware, network configurations, and security measures.
- **User Instructions and Interfaces Documentation:** This documentation includes user manuals, interface guides, and training materials that facilitate effective interaction with the system.

The documentation for each documentation item should be composed of several nested information elements, arranged in a logical order. For instance, a datasheet documenting training data for a high-risk AI system might start with an overview (intended purpose and scope) and then drill down into technical specifics such as data provenance, preparation techniques, risk management, and security protocols. This structure not only aids in clarity and ease of access but also supports incremental updates, allowing stakeholders to modify individual elements without having to overhaul the entire document.

In summary, by defining documentation items as comprehensive records built from discrete information elements, organizations can ensure that all critical dimensions - from technical design and development to regulatory compliance and user guidance - are transparently and systematically recorded.

6.3 Documentation stakeholders (roles)

6.3.1 General

Different stakeholders involved in AI development, deployment, and regulation have specific responsibilities that should be supported by transparent, clear documentation. These stakeholders range from those who create and provide AI technologies to those who integrate, use, or are impacted by them.

ISO/IEC 22989 [i.4] offers a framework that defines these roles, outlining the various entities that contribute to AI systems - from developers and producers to customers and regulators. The roles are organized hierarchically in a tree-like structure (see Figure 1) to reflect their relationships and subcategories. This structure helps clarify how broad stakeholder categories break down into more specific roles, which is important for understanding responsibilities and interactions across the AI life cycle.

Each stakeholder's role is associated with distinct documentation requirements to support trustworthiness, from ensuring system accuracy and fairness to verifying compliance with data protection laws like the GDPR. Furthermore, documentation requirements for AI providers, producers, and users emphasize transparency, particularly around data handling, algorithmic decision-making, and system monitoring.

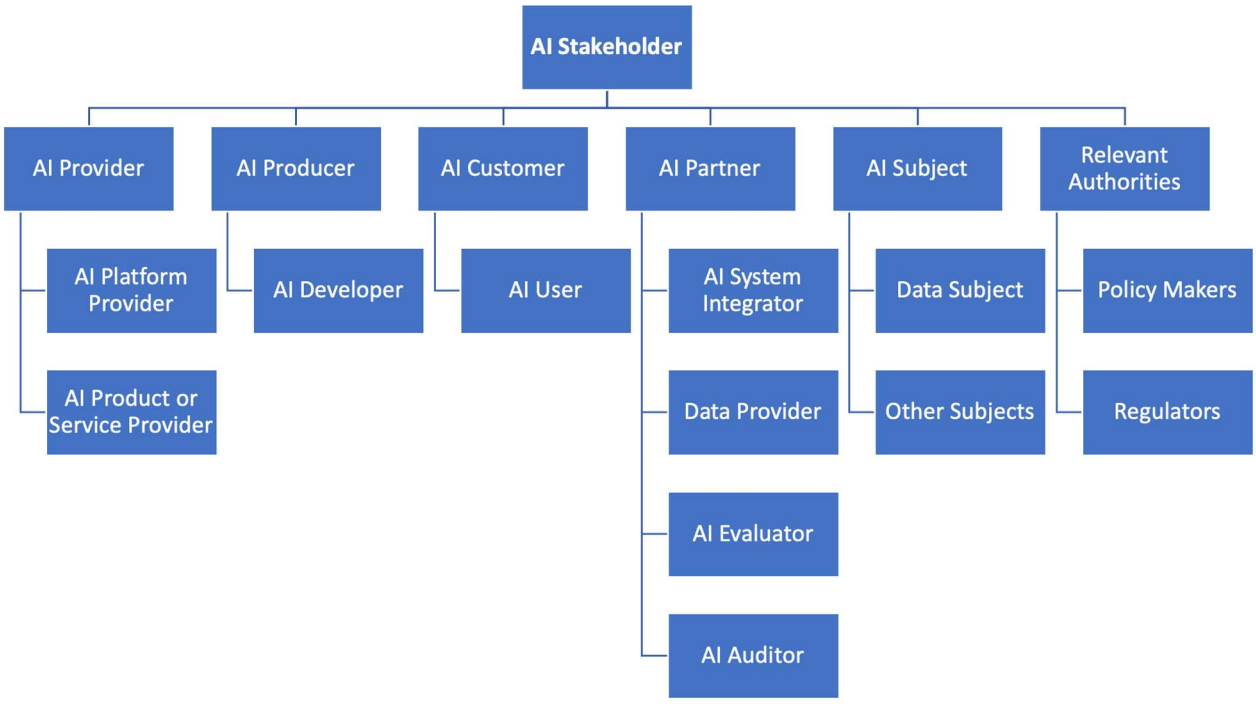

Figure 1: ISO/IEC 22989 [i.4] Stakeholders

Since different stakeholders have different documentation needs/requirements, it is difficult to identify a one-size-fits-all approach, i.e. a single documentation file that meets everyone's needs/requirements at the same time. For this reason, customizing AI documentation to stakeholders is recommended, in line with best practices in technical writing [i.87].

6.3.2 Audience analysis

In ISO/IEC/IEEE 26514 [i.24], *Audience analysis* is the process of determining who will use the information included in the documentation. The standard requires this process to be conducted taking into consideration factors such as users' background, experience, and education, familiarity with technical language, the ways in which they might use the software, their learning stages (e.g. novice, expert), and how often they use the software. Groups of users who share characteristics and needs constitute an audience.

Audience analysis is an important step in planning, writing, and reviewing technical documentation, as it determines the content, structure, and use of the intended information. Consequently, customizing AI documentation on the stakeholders can impact the modality and techniques used for the documentation, as well as the items included in the documentation and how they are presented.

While specific customization depends on the results of the audience analysis, a sketch can be provided of how modalities, techniques, and items can be tailored on the needs of stakeholders by assuming they constitute specific audiences. For each type of stakeholder, it is provided, when possible, a reference to standards representing a starting point for structuring an appropriate documentation.

6.3.3 Stakeholder categories and documentation requirements

6.3.3.1 AI Provider

Documentation Requirements: For AI providers, documentation should ensure transparency about the AI technologies being offered, including detailed descriptions of the AI models, algorithms, and data processing techniques used. Additionally, this documentation should include comprehensive records of testing methods, validation procedures, and ongoing monitoring protocols to maintain system trustworthiness over time. Documentation should also explain how regulatory standards (such as the EU AI Act or GDPR) have been integrated into the system's design and how risk management practices address potential harm or bias. Providers should offer explicit documentation on how updates or modifications are communicated to stakeholders, ensuring continuous compliance. This documentation is crucial for establishing the trustworthiness of the AI solutions, especially when these solutions are integrated into larger systems.

ISO/IEC/IEEE 26514 [i.24] provides an analysis of the requirements for designers and developers of user documentation. It includes both approaches to standardization: a) process standards, which specify the way in which documentation products are to be developed; and b) documentation product standards, which specify the characteristics and functional requirements of the documentation. It is addressed to designers and developers of software user documentation.

EXAMPLE: An AI service provider might need to supply comprehensive documentation on how their model ensures fairness and accuracy, alongside performance metrics and bias mitigation strategies, to reassure customers and partners of the system's reliability and ethical integrity. In this case, the provider might also need to include detailed logs showing compliance with GDPR requirements for handling personal data and records of any third-party audits conducted to verify the system's fairness and transparency. This added level of detail helps reassure customers and regulators that the AI system operates within the defined legal and ethical boundaries.

6.3.3.2 AI Producer

Documentation Requirements: AI producers require highly detailed documentation throughout the development life cycle. This includes technical specifications, design documents, testing protocols, and deployment records. This documentation should cover the entire development pipeline, from data preprocessing techniques to model selection and training processes. Additionally, it should include comprehensive versioning control and change logs that track adjustments made throughout development, ensuring traceability for auditing purposes. Detailed error analysis, stress testing outcomes, and compliance with industry standards should also be documented, along with mechanisms for post-deployment monitoring and system maintenance. Such documentation is vital for internal audits, ensuring compliance with industry standards, and facilitating external evaluations.

Since AI producers manages user documentation, the ISO/IEC/IEEE 26511 [i.27] and ISO/IEC/IEEE 26513 [i.29] standards are relevant in this context. This former supports the interests of software users by driving the realization of consistent, complete, accurate, and usable documentation. It is addressed at managers responsible for the development and production of user documentation. The latter provides documentation requirements for testers and assessors of user documentation.

EXAMPLE: A model designer would need to document the entire model creation process, including data preprocessing techniques, model selection rationale, and validation results, to ensure that the AI system can be thoroughly reviewed for trustworthiness by auditors or evaluators. This documentation should also include details on how the system complies with medical regulations, such as Health Insurance Portability and Accountability Act (HIPAA) in the U.S., and should outline procedures for addressing patient privacy and ensuring the accuracy of diagnosis results. In this case, the producer would need to ensure that any updates to the model, such as retraining on new data, are thoroughly documented and retrievable for future audits.

6.3.3.3 AI Customer

Documentation Requirements: AI customers typically do not create their own documentation but they should ensure that the AI provider has supplied sufficient guidance and trustworthiness assurances. This should include instructions for deployment and use, certification reports, compliance assessments, or summaries of the system's capabilities and limitations to verify that the AI system meets their requirements. Additionally, documentation on how to handle exceptional cases, such as system errors or biased outputs, should be provided to ensure that the customer can implement the system in a controlled and compliant manner. AI customers should deploy the AI system according to the instructions for use supplied by the provider in the technical documentation. Under the AI-Act, this is mandatory for deployers of high-risk AI systems. A list of possible requirements for AI customer documentation can be found in the ISO/IEC/IEEE 26512 [i.28] standard.

EXAMPLE: An AI customer might assess an ISO 9001-like certification from the provider, verifying that it aligns with their quality and safety standards without needing deep technical expertise.

6.3.3.4 AI Partner

Documentation Requirements: For AI partners, the documentation should be detailed and precise to support their specialized tasks. AI auditors, for example, require extensive documentation on AI system design, data integrity, and quality and risk management systems, often in a machine-readable format like JSON, to perform thorough audits. Similarly, AI system integrators need detailed integration manuals and API documentation to ensure seamless incorporation of AI components into larger systems.

EXAMPLE: An AI auditor conducting a fairness audit of an AI recruitment tool would need access to extensive documentation on the system's training data, bias detection methods, and performance outcomes across different demographic groups. In addition to these details, the auditor may also require documented proof of external audits, certification from third-party evaluators, and records of any bias remediation actions taken. These materials help ensure that the system not only meets ethical standards but also operates within the legal frameworks for fair hiring practices.

6.3.3.5 AI Subject

For instance, a data subject might require documentation that explains how their personal data is processed by an AI system, including information on data retention policies, consent mechanisms, and privacy safeguards, presented in clear, non-technical language. In this case, it is also essential for the documentation to explain how users can revoke their consent for data usage, how long their data will be retained, and the specific measures the system uses to protect their privacy. This level of transparency builds user trust and aligns with privacy laws like GDPR.

6.3.3.6 Relevant Authorities

Documentation Requirements: Relevant authorities, such as regulators and notified bodies, do not generate their own documentation but instead receive and review documentation to ensure regulatory compliance and policy enforcement. This includes reports on AI system transparency, accountability measures, and ethical considerations submitted by AI providers. Documentation should also include detailed records of transparency measures, accountability protocols, and data protection strategies, particularly for high-risk AI systems. Regulators may also request audits of the AI system's ethical frameworks, such as bias detection and mitigation procedures, and evidence of periodic reviews to ensure continuous compliance.

EXAMPLE: A notified body need access to detailed documentation, including performance metrics, algorithmic transparency reports, and conformity assessment records, to verify that high-risk AI systems meet EU requirements before being placed on the market. Similarly, a regulator might review documentation that outlines how an AI system complies with applicable legislation, such as GDPR, including records of data protection impact assessments, to ensure that the system adheres to societal and ethical norms. In this context, the documentation should also outline any corrective actions taken in response to compliance failures, including updates to the AI system or adjustments to its deployment process. This helps regulators ensure that the system remains compliant over time and that any issues are promptly addressed.

6.4 AI-system life cycle

An AI system life cycle encompasses the comprehensive series of stages involved in the creation, deployment, and maintenance of an AI-based system. This life cycle helps in structuring the development process to ensure effective, reliable, and ethical AI-based solutions. The life cycle typically includes several phases, each critical to the success of the AI-based system.

There are various AI-system life cycle descriptions that have been developed for different purposes, and accordingly, emphasize different aspects and vary in their level of detail. One of the most well-known might be the Cross-Industry Standard Process for Data Mining (CRISP-DM) [i.43] that has become the de facto standard process model for data mining, analytics, and machine learning projects. The focus is on providing a clearly defined, process-orientated framework that enables companies to carry out data mining projects efficiently and effectively. At the same time, however, it lacks the perspective of other stakeholders, like the persons affected.

ISO/IEC 22989 [i.4], clause 6, describes a life-cycle process that covers all major aspects from the 'inception stage' (requirements engineering) to the 'retirement' of a system. As a draw-back, it lacks granularity that might help to inspect valuable aspects regarding transparency.

ISO/IEC TR 29119-11 [i.23], Figure A.2, describes a detailed machine learning workflow that covers all relevant aspects but omits some crucial feedback loops.

To be as generically applicable as possible, an adaption of a generic AI-system life cycle is inspected for the present document, called the long chain of responsibilities [i.42]. This concept emphasizes the necessity of considering a broad spectrum of responsibilities across different stages of AI system development and deployment, from the conceptualization and design phases through to their real-world applications and impacts. It is very similar to the machine learning workflow proposed by ISO/IEC TR 29119-11 [i.23] but uses the common workflow terminology as ISO/IEC 22989 [i.4] does. It is major drawback for the purpose of the present document is lack of taking the requirements engineering phase into account. Therefore, the long chain of responsibility proposed by [i.42] has been extended by [i.39], explicitly to analyse possible mechanisms that provide transparency. The life cycle model that is outlined in the present document distinguishes between a System Life Cycle, a Data Life Cycle, and a Model Life Cycle as depicted in Figure 2 and described below.

System Life Cycle: For the system life cycle, the phases are differentiated as **Inception, Analysis & Design**, **Implementation & Integration**, and **Deployment & Operation**. These phases further encompass phases from the data and model life cycle. For documentation purposes, special attention is given to *KPI and Requirements Gathering* Data life cycle and Model life cycle:

- **KPI and Requirements Gathering:** This phase is the first step in the system development process, involving a systematic approach to identify, specify, and manage both requirements and Key Performance Indicators (KPIs). The aim is to understand the needs of customers, legal obligations, and other relevant factors. The outcome of this process is a set of documents that detail various requirements for different stakeholders, specifying what the AI-based system is expected to achieve. These documents also include informal target criteria, benefit and risk assessments, as well as clearly defined KPIs, which serve as measurable benchmarks to assess the system's performance against its intended goals.

Data Life Cycle: Refers to the various stages that a dataset goes through, from its initial collection to its eventual deletion. Below, the different stages in the data life cycle are described:

- **Data Collection & Extraction:** Data can be freshly collected from various sources such as sensors, databases and surveys. However, data could be extracted or created through processes like simulations, experiments or computational models. The process of data collection may underly legal restraints and may also need to satisfy specific requirements. This phase of the data life cycle falls under the Inception, Analysis & Design phase of the system life cycle.

- **Data Preparation & Processing:** The construction of a data set consists of multiple phases that depend on the specific data at hand and task to be performed. This phase falls under the Implementation & Integration phase of the system life cycle. In context of a classification task, the following preparation and processing can be performed:

    - Data labelling: the output variable to be predicted needs to be identified if it is already part of the data or labelled by hand if it is not part of the data.

    - Data cleaning: redundant information in data as well erroneous or missing values needs to be dealt with.

- Data transformation: parts of the data may need to be transformed from one format or structure to another.
- Data integration: data collected or extracted from multiple sources will have to be combined to create a unified dataset.
- Data storage: storing the processed data in a way that ensures it is secure, organized and accessible.

- **Data Monitoring & Maintenance:** it is important to ensure that data remains accurate, up-to-data and usable over time. This phase falls under the Deployment & Operation phase of the system life cycle. This may be achieved through the following activities:
  - Data updates: regularly updating the dataset with new data or correcting outdated information.
  - Data quality monitoring: continuously checking data for issues such as errors, inconsistencies, degradation or drifts.
  - Data deletion and destruction: permanently removing data that is no longer needed or should be deleted for compliance or privacy reasons.

Model Life Cycle: Refers to the stages an AI or machine learning model goes through, from its initial design through to its eventual retirement. Below, the different stages of the model life cycle are described:

- **Experimentation:** There is a plethora of choices to be made when settling for a machine learning method. Various model types like an artificial neural network, a support vector machine, a decision tree, etc. might be viable choices. Each type needs to be specified in terms of hyperparameters (in the context of ANN, for example, the number of layers, the number of neurons per layer, the activation function(s), the learning rate, the batch size and the stopping criterion). There are various software packages and tools, that may hide some of the actual complexity behind such approaches and set parameters to default values. One can either choose a method for which the data is suitable and/or which requires little preprocessing or design the details of the preprocessing with the chosen procedure in mind. This phase of the model life cycle falls under the Inception, Analysis & Design phase of the system life cycle.

- **Model Training & Model Evaluation:** involves both training the model and evaluating its performance on the training set to assess its ability to learn patterns. This phase falls under the Implementation & Integration phase of the system life cycle. Below are some activities undertaken in this stage:
  - Training the model: involves feeding training data to the model to learn patterns and relationships.
  - Hyperparameter tuning: optimizing hyperparameters (e.g. learning rate, number of layers) to maximize performance.
  - In-training evaluation: assessing model performance on the training data by measuring metrics like accuracy, loss or error rates.

- **Model Validation:** has to do with testing the model on a separate validation or test dataset to ensure it generalizes well to unseen data. This phase falls under the Implementation & Integration phase of the system life cycle. Some activities involved in model validation include:
  - Cross-validation: techniques like k-fold cross-validation ensure that the model doesn't overfit the training data and works well with new data.
  - Bias and fairness checks: examining the model for potential biases in its predictions to ensure fair outcomes, especially in critical applications.

- **Model Integration & Deployment:** in this stage, the trained model is deployed into production environment and integrated into real-world systems to perform an intended task. This phase falls under the Implementation & Integration phase of the system life cycle. Below are some activities undertaken at this stage of the life cycle:
  - Infrastructure setup: ensuring computational resources are adequate for production.
  - Security: implementing robust security protocols to protect the model and the underlying data.

- **Model Monitoring & Maintenance**: involves continuously tracking the model's performance and maintaining its quality over time. This phase falls under the Deployment & Operation phase of the system life cycle. The following activities are involved in this:
  - Performance monitoring: tracking real-time performance metrics to detect issues like data drift.
  - Error handling: managing and addressing any performance drops or issues detected post-deployment.
  - Model retraining: regularly updating models with new data to keep it relevant.

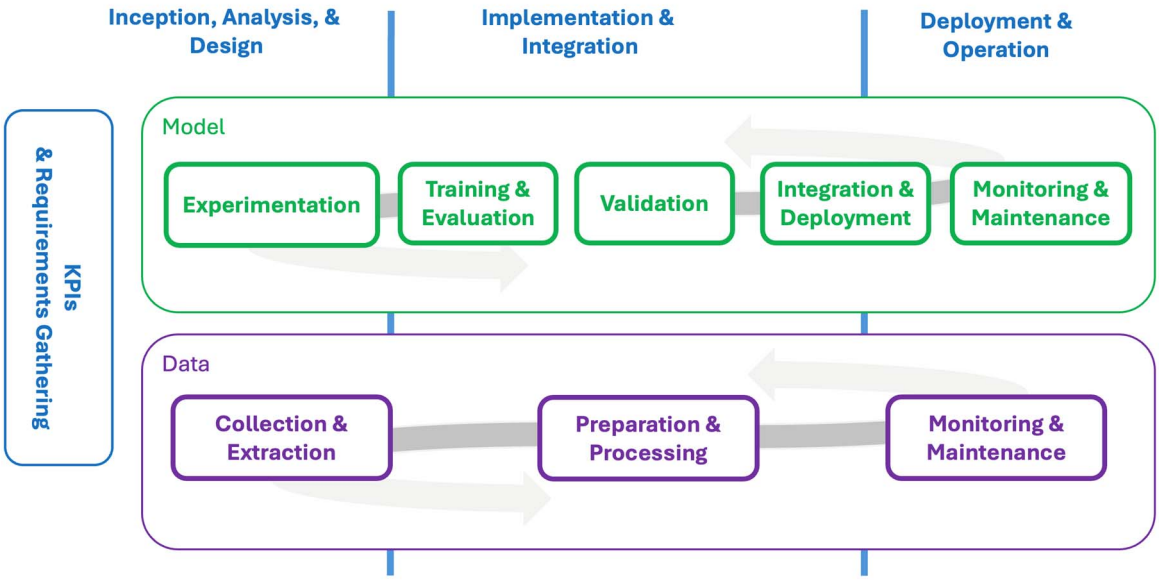

Figure 2: High level overview on system, model and data life cycle phases

While procedures for each phase can and should be documented once they are decided upon, they may be adapted during execution, which makes it important to revise the documentation after its execution, in case of recurrent procedures, each time. In case of continuous procedures, the documentation needs to be revised on a regular basis.

6.5 Documentation Techniques for Effective Information Management

6.5.1 General

An **AI documentation approach** serves as a high-level strategy or framework that outlines how documentation is created, organized, and maintained. It defines the **what** (the content and scope) and **why** (the purpose and goals) of documentation. In contrast, a **documentation technique** refers to the specific methods or tools used to create, organize, or present documentation. Techniques represent the **how** of documentation - they are the practical steps, formats, or tools employed to bring the broader documentation approach to life.

Because of this complementary relationship, an AI documentation approach frequently incorporates one or more documentation techniques to achieve its objectives. For example, a high-level approach like **Model Cards** (used for documenting machine learning models) might employ techniques such as **questionnaires**, **templates**, and **visual documentation** to implement the strategy effectively. In this way, the approach provides the overarching framework, while the techniques offer the practical means to execute it.

This clause provides a comprehensive overview of key documentation techniques, their advantages, and challenges, helping teams choose the right tools and methods to create clear, accessible, and effective AI documentation.

6.5.2 (Motivation and) Overview

A documentation technique is a combination of a specific technique to represent information (e.g. text) and a specific format (e.g. a list). Documentation techniques are the foundation of specific documentation approaches and methodologies (e.g. Model Cards). Each approach is based on at least one documentation technique. While the documentation item determines which documentation techniques might be applicable, the fitting documentation approaches can be chosen to address the specific needs of stakeholder groups.

This clause provides an overview of common modalities and formats, as well as known documentation approaches, and how they are related to the different stakeholders, documentation items, and regulatory requirements. The relationships can serve as a guidance on choosing suitable options tailored to specific needs. The relationships also provide an overview of compatible and complementary approaches. If one approach covers only a subset of the documentation items, other suitable approaches can be identified to complement it in order to cover the remaining items. Additionally, the relationships help identify which items may need to be requested in addition to those provided by a supplier relying on a given approach. This ensures that specific stakeholder needs or regulatory requirements are fully accommodated. The relationships can also indicate which items can be represented in different approaches in case the documentation needs to be converted or integrated from one approach into another. Existing standards that can be considered with regard to the modalities, approaches, stakeholders, as well as quality aspects, are summarized as well.

Documentation techniques enable to depict both, how quality requirements and legal obligations for AI systems have been met. In research, legislative practice, and organizational collaboration, documentation techniques can be differentiated with regard to text, image and interactive paradigms. On the one hand, text documentation ensures that the architecture and processes of the system are transparent and compliant with industry standards, supporting consistent performance. Datasheets for data sets, on the other hand, verify the quality of the data, e.g. to ensure that a data set is accurate, representative and free from unwanted bias, which is critical for compliance with ethical and legal standards. Additionally, process flowcharts and diagrams increase system clarity, make performance issues assessable and ensure continuous quality improvement.

6.5.3 Questionnaires

Questionnaires are structured sets of questions designed to gather specific information in a systematic way. They are often used to collect metadata, feedback, or details about datasets, models, or processes.

- Advantages:
  - Ensures consistency in data collection.
  - Easy to distribute and analyse.
  - Useful for large teams or standardized processes.
- Challenges:
  - Limited flexibility in capturing nuanced or domain-specific details.
  - May require follow-up for clarification.

6.5.4 Information Sheets

Information sheets are static documents that provide detailed information in a narrative or report-like style. They are often used to communicate key details about a system, model, or dataset.

- Advantages:
  - Comprehensive and detailed.
  - Highly customizable for specific audiences.
  - Useful for regulatory compliance and formal reporting.
- Challenges:
  - Static nature limits real-time updates.

- Can become lengthy and difficult to maintain.

6.5.5 Checklists

Checklists are lists of items, tasks, or requirements that need to be completed or verified. They ensure consistency and completeness in processes.

- Advantages:
  - Simple and easy to use.
  - Ensures no steps are missed.
  - Useful for compliance and quality assurance.
- Challenges:
  - May oversimplify complex processes.
  - Requires regular updates to remain relevant.

6.5.5 Templates

Templates are predefined structures or formats for documenting information. They ensure consistency across documents and make it easier to create new documentation.

- Advantages:
  - Saves time and effort.
  - Ensures uniformity across documents.
  - Easy to customize for different use cases.
- Challenges:
  - May not fit all documentation needs.
  - Requires initial setup and maintenance.

6.5.6 White Papers

White papers are authoritative reports that provide in-depth information on a specific topic, often used to explain methodology, results, and implications.

- Advantages:
  - Highly detailed and formal.
  - Useful for communicating complex ideas to a technical audience.
  - Builds credibility and authority.
- Challenges:
  - Time-consuming to produce.
  - May not be accessible to non-technical stakeholders.

6.5.7 Knowledge Graphs

Knowledge graphs are network representations of information that show relationships between different entities. They help in understanding complex systems and their interconnections.

- Advantages:
  - Provides a holistic view of complex systems.
  - Can be queried programmatically for insights.
  - Useful for organizing and visualizing relationships.
- Challenges:
  - Requires expertise in graph theory and knowledge representation.
  - Possibly oversized for simple systems.

6.5.8 Visual Techniques (Diagrams, Flowcharts, Infographics)

Visual documentation uses elements like diagrams, flowcharts, and infographics to communicate complex information in an intuitive way.

- Advantages:
  - Simplifies complex systems and processes.
  - Improves accessibility for non-expert stakeholders.
  - Enhances understanding through visual representation.
- Challenges:
  - May oversimplify details.
  - Requires design skills to create effective visuals.

6.5.9 Interactive Techniques

Documentation that allows users to interact with the content, such as running code, exploring data, or navigating through dynamic elements.

- Advantages:
  - Provides hands-on learning experiences.
  - Encourages exploration and experimentation.
  - Supports real-time updates and collaboration.
- Challenges:
  - Requires technical infrastructure and expertise.
  - May not be accessible to non-technical users.

6.5.10 Domain-Specific Language (DSL)

A specialized programming or markup language designed for a particular application domain. Used to create structured, machine-readable documentation.

- Advantages:
  - Ensures standardization and precision.

- Easy to integrate into automated pipelines.
- Tailored to the specific needs of a domain.

- Challenges:
  - Requires domain expertise to create and understand.
  - Limited usability for non-technical stakeholders.

6.5.11 Summary

A **documentation approach** and **documentation techniques** work hand in hand. The approach defines the overall strategy and goals, while the techniques provide the practical tools and methods to implement that strategy. For example, a **Model Card** (approach) might use **questionnaires**, **templates**, and **visual documentation** (techniques) to achieve its goal of providing transparent and standardized documentation for a machine learning model.

By understanding this relationship, teams can effectively combine high-level strategies with practical tools to create documentation that is both comprehensive and accessible.

In Tables 1 and 2, these documentation techniques are mapped to the existing AI documentation approaches as listed in Annex D. The techniques that are best aligned with the specific needs and responsibilities of the relevant stakeholders are proposed.

Table 1: Mapping of Documentation Techniques to AI Documentation Approaches

Documentation Approaches	Documentation Technique/Format
Datasheet for Datasets (see clause D.4.3)	Questionnaire-Based
Model Facts Label (see clause D.4.1) Model Cards (see clause D.2.1) Method Card (see clause D.2.2) Risk Cards (see clause D.4.2) FactSheets (see clause D.3.1)	Information Sheet (Static Document)
Dataset Nutrition Lable (see clause D.1.3) Data Cards (see clause D.1.4)	Interactive Techniques
DescribeML (see clause D.1.2)	Domain-Specific Language
System Cards (see clause D.3.2)	Visual Techniques

Table 2: Documentation Techniques Aligned with Stakeholder Needs and Responsibilities

Stakeholder	Suitable Documentation Technique/Format	Reason for Mapping
AI Provider	Interactive Techniques Visual Techniques	**Reason:** AI providers need to ensure transparency for a wide range of stakeholders, including customers, regulators and partners. Web-based and Google Docs formats allow for real-time updates, collaboration, and version control, which are critical for maintaining compliance with evolving regulations like the EU AI Act and GDPR. Also, Visual formats simplify complex technical details for non-technical audiences. **Example:** An AI provider offering a facial recognition system could use a web-based dashboard to document model performance metrics, bias mitigation strategies, and GDPR compliance. Infographics could summarize how the system handles data privacy and user consent.
AI Producer	Domain Specific Language Interactive Techniques Visual Techniques	**Reason:** AI producers require highly detailed and technical documentation to track the development life cycle, including design, testing, and deployment. DSL ensures precision in documenting technical specifications, while web-based formats support traceabilty and version control. Visual formats help communicate testing outcomes and compliance records to internal and external auditors. **Example:** A producer developing a medical AI system could use DSL to document data preprocessing techniques, model selection and validation results. Web-based formats could be used to document updates and retraining processes, while infographics could be used to illustrate preprocessing pipelines and model architecture for clarity during audits.
AI Customer	Information Sheet Interactive Techniques Visual Techniques	**Reason:** AI customer needs clear, concise and user-friendly documentation to understand how to integrate and operate AI systems within their workflows. Static documents and infographics provide easy-to-digest summaries of system capabilities, limitations, and compliance certifications. Web-based formats ensure access to the latest updates and operational guides. **Example:** A customer using an AI-powered recruitment tool could receive a static document summarizing the system's fairness metrics and compliance with hiring regulations. A web-based portal could provide step-by-step instructions for integrating the tool in the HR systems.
AI Partner	Domain Specific Language Questionnaire-based	**Reason:** AI partners, such as system integrators and auditors, require detailed and structured documentation to perform their specialized tasks. DSL ensures precision in integration manuals and API documentation, while questionnaires help auditors gather specific information for compliance assessments. **Example:** An AI auditor evaluating a recruitment tool could use a questionnaire to document details on training data, bias detection methods, and performance outcomes.
AI Subject	Information Sheet Visual Techniques	**Reason:** AI subjects, such as data subjects or end-users, need transparent and accessible documentation to understand how their data is used and the implications of AI-driven decisions. Static documents and infographics simplify complex concepts and ensure compliance with transparency requirements under regulations like GDPR. **Example:** A data subject using a healthcare AI app could receive a static document explaining how their data is processed, their rights to opt-out, and the measures in place to protect their privacy. Infographics could illustrate the data life cycle and anonymization techniques.
Relevant Authorities	Interactive Techniques Visual Techniques	**Reason:** Regulators and policymakers require comprehensive and accessible documentation to verify compliance with legal and ethical standards. Web-based and Google Docs formats facilitate the submission, review, and updating of compliance reports. Visual formats help present transparency measures, accountability protocols, and ethical considerations in a clear and concise manner. **Example:** A notified body assessing a high-risk AI system could review web-based documentation detailing performance metrics, algorithmic transparency, and conformity assessments. Infographics could summarize bias mitigation strategies and data protection measures.

6.6 Quality aspects of documentation

6.6.1 General

The information contained in the technical documentation should follow established principles of information quality. ISO/IEC/IEEE 26514 [i.24], clause 7, identifies six key principles: correctness, consistency, comprehensibility, conciseness, minimalism, and accessibility.

6.6.2 Correctness

The information provided in the technical documentation should accurately reflect the AI systems' actions and expected results for the specific version being documented. This includes details on functionalities, limitations, and behaviour. Any updates or changes made to the AI system (e.g. new features, bug fixes) should be reflected promptly and correctly in the corresponding documentation.

6.6.3 Consistency

The technical documentation should maintain a consistent structure and layout. Consistency applies to all elements including screens, pages, text formatting (headings, spacing, fonts), graphics, icons, colours, signal words, and audio-visual elements. Additionally, consistent terminology should be used for user interface elements, data, fields, tasks, pages, and processes within the documentation and the AI system itself.

6.6.4 Comprehensibility

The technical documentation should be easily understood by all relevant stakeholders. Information should be readily understood by the least experienced stakeholder within the expected audience. This is particularly important when serving a diverse user base with varying levels of experience, skills, training and knowledge. Terminology selection plays a vital role in achieving comprehensibility. Technical documentation should opt for terms commonly used within the stakeholder's environment or the application domain. For instance, it is preferable that documentation for a medical AI system employs medical terminology readily understood by healthcare professionals, rather than complex technical terms related to the underlying algorithms. Usability testing can be employed to validate the comprehensibility of the documentation.

6.6.5 Conciseness

Information within the technical documentation should be presented concisely, both in terms of format and media, avoiding unnecessary repetition or duplication. While repetition can be a useful tool for educational purposes, technical documentation should prioritize clarity and efficiency.

6.6.6 Minimalism

The technical documentation should be minimal, containing only essential information needed for stakeholders to understand concepts, perform tasks, and troubleshoot issues. Technical documentation should avoid including content that is not strictly necessary for accomplishing these objectives. A minimalist approach ensures stakeholders are not overwhelmed by extraneous information and can focus on the core functionalities of the AI system. At the same time, however, the minimalism should not compromise the completeness of the technical documentation.

6.6.7 Accessibility

The technical documentation should be accessible to all expected stakeholder groups, considering factors like language, format, and accessibility needs and regardless of their abilities or environments. This includes ensuring technical availability, legibility, and findability of the information. For example, documentation for visually impaired stakeholders may require alternative formats such as screen reader compatible text. Websites and mobile applications containing the documentation should adhere to accessibility guidelines outlined in relevant standards and established good practices. Guidelines concerning accessibility are for example provided in ETSI EG 204 061 [i.25]. The principles of information quality may need to be implemented differently depending on the target audience, the risk level of the AI application, and the domain. In addition, trade-offs between principles should be considered, e.g. conciseness vs. comprehensibility, correctness vs. accessibility, comprehensibility vs minimalism. Finally, EN 301 549 [i.26] provides some hints on documenting accessibility and compatibility features, as well as making documentation accessible (clause 12.1 in particular), relationships to Directive 2016/2102 on Web Accessibility [i.51], and assessment criteria for determining conformance.

6.6.8 Systematic understanding

A systematic understanding refers to a structured and comprehensive approach to comprehending complex systems, processes, or subjects. In the context of AI systems, it involves an organized knowledge of how various components such as data, algorithms, and the corresponding infrastructure is in interaction.

6.7 Documentation Approach

To derive a structured documentation approach, the present document describes the basic understandings and prerequisite considerations for this, as described in clause 4, 6 and 7. Already existing approaches are discussed in clause 5.

From this foundation the following structured documentation approach is compiled in three main steps (see also Figure 3 for a visualization):

Step 1: Understand and identify the purpose of the documentation artifacts

In this step an understanding of the motivation and purpose for the documentation activity (see clause 4) is developed and defined. Especially the applicable requirements from regulatory obligations (see clause 7.3 with reference to the EU AI Act) should be selected.

Step 2: Identify the selected documentation aspects per document

During this step one or more documents (documentation artifacts) should be identified. To structure this identification process, several aspects of the intended document(s) should be selected. Each of the following aspects depends on the identified documentation purpose of step 1:

- Identification of the documentation item (i.e. the subject of documentation) - see clause 6.2.
- Identification of the documentation stakeholders (audience, authors, involved) - see clause 6.3.
- Identification of the phase(s) within the AI-system life cycle when the document is to be created/maintained - see clause 6.4.
- Identification of the documentation technique(s) which serves best the intended purpose for the identified stakeholders - see clause 6.5.

The above listed aspects can be considered specific for each identified document (documentation artifact) and furthermore will have cross-dependencies among each other. E.g. the identified audience (stakeholders) may suggest particular documentation techniques fitting to their skills of understanding, or the type of document in focus can only be created within a specific phase of the AI-life cycle.

Step 3: Identify the document contents (information elements) and create/assemble the document

In this step each of the identified document is detailed with reference to its contents (information elements, see clause 6.2) and then finally created or compiled. To support the process of deriving the documents contents, the following activities are suggested to follow:

- For high-risk AI systems and the need to comply with the EU AI Act: Consideration of recommended documentation approaches as described within clauses 7.3.2 to 7.3.8.
- Selection of an already existing documentation scheme (see clause 5.1 and Annex D), which fit to the defined documentation aspects of step 2.
- For technical documentation, required by EU AI Act for high-risk AI systems (see clause 7.3.8): Selection of applicable documentation templates.
- For other documentations: Identification of the information elements according to the defined documentation aspects of step 2.
- Creation of the documents by describing or filling-in the information elements.

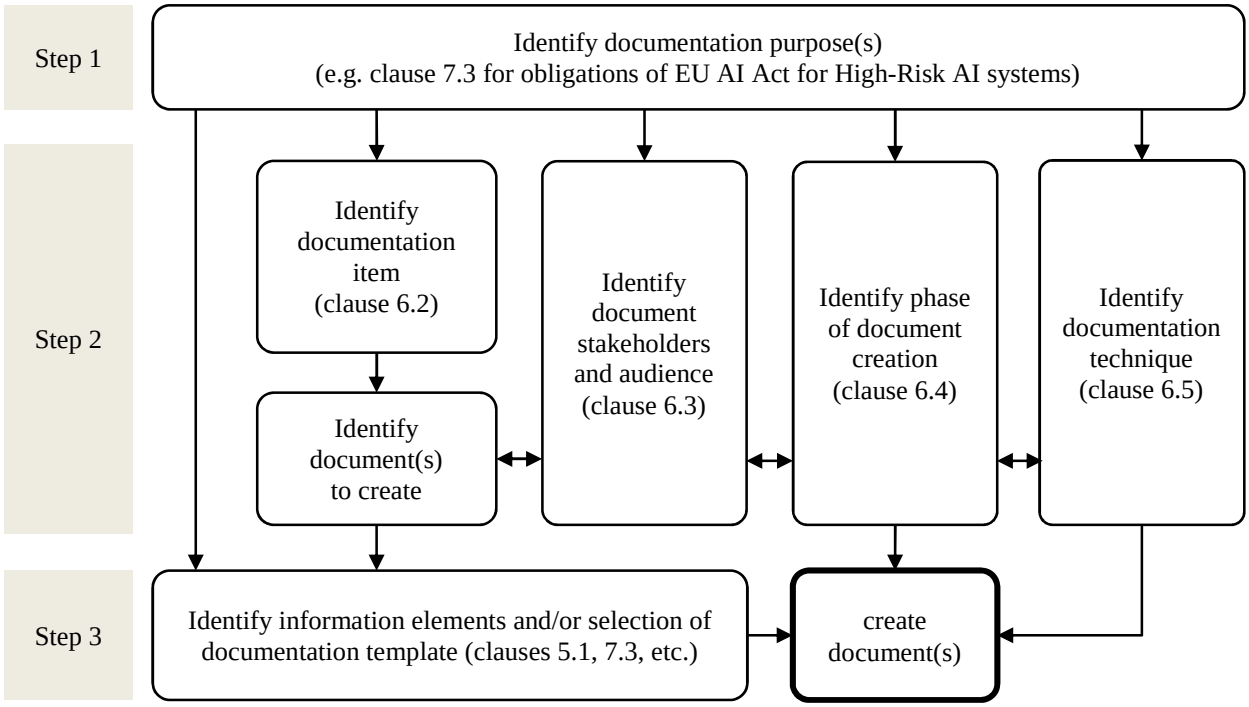

Figure 3: Documentation approach depicted as structured workflow

7 Guidance for EU AI Act Compliant Documentation

7.1 Introduction to the EU AI Act

The EU Artificial Intelligence Act (AI Act), officially Regulation (EU) 2024/1689 [i.2] , is a comprehensive regulatory framework aimed at ensuring **trustworthy**, **human-centered AI** in Europe. Its goal is to protect health, safety, fundamental rights, and EU values while encouraging AI innovation and implementation. The AI Act applies to any AI system marketed or used within the EU, regardless of the provider's geographical location. It adopts a risk-based approach to regulation, classifying AI systems into four categories: prohibited, high-risk, limited-risk, and minimal-risk. These are often illustrated as a pyramid (see Figure 4), where regulatory obligations increase with risk severity.

At the top of the pyramid are **prohibited AI practices (Art. 5)**, which are banned outright due to their inherent threat to fundamental rights or safety. These include systems for real-time biometric identification in public by law enforcement (with narrow exceptions), social scoring, and manipulative or exploitative AI targeting vulnerable groups. Such systems cannot be placed on the market under any conditions. **High-risk AI systems (Chapter III)** form the core of the regulatory framework. These include AI used in critical domains such as law enforcement, critical infrastructure, employment, education, and health. The EU AI Act requires that providers of high-risk systems comply with stringent obligations, including conformity assessment and comprehensive documentation. These systems are the focus of the documentation guidance in this chapter. Below that are AI systems subject to **transparency obligations (Art. 50)**, such as chatbots, deepfake generators, or biometric categorization tools. While not high-risk, the EU AI Act requires to inform users about their AI nature to ensure minimal transparency. These fall under the **limited-risk** tier. At the base of the pyramid are *minimal-risk* systems, including most AI used for personal, recreational, or low-impact applications. These systems are not regulated under the Act, though voluntary codes of conduct and good documentation practices are encouraged. Obligations for **General Purpose AI (GPAI)** models, especially those with **systemic risk (Art. 51)**, are addressed in a dedicated subsection, in line with Chapter V of the Act. For a concise overview of the foundational pillars and operationalization of Trustworthy AI in alignment with the EU AI Act, see Annex B.

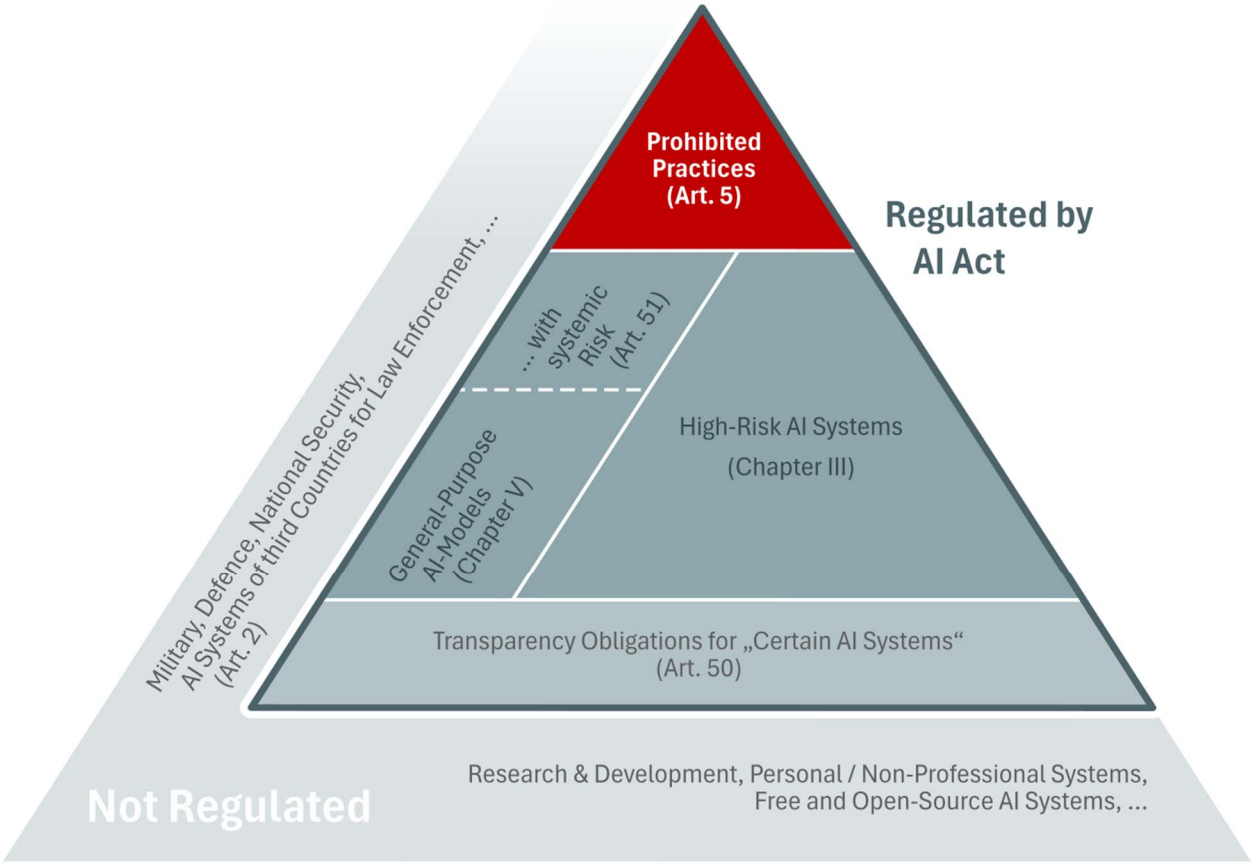

Figure 4: EU AI Risk Pyramide

The predominant regulatory responsibilities under the AI Act are imposed on **providers**, entities that develop or market AI systems, especially concerning high-risk AI systems. This encompasses providers based outside the EU when their systems or outputs are utilized within the EU. Although **deployers** (professional users) have responsibilities, these are more limited in scope. The AI Act delineates the necessary compliance components in legal terminology, but it fails to provide recommendations on how requirements should be documented.

This clause examines the responsibilities of **high-risk AI system providers**, offering a comprehensive analysis of the documentation requirements mandated by Art. 9 to 15. The present document offers practical guidance for organizing documentation in compliance with international standards (ISO/IEC 22989 [i.4], ISO/IEC 24028 [i.1] and ISO/IEC 42001 [i.10]) and incorporates prominent documentation techniques and approaches evaluated in the present document. The objective is to assist providers in generating thorough, verifiable proof of compliance.

7.2 Mapping of EU AI Act Stakeholder

The EU AI Act introduces a legal framework that closely maps to the roles given in clause 6.3, assigning specific responsibilities to each stakeholder to ensure AI systems meet safety, transparency, and ethical standards. Both the ISO and the EU AI Act frameworks aim to clarify who is accountable for different aspects of AI systems' development and usage, promoting trust through detailed record-keeping and compliance documentation.

While ISO/IEC 22989 [i.4] focuses on functional roles (who does what in practice), the AI Act focuses on legal accountability (who is responsible under the law). Thus, some EU AI Act roles, such as **provider**, map to multiple ISO roles, depending on whether the stakeholder is creating, modifying, integrating, or deploying the AI system. Regulatory bodies under the AI Act (**market surveillance authorities**, **notified bodies**, etc.) are clearly represented in the ISO under "Regulators" and "Policy Makers."

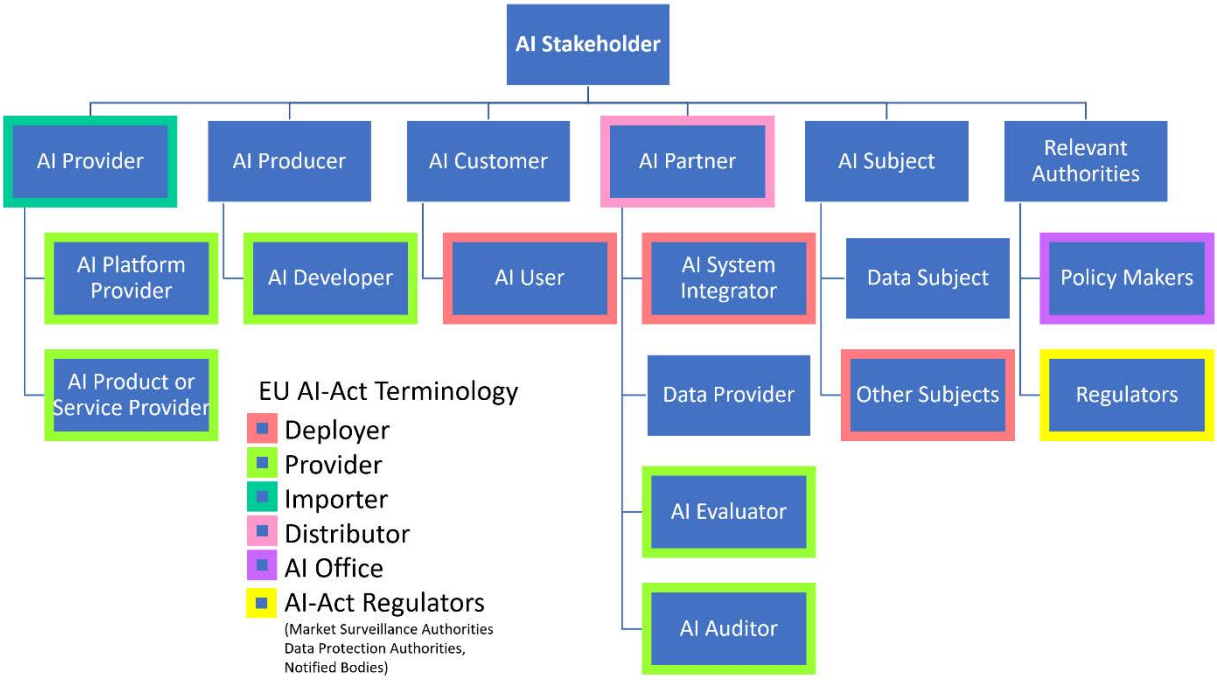

Figure 5: Mapping of ISO/IEC 22989 [i.4] Stakeholders to EU AI Act Roles

Figure 5 illustrates how the stakeholder roles defined in the ISO/IEC 22989 [i.4] standard correspond to the roles established under the EU AI Act. It provides a visual mapping of these stakeholders, showing overlaps and distinctions, and uses colour coding to highlight specific responsibilities, such as "Providers," "Distributors," "Deployers", and "Relevant Authorities." This visual helps clarify how the ISO standard and EU AI Act terminology align to support accountability across the entire AI life cycle. Table 3 provides additional explanations regarding the mapping.

Table 3: Mapping of EU AI Act Roles to ISO/IEC 22989 [i.4] Stakeholders

EU AI Act Roles	Corresponding ISO/IEC 22989 [i.4] Stakeholder(s)	Explanation
Provider	• AI Provider • AI Producer • AI Partner (Auditor/Evaluator, when ensuring conformity) • AI Customer (in internal deployment scenarios)	The entity placing the AI system on the market or putting it into service under their name. The ISO *AI Provider* is the most direct match. If the system is developed in-house, the *AI Producer* or *AI Custome*r can also act as the Provider.
Importer	• AI Provider (when importing and assuming compliance responsibilities)	If the importer places the system on the EU market under their own name or brand, they functionally become an *AI Provider* under ISO.
Distributor	• AI Partner • AI Provider (if distributing under their own name or modifying the system)	If they distribute unchanged systems, they act more as a commercial partner. If they modify or rebrand, they align with the ISO *AI Provider*.
Authorized Representative	• AI Partner	Represents non-EU Providers for regulatory compliance. Acts as an intermediary stakeholder in ISO but typically supports Provider obligations.
Deployer	• AI Customer • AI Producer • AI Partner • (e.g. system integrator)	Entities that use AI systems under their authority. In ISO, *AI Customer* is the closest equivalent. *AI Producer* or *AI Partner* may also deploy systems internally or as part of integration.
Affected Person	• AI Subject	Individuals impacted by the AI system's outputs or decisions. This is a direct mapping to the ISO's *AI Subject*.
Market Surveillance Authority	• Regulator (ISO)	Ensures marketplace compliance. Directly aligns with the ISO role of *Regulator*.
Data Protection Authority	• Regulator (ISO)	Oversees data governance and privacy compliance, particularly for systems processing personal data.
Notified Body	• AI Partner (Evaluator, Auditor) • Regulator • (in conformity roles)	Performs third-party conformity assessments under the EU AI Act. ISO refers to these stakeholders as either evaluators (*AI Partner*) or part of the regulatory oversight framework.
European Commission / AI Office	• Policy Maker	Coordinates the implementation and governance of the EU AI Act across the EU.Directly matches ISO's *Policy Maker*.

NOTE: Although ISO roles are used almost throughout the present document, the EU AI Act nomenclature is used in this clause to simplify the reference to the AI Act.

7.3 Documentation Guidance for High-Risk AI Systems

7.3.1 General

High-risk AI systems are permitted on the EU market *only* if they comply with a series of essential context requirements set out in **Chapter III, Section 2** of the AI Act [i.2]. These requirements span risk management, data governance, technical documentation, record-keeping, transparency, human oversight, and performance (accuracy, robustness, cybersecurity). Each context requirement corresponds to an Article (Art. 9 through 15). **Providers of conforming high-risk AI systems document the compliance with each of these obligations.** Below, each Article's context requirement is analysed, the needed specific **documentation tasks**, and recommend **documentation approaches or techniques** are identified. The goal is to guide providers in creating documentation that not only meets the legal minimums but is organized and effective for compliance demonstration. The mandated context requirements for AI providers are presented in Figure 6 and further detailed in the following clauses. The identification and specification of these contextual requirements directly forms the structure and contents of the technical documentation, ensuring traceability from system objectives and constraints to documented design decisions, risk controls, and life cycle artifacts.

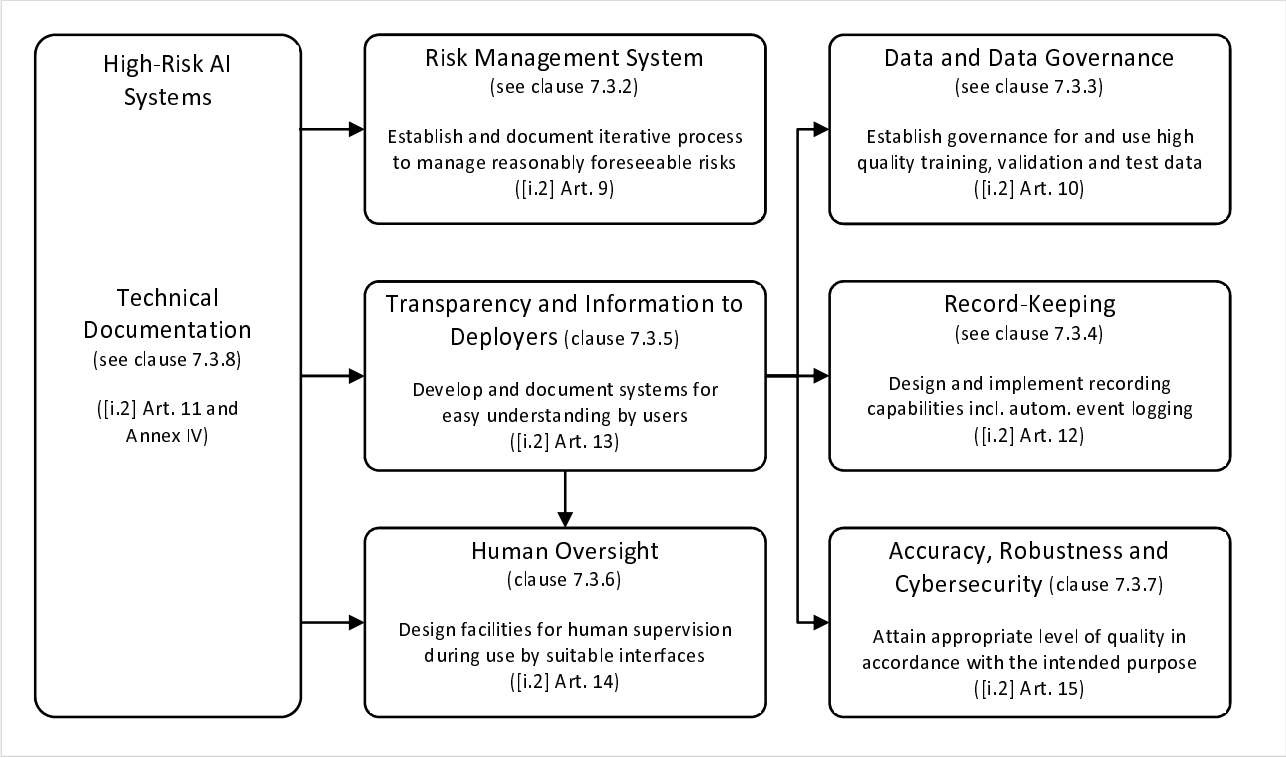

Figure 6: Overview of context requirements for technical documentation of High-Risk AI Systems

7.3.2 Risk Management System (Art. 9)

Context Requirement: According to EU AI Act providers of high-risk AI systems establish **a risk management system** and operate it throughout the AI system's life cycle. This is a continuous, iterative process of identifying, analysing, and mitigating risks. The provider document the risk management and keep it up-to-date. In practice, the AI Act's Art. 9 requires providers to perform a thorough hazard and risk assessment before market release and to update it based on post-market monitoring. Key steps include identifying reasonably foreseeable risks to health, safety, and fundamental rights; estimating and evaluating their severity and probability; implementing measures to mitigate or eliminate those risks; and testing the AI system to validate risk mitigations. The identified risks cover not just the intended use but also reasonably foreseeable misuse of the AI system.

Documentation Tasks: The provider establishes a comprehensive **risk management documentation** by adhering to a structured documentation process (see clause 6.7). At a minimum, the following documents are included [i.2]:

- *Risk identification:* The document lists identified risks to health, safety, or fundamental rights (e.g. discriminatory bias, technical malfunctions), including intended purpose, context of use, and vulnerable groups affected. *(Art. 9(2)(a) and Art. 9(9))*
- *Risk analysis and evaluation:* The document describes each identified risk under both intended purpose and reasonably foreseeable misuse scenarios. Additionally, the document sets out other risks possibly arising, based on data from post-market monitoring. *(Art. 9(2)(b), Art. 9(2)(c))*
- *Risk mitigation:* The document describes the mitigation measures implemented for each risk that cannot be eliminated (e.g. design modifications, safeguards, training data improvements, warnings in the user instructions, etc.). It should map each mitigation to the corresponding risk and indicate the resulting residual risk. *(Art. 9(4), Art. 9(5)(a)-(c))*
- *Residual risk justification:* The document justifies why the residual risk is judged acceptable. (*Art. 9(5))*
- *Risk-based testing:* The document includes a description of the testing carried out to identify the most appropriate and targeted risk management measures, ensuring that the high-risk AI system performs consistently for its intended purpose. This involves testing in real-world conditions, throughout development and in any event prior to placing on the market or putting into service. Additionally, the document reflects metrics and probabilistic thresholds related to the testing procedures. *(Art. 9(6), Art. 9(7), Art. 9(8), Art. 60)*

Recommended Documentation Approaches: Providers are encouraged to adopt structured, standardized documentation methods to effectively document risk management practices. Recommended approaches include:

- *Risk Management Standards (ISO 31000 [i.13], ISO 14971 [i.15])*: Adapt established safety engineering and medical device risk management frameworks for structured risk plans and comprehensive logging in AI-specific contexts.
- *Datasheets for Datasets:* Link data-related risk mitigations directly into Datasheets, clearly documenting data quality, representativeness, and bias reduction measures.
- *Model Cards:* Employ Model Card templates (see clause D.2.1) to document system performance, robustness checks, and bias evaluation results, providing clear evidence for risk-based testing.
- *Assurance Cases*: Develop structured safety arguments (goal → argument → evidence) to comprehensively integrate risk documentation (see clause D.4.4). Assurance Cases should reference Risk Cards, Model Cards, and test reports, systematically demonstrating how identified risks are mitigated and safety objectives are achieved.

7.3.3 Data and Data Governance (Art. 10)

Context Requirement: According to EU AI Act high-risk AI systems that use data for **training**, **validation**, or **testing** meet strict data quality and governance requirements as under Art. 10. The used datasets are relevant, representative, complete, and as accurate as possible for the AI system's intended purpose. They have appropriate statistical properties and do not introduce unjust bias, especially toward demographic groups. Data collection and processing follow clear governance procedures and comply with data protection laws and ethical standards. The goal is to ensure that the AI system relies on high-quality, well-managed data that supports fair and reliable outputs.

Documentation Tasks: The provider establishes a comprehensive **data documentation** by adhering to a structured documentation process (see clause 6.7). At a minimum, the documentation includes the following key information elements for all datasets used in training, validation, and testing [i.2]:

- *Data collection and origin*: The document describes design choices made during the development and management of datasets that affect how the AI system is trained, validated, and tested. Furthermore, the document demonstrates the data's origin and how data is collected, including a description of the data collection protocols (e.g. web scraping, public datasets, or user-generated data) Additionally, the document lists annotation, labelling, cleaning, updating, enrichment and aggregation procedures, and any data augmentation or synthesis techniques used. Moreover, the assessment results of the availability, quantity and suitability of the data sets are included. (*Art. 10(2)(a), 10(2)(b), 10(2)(c), 10(2)(e))*
- *Representativeness and relevance*: The document describes the data's intended purpose, relevance, errors, statistical properties and representativeness. The documentation reflects alignment with the AI system's intended purpose and context of use. It specifies data context, including properties specific to the contextual, geographical, behavioural, and functional setting of the AI system's intended use. *(10(3), 10(4))*
- *Data Preparation:* The document demonstrates that datasets are, "*to the best extent possible, free of errors and complete*." Additionally, the document includes summary statistics (e.g. class distributions, missing data rates, label error checks), data-preparation procedures (e.g. updating, labelling, annotation, enrichment, aggregation, and cleaning via outlier removal), and any known limitations. Providers document the resolution of any existing quality issues. *(Art. 10(2)(c), 10(3))*
- *Bias Assessment and Mitigation:* The document reflects potential dataset bias and the respective detection, mitigation and prevention measures. Additionally, the document includes findings of negative impact on the health and safety of persons, causes of discrimination as well as fundamental rights, and corrective actions such as data augmentation, training adjustment or targeted data collection. Where personal data is processed, the documentation indicates reasons why the processing of special categories of personal data was strictly necessary as well as demonstration of compliance with applicable data protection laws. *(Art. 10(2)(f), 10(2)(g), 10(3), 10(5) with cross-references to GDPR (Reg. 2016/679), Reg. 2018/1725, and Dir. 2016/680))*

Recommended Documentation Approaches: To systematically fulfil these documentation requirements, providers are advised to employ the following proven techniques:

- *Datasheets for Datasets*: Utilize structured templates or questionnaires to capture detailed metadata on data origin, composition, quality metrics, and bias mitigation efforts. Datasheets provide comprehensive evidence supporting regulatory reviews (Gebru et al. [i.19], clause D.1.1).
- *Data Statements or Data Cards:* Offer concise yet thorough documentation focusing particularly on ethical aspects, data representativeness, and bias considerations, suitable especially for NLP and sensitive-data scenarios (clause D.1.4).
- *Data Nutrition Labels:* Present key data quality indicators concisely, offering quick readability and clarity on dataset characteristics and representativeness (clause D.1.3).
- *Bias Mitigation Logs*: Maintain explicit records of bias assessments, adjustments, and corrective actions. Such logs enhance transparency and support risk management documentation aligned with ISO/IEC TR 24028 [i.1].

All data governance and quality documentation feed directly into the Technical Documentation (Annex IV), explicitly fulfilling the requirement of the EU AI Act to document dataset characteristics. Providers substantiate all claims regarding data quality and representativeness with quantitative evidence, such as demographic analyses or statistical breakdowns.

This structured documentation not only ensures compliance with Art. 10 but also provides crucial evidence for risk management purposes (Art. 9), particularly regarding bias mitigation and data integrity, strengthening the overall AI system's transparency, trustworthiness, and regulatory compliance.

7.3.4 Record-Keeping (Art. 12)

Context Requirement: High-risk AI systems which comply with the EU AI Act are designed to facilitate the recording of events ("logs") during operation, as appropriate for their intended purpose. The logs facilitate traceability, allowing for the reconstruction of system functionality, especially during instances of failure or unexpected behaviour. Providers establish an automatic logging mechanisms and guarantee the retention of logs for subsequent review.

Documentation Tasks: The provider establishes comprehensive **record-keeping and logging documentation** by adhering to a structured documentation process (see clause 6.7). At a minimum, the documentation includes the following key information elements [i.2]:

- *Logging Specifications:* The document includes the events logged during the AI system's operation. For biometric systems, the provider indicates at least how the period of use of the AI system, the input data compared with a reference database, the matches found during the comparison and the identification of the natural persons involved in the verification of the results are recorded. (*Art. 12(1), 12(2), 12(3), Annex III (1), Art. 79(1))*
- *Log Access and Analysis:* If a competent authority requests generated logs, the document describes access methods and analyses tools provided to access the generated logs. This includes any available interfaces, such as administrative dashboards or APIs, as well as tools provided for audit or incident response purposes. If logs are encrypted or require special handling, particularly in cases involving personal data, those procedures are described. Providers ensure that sufficient information is included to enable competent authorities, auditors, or incident responders to obtain and interpret the logs effectively. *(Art. 12(2), Art. 21(2))*
- *Data Protection Considerations:* If logs include personal data, the document addresses compliance with applicable data protection laws (e.g. GDPR), including lawful basis, storage security, and access restrictions. *(Art. 19(1), GDPR (Reg. 2016/679))*

Recommended Documentation Approaches: Providers are advised to adopt the following documentation techniques:

- *Log Structure and Sample Entries*: Clearly document log formats with illustrative examples, such as: *"[2025-05-01 10:30:15] INPUT ID=abc123, Decision=Approved, Score=0.87"*. These entries clarify how logs directly support traceability and accountability.
- *Integration with Risk Management*: Demonstrate the linkage of logging mechanisms with risk mitigation strategies. For example, logs can provide auditability for explainability-related risks, reinforcing oversight and compliance with ISO/IEC 42001 [i.11].

7.3.5 Transparency and Information to Deployers (Art. 13)

Context Requirement: High-risk AI systems which comply with the EU AI Act are transparent enough to allow deployers to understand and use their outputs correctly. This includes clear, accurate, and accessible instructions for use and information to interpret systems output and behaviour. Providers also ensure deployers receive adequate documentation and training.

Documentation Tasks: The provider establishes comprehensive **Instructions for Use (User Manual)** as part of structured documentation process. At a minimum, the documentation includes the following key information elements [i.2]:

- *Intended Purpose:* The document states the intended purpose of the AI system and reasonably foreseeable misuse, which may lead to risks to the health, safety, or fundamental rights. *(Art. 13(3)(b)(i), (iii))*
- *Installation, Maintenance and Updates:* The document specifies required computational hardware resources, expected lifetime of the AI system, and essential maintenance measures including software updates. (*Art. 13(3)(e))*
- *Instructions for use:* The document includes concise, complete, correct, clear and accessible information guiding the use of the system, addressed to deployers, e.g. in digital format. (*Art. 13(2))*
- *Output Interpretation Guide:* The document explains the meaning of outputs and behaviour using AI system's capabilities and human oversight. *(Art. 13(3)(d), b(iv); Art. 14(4) (b)-(d))*
- *Performance:* The document sets out the AI system's level of accuracy, robustness and cybersecurity. If persons are affected, the document describes the AI system's performance regarding the affected parties, on which the system is intended to be used. (*Art. 13(3)(b)(ii))(v))*

Recommended Documentation Approaches: Providers are advised to adopt established user documentation standards and techniques, including:

- *Adopt ISO/IEC/IEEE 26511:2018 [i.27]*: Use established standards for structured, clear, and user-oriented documentation.
- *Include practical Aids*: Add quick reference guides, threshold tables, flowcharts, and explainability summaries to support usability.
- *Ensure Documentation Consistency:* Align user instructions with technical documentation, clearly reflecting system features, limitations, and oversight mechanisms.

7.3.6 Human Oversight (Art. 14)

Context Requirement: High-risk AI systems which comply with the EU AI Act are designed to allow effective **human oversight** to prevent or reduce risks to health, safety, or fundamental rights. When designed appropriate human overseers are able to understand the system, interpret its output, and intervene or shut it down when necessary. Human oversight is documented in two ways: (1) in the design documentation, showing what oversight measures are built into the system, and (2) as guidance in the user instructions, detailing how the oversight is to be performed. As user guidance is addressed in clause 7.3.5, this clause focuses on documenting the design rationale, mechanisms, and technical implementation of human oversight in accordance with Art. 14.

Documentation Tasks: The provider establishes comprehensive **human oversight** documentation as part of structured documentation process (see 6.7). At a minimum, the documentation includes the following key information elements [i.2]:

- *Oversight Measures:* The document specifies the specific technical and organizational human oversight measures, including into the AI system built-in measures and measures to be implemented by the deployer. *(Art. 14(3))*
- *Risk mitigation:* The document explains how the oversight measures effectively mitigate risks to health, safety, and fundamental rights. *(Art. 14(2))*
- *Interface for Oversight:* The document describes user interface elements that enable interpretation of system behaviour, monitoring and controlling. *(Art. 14(1), 14(4)(c), 14(4)(d))*

- *Override and Stop Controls:* The document defines how human operators can override or reverse the system's outputs and safely interrupt its operation. This includes the design, logic, and accessibility of override functions, as well as the implementation and operation of the stop function. (*Art. 14(4)(d), (e))*

- *Automation Bias Mitigation:* The document describes measures implemented to reduce the risk of automation bias, where human overseers may over-rely on AI outputs. This can include interface design choices such as requiring human confirmation for critical decisions, displaying confidence levels, or using alerts to encourage the operator's involvement. (*Art. 14(4)(b))*

Recommended Documentation Approaches: Providers are advised to adopt established human-system interaction standards and practices, including:

- *Oversight Scenarios:* Use real-world use cases to illustrate oversight procedures.

- *Human Factors Standards:* Apply ergonomic and usability documentation methods (e.g. ISO 9241-210 [i.14]) to ensure interface accessibility and interpretability.

- *Quality Management Linkage*: Reference quality management system procedures (Art. 17), such as post-market oversight reviews and operator feedback loops.

- *Training Annex*: Reference or annex any training resources developed for human overseers to support operational readiness and compliance.

- *Oversight Plan Template*: Include a reusable plan outlining deployer oversight tasks (e.g. monitoring frequency, response actions, warning signs).

Consistency is maintained between design documentation and user instructions, such that oversight features described in system design (e.g. override buttons, alert systems) are also reflected in the Instructions for Use.

7.3.7 Accuracy, Robustness, and Cybersecurity (Art. 15)

Context Requirement: Art. 15 of the EU AI Act mandates that high-risk AI systems achieve and maintain an appropriate level of *accuracy*, *robustness*, and *cybersecurity* throughout their life cycle. These characteristics are essential to ensure the system operates reliably under expected conditions, withstands disturbances, and is protected against manipulation or misuse.

Documentation Tasks: At a minimum, the provider includes the following information elements as part of a structured documentation process [i.2]:

- *System Performance Summary:* The document provides an overview of how the AI system ensures accuracy, robustness, and cybersecurity throughout its life cycle. It declares achieved accuracy levels, along with relevant metrics, benchmark and measurement methodologies. *(Art. 15(1), Art. 15(2)) Art. 15(3))*

- *Robustness Report:* The document demonstrates the AI system's resilience to internal faults, environmental variations, and interaction with users or other systems. It describes technical design (e.g. redundancy solutions, fail-safes, alerts, recovery modes) and organizational measures *(Art.15(4))*

- *Learning Feedback:* If AI systems continue to learn after being placed on the market, the document specifies mitigation measures to prevent biased feedback loops. *(Art.15(4))*

- *Vulnerability Mitigation*: The document includes implemented measures to prevent, detect, and respond to, resolve and control for attacks and inputs designed to cause the AI model to make a mistake. *(Art.15(5))*

Recommended Documentation Approaches:

- *Model Cards***:** Use Model Cards to document accuracy metrics, robustness considerations, intended use conditions, and evaluation results (e.g. precision, recall, calibration). These facilitate clarity and consistency across system documentation (clause D.2.1).

- *Validation Reports*: Maintain detailed test reports and performance logs covering both standard and stress conditions.

- *Cybersecurity Standards Integration*: Align mitigation documentation with best practices from AI-specific cybersecurity frameworks (e.g. ISO/IEC 27001 [i.12], ISO/IEC TR 24028 [i.1]) to strengthen conformity.

Documented evidence of system accuracy, resilience, and security directly supports the technical documentation required under Annex IV and strengthens the system's conformity assessment and certification under Art. 43-44 of the EU AI Act [i.2].

7.3.8 Technical Documentation (Art. 11)

High-risk AI systems which comply with the EU AI Act have a comprehensive and up-to-date technical documentation prior to market placement. The technical documentation builds upon the contextual requirements described above, operationalizing them into structured evidence required by Art. 11 and Annex IV of the AI Act. This documentation provides sufficient information to authorities to verify conformity (*Art. 11, Annex IV*). Annex IV lists a detailed but minimum set of required documentation items. SMEs and start-ups may utilize a simplified EU-prescribed documentation form (*Art 11 (1)*). Well formed technical documentation are **clear, comprehensive, and up to date**, are retained for 10 years (Art. 18) and updated to reflect any system changes.

Documentation Requirements: According to Annex IV, at a minimum, the technical documentation includes [i.2]:

- *General AI System Description*: The document defines the intended purpose of the AI system, provider details, system version, interactions with external hardware/software or other AI systems, and deployment formats (e.g. embedded hardware, APIs). It also includes user-interface descriptions, hardware environment specifications, and, where applicable, illustrations of physical products containing the AI. *(Art. 11(1), Annex IV(1)(a-h); see also clause 7.3.5)*

- *Design and Development*: The document details the process for the AI system's development, including methods and procedural steps, usage and integration of pre-trained systems, system logic, algorithms, key assumptions, classification strategies, optimization objectives, and any significant technical trade-offs. It also describes the AI system's architecture, component interactions, and computational resources utilized. *(Art. 11(1), Annex IV(2)(a-c, f); see also clauses 7.3.6 and 7.3.7)*

- *Data Documentation*: The document describes the datasets used for training, detailing their provenance, selection, representativeness, labelling, cleaning, and enrichment methodologies. It provides evidence supporting data quality, representativeness, and suitability to the intended purpose, including data protection measures applied. *(Art. 11(1), Annex IV(2)(d); see also clause 7.3.3)*

- *Human Oversight Measures*: The document presents an assessment of technical and organizational measures enabling human oversight as defined in Art. 14. It describes how the system facilitates human interpretation and appropriate responses to system outputs. *(Art. 11(1), Annex IV(2)(e); see also clause 7.3.6)*

- *Validation and Testing Reports:* The documents summarizes all validation and testing procedures (incl. data), accuracy and robustness metrics as well as cybersecurity and bias assessments. They refer to acceptance criteria on quality characteristics from e.g. ISO/IEC standards on software quality [i.6], [i.7], or [i.8]. It includes signed and dated test reports along with test logs. *(Art. 11(1), Annex IV(2)(g-h); see also clauses 7.3.3 and 7.3.7)*

- *Cybersecurity Measures*: The document describes cybersecurity protocols implemented to protect the AI system against AI-specific vulnerabilities, such as adversarial attacks and data manipulation. *(Art. 11(1), Annex IV(2)(h); see also clause 7.3.7)*

- *AI System Monitoring and Control*: The document describes the system's operational capabilities and limitations, highlighting accuracy across targeted user groups, foreseeable unintended outcomes, and risks to health, safety, fundamental rights, or discrimination. It also specifies input data requirements clearly. *(Art. 11(1), Annex IV(3); see also clauses 7.3.5 and 7.3.6)*

- *Performance Metrics Appropriateness*: The document provides a justification and rationale for selecting specific performance metrics, demonstrating their suitability for evaluating the AI system's intended functionalities and outputs. *(Art. 11(1), Annex IV(4); see also clause 7.3.7)*

- *Risk Management System*: The document summarizes the risk management procedures implemented in compliance with Art. 9, including identified risks, applied mitigations, and justification for residual risk acceptability. *(Art. 11(1), Annex IV(5); see also clause 7.3.1)*

- *Life cycle Changes Record*: The document maintains an ongoing record of all significant modifications and updates made to the AI system throughout its life cycle. *(Art. 11(1), Annex IV(6); see also clause 7.3.3)*

- *Standards and Compliance Declaration*: The document includes references to all fully or partially applied harmonised standards or an alternative measure employed for compliance. It also contains a copy of the official EU Declaration of Conformity. *(Art. 11(1), Annex IV(7-8); see also clause 7.3.8)*

- *Post-Market Performance Evaluation System*: The document details the processes established for ongoing post-market performance monitoring and evaluation of the AI system, including a monitoring plan as required by Art. 72(3). *(Art. 11(1), Annex IV(9); see also clause 7.3.4)*

Information elements: Through an analysis of the AI Act, and in particular Annex IV, the following list identifies necessary information that the AI Act requires to be included in the technical documentation. Such information is referred to as "information elements". Given the substantial number of these elements, Figure 7 presents a visual overview for a more efficient navigation. To enhance clarity and facilitate the understanding of technical documentation obligations under the AI Act, the documentation items and corresponding information elements are structured into three primary categories:

1) **AI System information:** This category encompasses details pertaining to the AI system itself. It includes, but is not limited to, information about the system's architecture, its development life cycle, and its intended purpose.

2) **Data information:** This category focuses on the data utilized in the training, validation, testing, and operation of the AI system. It covers details such as data collection methods, processing procedures, and test reports, among others.

3) **Controls information:** This category addresses the safeguards (i.e. controls) implemented to mitigate risks associated with the AI system. These controls can also be defined as risk mitigation measures and apply to various stages and components of the AI system. For instance, controls may target the system itself (e.g. human oversight, accuracy), the underlying data (e.g. data transparency, quality), or the development process (e.g. risk management, quality assurance).

It is important to note that this categorization is not explicitly presented in the AI Act, nor are the documentation items and information elements presented hierarchically in the legislative text as suggested in Figure 7. However, this hierarchical structure is very useful for understanding the documentation requirements in the AI Act, as it provides a structured approach compared to the simple list of items in Annex IV.

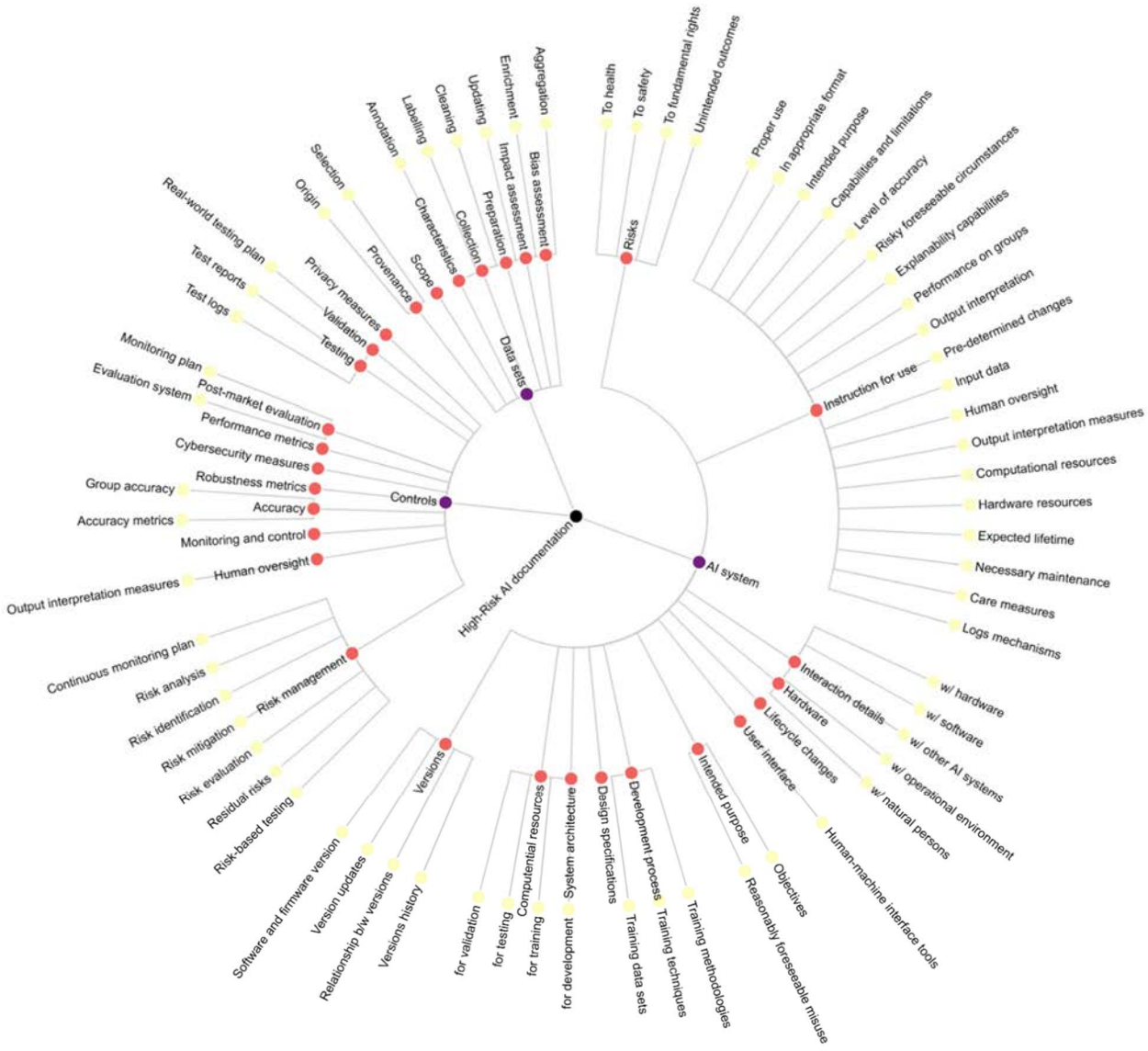

Figure 7: High-Risk: Documentation Items and information Elements required by the AI Act

Recommended Documentation Approaches: While the AI Act provides an extensive list of documentation items and corresponding information elements to be documented, it does not offer technical details on how to do so. Therefore, it is crucial to have documentation schemes that provide more detailed guidance on how to document each of these items. Clauses 7.3.2 to 7.3.7 offer structured guidelines for documentation of AI Requirements set forth in Art. 9 to 15.

- *Structured Technical Dossier*: Maintain a clearly structured, regularly updated master technical file explicitly aligning documentation elements to the Art. 11 and Annex IV requirements.
- *Compliance Traceability Matrix*: Provide a clear mapping matrix linking each AI Act requirement (Art. 9 to 15) directly to corresponding evidence sections within the technical documentation.
- *Use of Simplified Forms for SMEs*: Small and medium-enterprises may opt to fulfil the Annex IV documentation requirements via a simplified form developed by the European Commission. When used, this form is accepted by notified bodies for the purposes of conformity assessment, in accordance with Art. 11(1).
- *Reference to Documentation Techniques*: Specific documentation techniques suitable for each requirement category (e.g. risk management, data governance, human oversight, robustness) are further discussed in the corresponding clauses of clause 7.

7.4 Documentation Requirements for GPAI models

7.4.1 General

Under the AI Act, **General-Purpose AI (GPAI) models** are defined as models that display **significant generality** and are capable of competently performing a **wide range of distinct tasks** (Art. 3(63)). All GPAI providers should comply with obligations in Art. 53, including keeping up-to-date technical documentation, publishing a summary of the training data used, establishing a policy to respect copyright and providing documentation to downstream deployers (Art. 53).

Models released under a free and open-source license, with publicly available weights and architecture, are exempt from certain obligations unless they are classified as **systemic-risk** GPAI (Art. 53(1)). A GPAI model is presumed to pose **systemic risk** if it has **high-impact capabilities** based on benchmarks (e.g. when it was trained using a total computational power greater than $\mathbf{10^{25}}$ **FLOPs or** is designated by the Commission as such (Art. 52).

Systemic-risk GPAI model providers should comply with additional obligations under Art. 55, including model evaluation and adversarial testing, conducting a model-specific systemic **risk assessment** and **mitigation**, **reporting** serious **incidents**, and **maintaining** adequate **cybersecurity**. Providers should notify the Commission within two weeks if they determine a model meets the systemic-risk criteria (Art. 52(1)).

7.4.2 GPAI Models without Systemic Risk

Under **Art. 53 of the AI Act**, all GPAI model providers, regardless of systemic risk status, should maintain the following **technical documentation** and provide **sufficient information to deployers**, at a minimum [i.2]:

- *Technical Documentation*: The document describes the model's architecture, intended tasks, training process, computational and energy resources used, evaluation results, known limitations, technical means to integrate the GPAI model. (Art. 53(1)(a); see also Annex XI)
- *Training Data Summary*: The document includes a public summary, published by the provider, describing the datasets used to train the model. (Art. 53(1)(d))
- *Copyright Compliance Policy*: The documentation explains how the provider complies with copyright related to Art. 4(3) of Directive (EU) 2019/790. (Art. 53(1)(c))
- *Downstream Documentation*: The document includes technical documentation for downstream providers. (Art. 53(1)(b); Annex XII)

7.4.3 GPAI Models with Systemic Risk

In addition to the baseline requirements listed in clause 7.3.2 GPAI models with systemic risk should meet the following requirements [i.2]:

- *Model Evaluation and Adversarial Testing*: The document includes test and evaluation results, including adversarial testing. (Art. 55(1)(a))
- *Systemic-Risk*: The document describes the systemic risks, the origins and the mitigation measures applied. (Art. 55(1)(b))
- *Serious Incident Reporting*: In case of serious incidents and possible corrective measures, the document describes the procedures in place for tracking and documenting serious incidents. (Art. 55(1)(c))
- *Cybersecurity Protections*: The document details the cybersecurity protection for the GPAI model as well as its infrastructure. (Art. 55(1)(d))

7.4.4 Documentation

Figure 8 gives an overview of information elements for GPAI models without and with systemic risk:

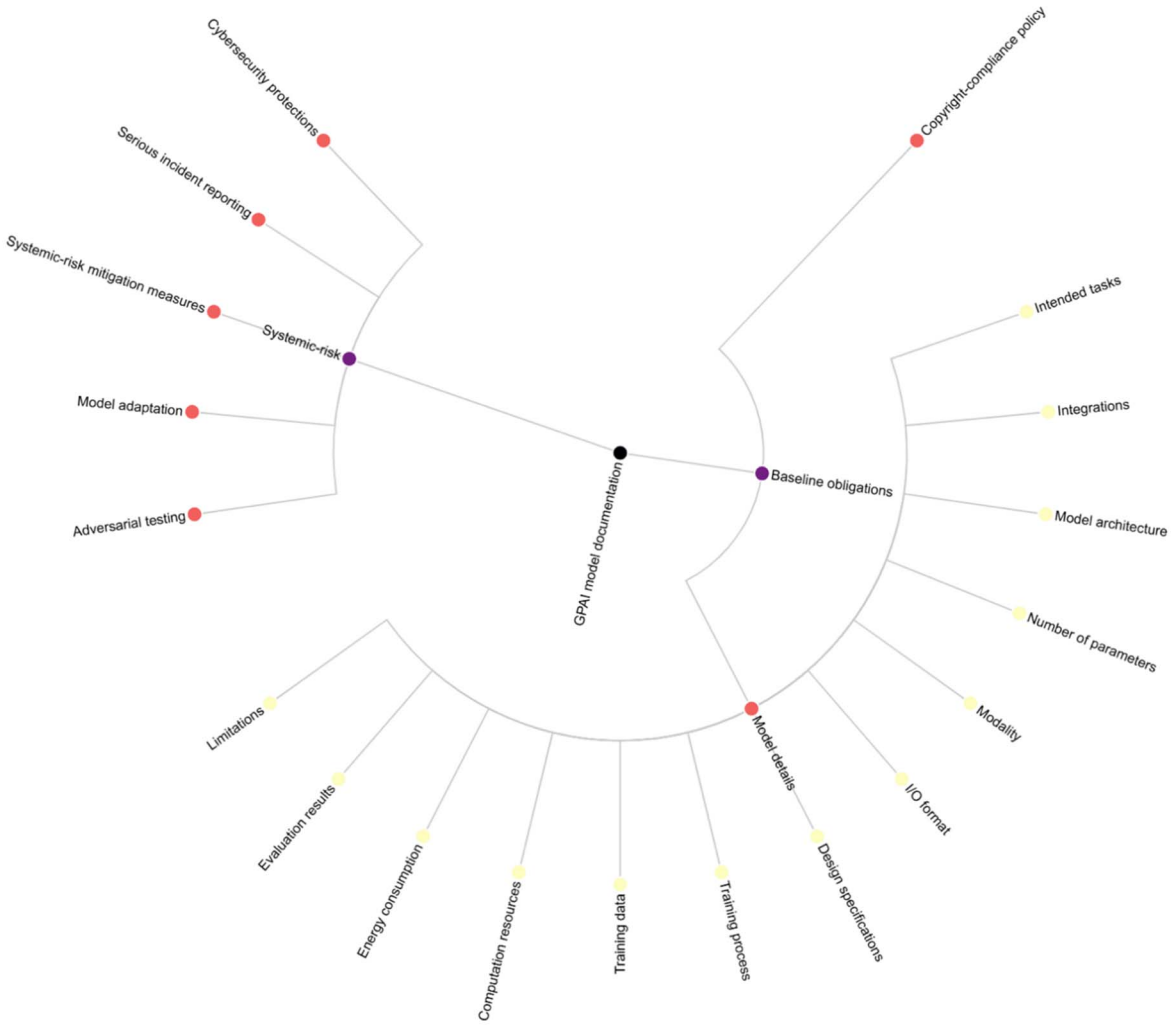

Figure 8: GPAI models: Documentation Items and information Elements required by the AI Act

Recommended Documentation Approaches: Providers are encouraged to adopt structured, standardized documentation methods to efficiently comply with the technical and risk-based obligations under Art. 53 and 55 of the AI Act. Recommended approaches include:

- *Risk Management Frameworks (ISO 31000 [i.13], ISO 14971 [i.15])*: Use general and sector-specific risk management standards to structure risk identification, mitigation, and documentation processes. These frameworks help produce a comprehensive risk file in line with Annex IV(5), covering known risks, mitigation measures, and residual risk justifications.

- *Datasheets for Datasets*:
  Integrate datasheets into documentation workflows to capture dataset origin, representativeness, data processing methods, and bias mitigation strategies. These are directly relevant for fulfilling the public training data summary under Art. 53(1)(d) and Annex IV(2)(d).

- *Model Cards*: Use standardized Model Card formats to document model purpose, architecture, limitations, performance metrics, and robustness evaluations. Model Cards support both technical documentation (Art. 53(1)(a)) and transparency obligations toward deployers (Annex XII).

Annex A:
Sample Documentation Scenarios

A.1 Healthcare Use Case

A.1.1 Use Case Description

Health Educational Conversational Agent (HECA) v2 Virtual Assistant (VA) is a Generative AI-powered conversational agent that specializes in providing information about medical products and services. The medical assistant undergoes training in the comprehension of medical documentation using advanced Natural Language Understanding (NLU) techniques implemented by Large Language Models (LLMs). The primary objective is to effectively comprehend user inputs by leveraging NLU principles and provide precise and contextually relevant responses related to the field of health and medicine. In addition, this agent has a fallback system using LLMs trained on peer-reviewed medical articles. This ensures that users have a seamless experience, even when they ask unclear or unidentified questions. The HECA v2 Virtual Assistant is specifically designed to manage confidential medical information with utmost confidentiality.

HECA v2 is applied within Horizon Europe AI4HF project to support the development of a comprehensive and standardized methodological framework for trustworthy and ethical provision of personalized risk assessment and care plans for individuals living with Chronic Heart Failure (https://www.ai4hf.com/about-ai4hf). According to the EU Artificial Intelligence Act (Art. 6 and Annex III), AI systems used in healthcare for diagnosis, prognosis, or clinical decision support are classified as high-risk [i.3]. As AI4HF falls under this category, HECA v2, being an integral part of the framework, should be fully documented to ensure compliance with the AI Act requirements for high-risk AI systems.

A.1.2 Documentation Approach

A.1.2.1 Key Documentation Requirements

High-risk AI systems should maintain comprehensive and up-to-date technical documentation, as required in clause 7.3.8. This clause provides illustrative examples for documenting the first two key documentation requirements:

- General AI System Description.
- Design and Development Documentation.

A.1.2.2 Example 1: General AI System Description

The **General AI System Description** is mandatory documentation element derived from Art. 11 and Annex IV of the AI Act. The following steps outline the application of the proposed documentation approach described in clause 6.7:

Step 1: Understand and identify the purpose of the documentation artifacts

Purpose: To communicate essential general information about the AI System to downstream users and stakeholders (researchers, patients, clinicians), enabling a common understanding of system capabilities, intended use, and limitations.

Step 2: Identify the selected documentation aspects per document

Documentation Item (clause 6.2): AI System

Documentation Stakeholder (clause 6.3): Provider (AI Developer)

Phase of Documentation (clause 6.4): Monitoring & Maintenance (living document, updated as required)

Documentation Technique (clause 6.5): Structured tabular format

Step 3: Identify the document contents (information elements) and create/assemble the document

Document: Table "General AI System Description"

Information items are presented in the following structured table.

Table A.1: General AI System Description

Purpose	
System Name	HECA V2 Virtual Assistant
Provider Name	CERTH/ITI
Version	Current version of the system is HECA Version B: Generative AI-powered conversational agent with two-LLM architecture for healthcare applications https://heca.iti.gr/e086a922-812c-4f9c-91bc-94e22d431c1f Previous versions: Generation A / CERTH Intelligent Personal Agent (CIPA): Focused on natural language interaction (text/voice) for indoor tasks, without LLMs. Generation B / CIPA Educational Virtual Assistant (EVA): Smartphone-based conversational agent using RASA framework, NLU, dialogue flow, and AI planning. Deployed in Smart Home for energy and health domains. HECA Version A / Health Education Agent v1: Educational Virtual Assistant for diabetes management and patient education. Integrated into self-management app, based on NLP/NLU models (Chrodis Plus JA).
Intended Purpose	The primary objective of the LLM-based virtual medical assistant is to enhance the management, accessibility, and reuse of healthcare data, particularly for critical chronic conditions such as rare cancers and heart failure. The assistant is designed to respond to user inquiries in an intuitive and user-friendly manner, supporting improved data governance and providing information to aid clinical decision-making, while not performing autonomous clinical decisions.
Use Case	HECA v2 serves as a core component within the AI4HF project, supporting the development of a standardized framework for ethical and trustworthy AI. It provides a conversational interface for personalized risk information retrieval and data interpretation, aligned with FUTURE-AI guidelines [i.86].
Target Audience	Researchers, Patients, Health Professionals
Scope of the application	The system is intended for public use in the healthcare domain as an informational tool to support research, patient education, and clinical decision-making processes.
System Description	
General functionality	HECA v2 is an AI-powered virtual medical assistant that provides accurate, domain-specific responses to user queries about chronic heart failure. It uses a two-LLM architecture: the first LLM retrieves relevant Question-Answer (QA) pairs from pre-stored embeddings in ChromaDB, while the second LLM generates the final answer based on these retrieved pairs. The system operates on pre-collected and processed medical data, with no live interaction with IoT devices or external sensors. A specialized scoring mechanism ensures that responses are contextually relevant and semantically aligned with the user's query, enabling trustworthy and precise conversational support.
Scale of deployment	The system is designed for global use and is accessible without user registration, supporting anonymized interactions. As such, there is no predefined limit on the number of users or connected endpoints, enabling broad scalability across diverse geographic regions. No geographic, sectoral, or institutional restrictions are imposed by the system architecture.
Interaction with external systems	The AI system interacts with external software components through standardized communication protocols, specifically HTTP requests and WebSockets. It relies on an internal ChromaDB instance for data retrieval and is currently deployed in an in-house environment, with planned migration to cloud infrastructure. The system incorporates multiple integrated Large Language Model (LLM) components that interact internally to support its core functionality. The system does not interface with external hardware (e.g. IoT devices) or third-party AI systems beyond its defined architecture.
Software and firmware details	The system is accessed via a web browser and operates through standard HTTP and WebSocket protocols. It does not rely on specific client-side software or firmware versions. Updates are applied to the server-side components as needed, with no required user-side updates, ensuring compatibility across common web platforms.
Deployment formats	Web-based service (browser interface); no packaged deployment required.
Update and maintenance	Updates are performed as needed, in alignment with European regulatory requirements and applicable contractual obligations. Updates may include enhancements to LLM components, QA content, and supporting infrastructure. Updates are deployed to the server environment through controlled processes; no user-side downloads or actions are required.

System Description	
Hardware requirements	The system is accessed via client devices (PC or mobile) equipped with a standard web browser; no specific hardware is required on the user side beyond browser compatibility (Chrome recommended). On the server side, the system operates on infrastructure with GPU capabilities (minimum 48 GB GPU RAM), with optimal performance achieved using multiple GPUs to support LLM inference and database retrieval operations.
User Interface	The system provides a React-based web front end accessible through standard web browsers. It presents a conversational chatbot interface developed within the framework of the Horizon Europe AI4HF project. The interface allows users to enter free-text queries and receive natural language responses based on curated medical knowledge. The interaction is focused on guided question-answer dialogue to support informational needs and understanding of chronic heart failure; the system does not generate autonomous clinical recommendations or decisions.

A.1.2.3 Example 2: Design and Development Documentation

The Design and Development Documentation, as required by Art. 11(1) and Annex IV(2)(a-c, f) of the AI Act, constitutes a broad and complex obligation that encompasses multiple documentation artifacts spanning the entire development life cycle of the AI system, as outlined in clause 7.3.8. In this example, a proposed documentation approach (see clause 6.7) is illustrated for one core component of the system architecture: the QA model. To support clarity, first a high-level component diagram of the overall system architecture is provided, and subsequently it is demonstrated how the QA component can be documented using the Model Card technique. The component diagram is illustrated in Figure A.1.

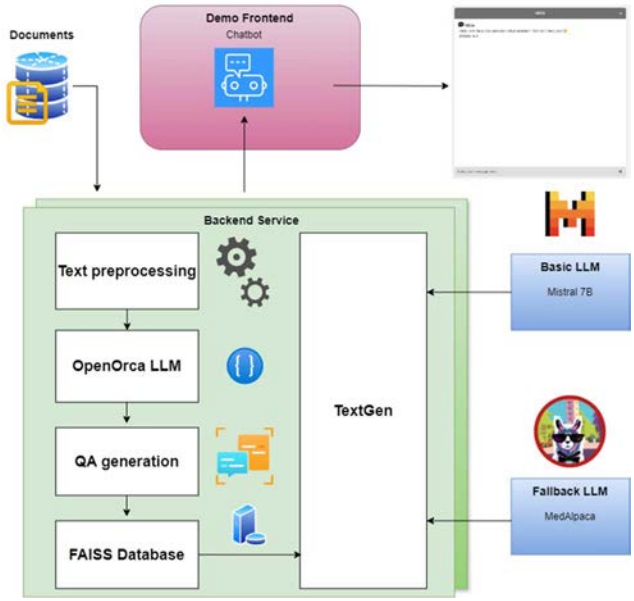

Figure A.1: HECA V2 Virtual Assistant system architecture

The **HECA V2 Virtual Assistant** is an AI-powered web-based system designed to provide reliable, domain-specific conversational support in the healthcare domain. Its architecture consists of a modular backend pipeline and an interactive frontend interface.

Conceptual Workflow:

- The system is initially trained on a curated set of healthcare-related documents.
- Once trained, users can interact with the assistant via a web-based frontend, which communicates with the backend to deliver accurate, context-aware responses.

System Components:

- **Frontend Interface:** Provides an intuitive, web-based conversational interface for users to submit queries and receive responses. Additionally, it manages communication between the user and backend services.

- **Backend Architecture:**
  - **Text Processing Module:** Analyses the input documents during system training, performing preprocessing steps to extract and structure relevant information.
  - **OpenOrca LLM Module:** Utilizes the OpenOrca Large Language Model to automatically generate Question-Answer (QA) pairs from the pre-processed documents.
  - **QA Generation:** Refines and prepares the generated QA pairs for storage and retrieval.
  - **FAISS Database (Vector Encoding Database):** Stores vectorized representations of the QA pairs using Facebook AI Similarity Search (FAISS), enabling efficient semantic retrieval during user interactions.
  - **Textgen Module:** Generates dynamic responses to user queries, leveraging relevant QA pairs retrieved from the FAISS database.
    - *Mistral 7B LLM:* Responds to queries directly related to information stored in the FAISS database.
    - *MedAlpaca LLM*: Handles queries that fall outside the trained knowledge base, providing responses to broader medical questions.

The following steps outline the application of the proposed documentation approach described in clause 6.7 for the documentation of the "**QA Generation**" component.

Step 1: Understand and identify the purpose of the documentation artifacts

Purpose: To communicate essential technical and ethical information about the QA model to downstream users (researchers, patients, and health professionals), enabling safe and effective use of the AI assistant.

Step 2: Identify the selected documentation aspects per document

Documentation Item (clause 6.2): QA Model (based on two LLMs with Chroma DB)

Documentation Stakeholder (clause 6.3): Provider (AI Developer)

Phase of Documentation (clause 6.4): Implementation & Integration (Model Training, Evaluation, and Deployment)

Documentation Technique (clause 6.5, D2.1): Model Card in structured tabular format

Step 3: Identify the document contents (information elements) and create/assemble the document

Document: Model Card for QA Model

Information items (1-7) are presented in the Model Card template below.

Table A.2: Model Card for QA Model

Model Overview	
Name	HECA V2 QA Model
Version	HECA Version B
Description	The QA Model is a core component of the HECA V2 Virtual Assistant. It enables the system to generate domain-specific, context-aware answers to healthcare-related queries by retrieving and processing QA pairs derived from curated medical documents.

<table>
<tr><th colspan="2">Purpose</th></tr>
<tr><td>Intended Use</td><td>The QA Model component is designed to automatically generate high-quality Question-Answer (QA) pairs from curated healthcare-related documents, enabling efficient semantic retrieval and explainable conversational support within the HECA V2 Virtual Assistant.
Realizable Capabilities
<table>
<tr><th>Sense</th><th></th><th>Process Knowledge</th><th></th><th>Act</th><th></th><th>Communicate</th><th></th></tr>
<tr><td>Visual</td><td>☐</td><td>Factual</td><td>■</td><td>Physical</td><td>☐</td><td>Visual</td><td>☐</td></tr>
<tr><td>Auditory</td><td>☐</td><td>Procedural</td><td>☐</td><td>Non-physical (Agents)</td><td>☐</td><td>Auditory</td><td>☐</td></tr>
<tr><td>Olfactory</td><td>☐</td><td>Conceptual</td><td>■</td><td rowspan="4"></td><td></td><td>Olfactory</td><td>☐</td></tr>
<tr><td>Gustatory</td><td>☐</td><td>Metacognitive</td><td>☐</td><td></td><td></td><td>Tactile</td><td>☐</td></tr>
<tr><td>Tactile</td><td>☐</td><td rowspan="2"></td><td></td><td></td><td></td><td>Textual</td><td>■</td></tr>
<tr><td>…</td><td>☐</td><td></td><td></td><td></td><td>Gestural</td><td>☐</td></tr>
</table></td></tr>
<tr><td>Primary Users</td><td>Researchers, health professionals, patients</td></tr>
<tr><td>Use Cases</td><td>Conversational support for healthcare education, research assistance, clinical support (non-decision-making).</td></tr>
<tr><td>Domain</td><td>Healthcare, with a focus on chronic heart failure.</td></tr>
<tr><td>Usage Scope</td><td>Public-facing web application, accessible globally via browser.</td></tr>
<tr><th colspan="2">Model Details</th></tr>
<tr><td>Architecture</td><td>Hybrid retrieval-generation architecture using:
1) OpenOrca LLM to generate QA pairs during training;
2) FAISS-based semantic retrieval engine;
3) Textgen module with Mistral 7B and MedAlpaca LLMs to generate responses.</td></tr>
<tr><td>Data Sources</td><td>Curated healthcare documents related to chronic heart failure. Data is pre-approved and processed internally. No live connection to clinical systems.</td></tr>
<tr><td>Training Process</td><td>A retrieval-augmented generation method is applied to generate Q/A pairs from the source documents using OpenOrca LLM. These Q/A pairs are post-processed and then stored to FAISS vector database. The process is repeated iteratively as new documents or updates become available.</td></tr>
<tr><td>Fallback Mechanism</td><td>If no matching QA pairs are found in FAISS, MedAlpaca LLM provides a general fallback answer to medical queries (with disclaimers for limitations).</td></tr>
<tr><th colspan="2">Evaluation & Performance</th></tr>
<tr><td>Metrics</td><td>Retrieval precision, response relevance (manual review), semantic similarity, user satisfaction ratings.</td></tr>
<tr><td>Validation</td><td>Manual validation by medical domain experts and NLP engineers during testing phase; iterative refinement based on test results.</td></tr>
<tr><td>Limitations</td><td>Model does not cover all possible medical questions. No real-time clinical data integration. Responses may not reflect most recent medical guidelines. Model cannot replace professional medical advice.</td></tr>
<tr><th colspan="2">Ethical Considerations</th></tr>
<tr><td>Fairness</td><td>Training data is curated to ensure diversity of medical content and representation across genders and demographic groups where applicable. Regular audits of QA pair generation are conducted to identify and mitigate potential biases in content or language.</td></tr>
<tr><td>Explainability</td><td>The architecture supports traceability; responses are generated via QA pairs retrieved from an indexed database (FAISS Database). Mistral 7B answers are based on retrievable inputs, and fallback answers (from MedAlpaca LLM) are flagged to indicate their more general nature.</td></tr>
<tr><td>Transparency</td><td>The system provides a clear description of its intended purpose, its two-LLM architecture, and how it generates responses based on curated medical knowledge. All updates are logged and versioned in internal technical system documentation, with major changes recorded in a public changelog for transparency.</td></tr>
<tr><td>Accountability</td><td>The system's compliance with the EU AI Act requirements and alignment with FUTURE-AI guidelines establish a framework for accountability, supported by comprehensive documentation and internal governance processes by CERTH/ITI.</td></tr>
<tr><td>Privacy & Security</td><td>User interactions are anonymized, and no personally identifiable information is stored. The model is trained and operates solely on pre-approved and processed internal medical data, with no live connection to clinical systems or direct use of patient data in training or inference. This design aims to prevent privacy intrusions related to data usage consent.</td></tr>
<tr><th colspan="2">Maintenance & Updates</th></tr>
<tr><td>Update Frequency</td><td>Planned quarterly updates to QA pair database and model tuning. Emergency updates as required (e.g. guideline changes).</td></tr>
<tr><td>Monitoring</td><td>System activity monitored through server logs; user feedback mechanisms planned; regular review of QA pairs and LLM outputs.</td></tr>
</table>

Maintenance & Updates	
Changelog	All updates logged and versioned in internal system documentation; major changes recorded in public changelog for transparency.
Contact & Governance	
Maintainer	CERTH/ITI AI Development Team.
Updates	Updates reviewed and approved by CERTH/ITI team leader and AI4HF coordinator.
Compliance	Aligned with requirements of the EU AI Act Art. 11 and Annex IV and with the FUTURE-AI guidelines.
Note	The capabilities listed in the section 'Intended Use' are an illustrative example for additionally provided information, to provide required information on AI systems capabilities to support transparency (see clause 7.3.5 and [i.89], [i.90], [i.91]).

A.2 AI-Based Person Detection for Construction Machinery

A.2.1 Use Case Description

This use case focuses on the creation and preparation of a domain-specific image and video dataset that can later be used for training and testing AI systems designed for detecting persons around construction machinery. The current stage of the project is centered on data collection and annotation under varied environmental and operational conditions. The dataset includes labelled frames extracted from videos recorded using Zed 2i stereo cameras in realistic construction scenarios.

Although no AI model has been fully trained or deployed within this use case, preliminary use of YOLOv8 models has supported pre-annotation and evaluation workflows. The dataset itself is being designed to serve as a foundation for the future development and validation of AI-based person detection systems. These systems may ultimately be integrated into mobile machinery to improve safety during reverse or swing operations.

This initiative is conducted in collaboration with the Construction Future Lab and the Federal Institute for Occupational Safety and Health (BAuA) and places emphasis on ethical data use, traceability, annotation quality, and regulatory preparedness. The resulting dataset and processes support the reproducibility of AI safety research and align with the EU AI Act requirements (see clause 7.3.3).

A.2.2 Documentation Approach

A.2.2.1 Key Documentation Requirements

High-risk AI systems should maintain comprehensive and up-to-date technical documentation, as detailed in clause 7.3.8, and should fulfil the obligations of documenting data governance. This clause provides illustrative examples for documenting the first two key documentation requirements:

- General AI System Description
- Data collection and origin

A.2.2.2 Example 1: General AI System Description

AI-based Person Detection for Construction Machinery refers to a potential AI-driven safety application that could use computer vision to identify individuals in hazardous zones around construction equipment. Although no such system has been developed within the scope of this project, the present documentation is intended to serve as a conceptual framework for how such systems might be described and assessed in the future.

The General AI System Description is a mandatory documentation element derived from Art. 11 and Annex IV of the AI Act. The following steps outline the application of the proposed documentation approach described in clause 6.7.

Step 1: Understand and Identify the purpose of the documentation artifacts

Purpose: to provide a clear and comprehensive overview of the AI system, including its intended use, functionality, and limitations. It explains how the system operates within its deployment context and outlines key aspects of its life cycle - such as data handling, integration, and safety relevance. This enables stakeholders, including researchers, industry partners, and regulatory authorities, to develop a shared understanding of the system's design, purpose, and implications for safe deployment in construction environments.

Step 2: Identify the selected documentation aspects per document

Documentation Item: AI System

Documentation Stakeholder: Provider (AI Developer)

Phase of Documentation: Monitoring & Maintenance (living document, updated as required)

Technique (clause 6.5): Structured tabular format

Step 3: Identify the document contents (information elements) and create/assemble the document

Information items are provided when such a model exists. It is recommended to structure the information as done in Table A.1.

A.2.2.3 Example 2: Data Documentation

The data documentation describes a curated and fully annotated visual dataset developed to support the development of AI systems for person detection around construction machinery in real-world construction site environments. Commissioned by the Federal Institute for Occupational Safety and Health (BAuA), the dataset was created as part of a research project carried out in collaboration with Construction Future Lab. Its purpose is to enable the evaluation and future implementation of AI-based person recognition systems tailored to the construction domain.

The dataset includes over 100 GB of video data, covering approximately 100 videos and 10 000 labelled images. These were collected using Zed 2i stereo cameras in diverse construction site conditions, encompassing scenarios such as wheel loader reversing, swivel and reversing area monitoring on excavators, and static field views using tripod-mounted cameras. The primary object class is "person," with data captured under varied lighting, weather, and body posture conditions to ensure representativeness.

With data collection and annotation now complete, the dataset is ready for use in training and evaluating AI models. Although no operational AI system has yet been built within this project, the dataset provides a robust and application-specific foundation for the future development and validation of safety-related AI solutions. It is intended for internal research, prototyping, and testing purposes, not for direct deployment or commercial use, and supports the transparent, standards-aligned advancement of high-risk AI systems focused on occupational safety.

Step 1: Understand and identify the purpose of the documentation artifacts

Purpose: To explain how data is collected, processed, and managed across its life cycle, providing stakeholders - including researchers, industry partners, and regulatory authorities - with a clear understanding of the system's capabilities, deployment context, and safety implications within construction environments.

Step 2: Identify the selected documentation aspects per document

Documentation Item: Conducted data collection and pre-processing steps

Documentation Stakeholder: Provider (AI Developer who uses the data for model training, etc.)

Phase of Documentation: Data Preparation & Processing (see clause 6.4)

Documentation Technique: Structured tabular format

Step 3: Identify the document contents (information elements) and create/assemble the document

Information items are presented in the following structured table.

Table A.3: Data Documentation

Purpose	
Name	Annotated Dataset for Person Detection in Construction Machinery Environments
Provider Name	Construction Future Lab GmbH
Version	1.0
Intended Purpose	To provide a curated dataset for training and evaluating AI-based person detection systems aimed at improving occupational safety in construction sites.
Use Case	Supports the development and validation of person recognition systems for use in reversing and swivel operations of mobile machinery such as excavators and wheel loaders.
Target Audience	Researchers, AI system developers, occupational safety experts, and regulatory auditors.
Scope of the application	The dataset is intended for internal research, prototyping, and validation phases of AI development. It is not designed for direct integration or commercial deployment, but as a foundational resource to support the safe and transparent development of high-risk AI systems.
Dataset Description	
Intended Use	Collection of annotated video/image data for training/testing AI models for person detection.
Content Volume	Over 100 GB of video data, approx. 100 video files, and 10 000 annotated images representing real-world construction site scenarios.
Relevant Object Classes	Person.
Collection Methods	Stereo video recordings with Zed 2i cameras under varying lighting, weather, and operational conditions (e.g. standing, walking, reversing zones).
Annotation Process	Pre-annotation using YOLOv8 models; manual verification and refinement to ensure labelling quality.
Scale of deployment	Experimental setup with stereo cameras; data stored on local devices and shared via secure academic platforms.
Interaction with external systems	Use of TU Dresden cloud, GitLab for versioning, local storage (HDDs).
Regulatory Context	The dataset is intended to support future development of AI systems that align with applicable regulatory and industry standards. These include: • **AI Act [i.2]** (classifying person detection as a high-risk application) • **GDPR [i.85]** (ensuring lawful and transparent data use) • **ISO/IEC 42001:2023 [i.10]** (AI management systems) • **ISO/IEC 23894:2023** [**i.10]** (AI risk management) • **ISO/IEC TR 5469:2024 [i.9]** (AI and functional safety) • **ISO 13849-1 [i.16]** (safety-related control system performance) • **DIN EN ISO 16001 [i.18]** (object detection in earth-moving machinery) • **ISO 21815-1:2022 [i.17]** (collision avoidance and interface protocols) • **Machine Regulation 2023/1230 [i.88]** (mandating third-party testing for machine learning-based safety components).
Software and firmware details	Label Studio, YOLOv8 variants (n-x), Python scripts for evaluation.
Deployment formats	Scripts, labelled datasets, and HTML/PDF reports for internal validation.
Update and maintenance	Manual updates by developers; version control via GitLab.
Hardware requirements	Zed 2i stereo camera, GPU-enabled workstation, minimum 1 TB HDD per session.
User Interface	Web-based annotation interface (Label Studio); role-based access for annotators and reviewers.
Data Privacy and Ethical Concerns	All individuals recorded in images/videos provided informed consent for data use, application in AI systems, and possible publication. No biometric or identifying data is used (in accordance with GDPR).

Annex B:
Trustworthy AI: Definition and core characteristics

B.1 Definition of Trustworthy AI

This annex provides a concise overview of Trustworthy Artificial Intelligence (AI), outlining foundational concepts, operational requirements, core characteristics, essential frameworks, and explicit alignment with the European Union Artificial Intelligence Act (EU AI Act).

Trustworthy AI, as defined by the High-Level Expert Group on AI (AI HLEG) in their Ethics Guidelines for Trustworthy AI, is built upon three pillars that form the foundation of trustworthy AI as indicated in Figure B.1, and necessitate adherence throughout the entire AI system life cycle:

- **Lawful:** AI systems rigorously comply with all applicable legal and regulatory frameworks. This encompasses adherence to national, international, and European Union legislation, including but not limited to the General Data Protection Regulation (GDPR) and relevant sector-specific directives. This adherence ensures AI operations remain within established legal parameters, safeguarding fundamental rights and societal values.
- **Ethical:** Beyond strict legality, AI systems are required to embody and uphold established ethical principles and values. This component is instantiated through four core ethical principles:
  - **Respect for Human Autonomy:** AI systems should augment human capabilities, facilitate informed decision-making, and preserve human control.
  - **Prevention of Harm:** AI systems are designed to preclude the infliction of physical, psychological, or economic detriment. Proactive identification and mitigation of potential negative impacts are imperative
  - **Fairness:** AI systems operate equitably, actively mitigating unjustifiable bias and discrimination, thereby ensuring impartial treatment across individuals and groups.
  - **Explicability:** The processes, functionalities, and decision-making mechanisms of AI systems exhibit transparency, interpretability, and comprehensibility to relevant stakeholders, thereby enabling scrutiny and accountability.
- **Robust:** AI systems are required to possess both technical and societal robustness. This necessitates that they be reliable, secure, and resilient, capable of consistent and safe operation within diverse real-world environments, while also adapting responsibly to evolving societal contexts. Technical robustness pertains to attributes such as accuracy, dependability, and cybersecurity, whereas societal robustness encompasses broader ethical considerations and societal impact.

For the operationalization of these three fundamental pillars, the AI HLEG introduces **seven key requirements**. Through a detailed analysis, the needed characteristics for each of these requirements, as proposed by the AI HLEG, have been identified and are indicated in Figure B.1.

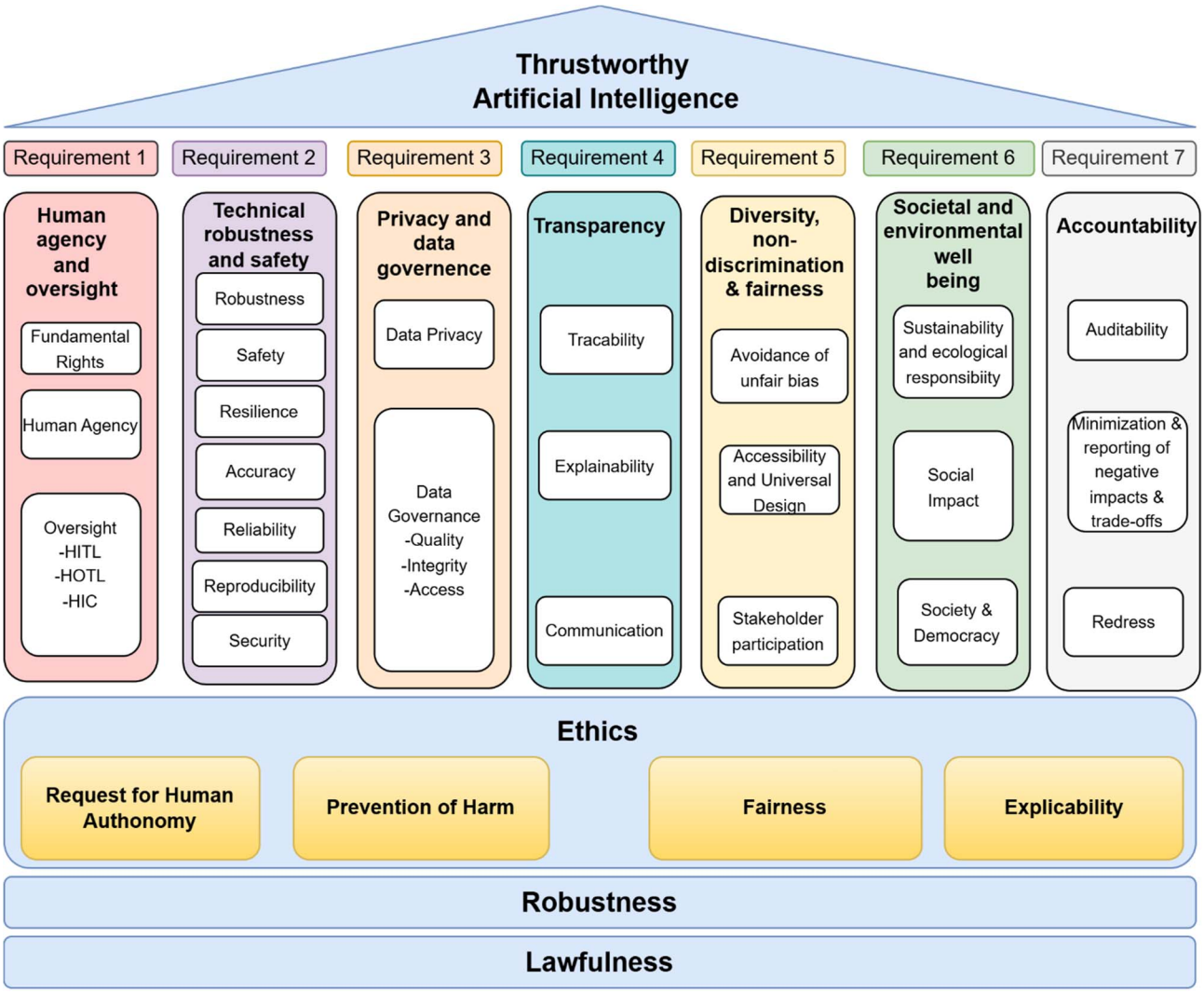

Figure B.1: Trustworthy AI pillars, requirements and characteristics (adopted based on [i.84])

B.2 Relevant frameworks and guidelines

A variety of international and European frameworks offer prescriptive guidelines and principles for the development and deployment of trustworthy AI. These frameworks collectively underscore the global consensus on the ethical and practical considerations necessary for trustworthy AI. The AI HLEG framework serves as a foundational basis, and elements from other prominent frameworks demonstrate a significant alignment, mapping onto its core structure. Table B.1 summarizes salient aspects, principles, or qualities derived from selected prominent frameworks.

Table B.1: Relevant frameworks and guidelines for AI trustworthiness

Framework/Guidelines	Overview
Ethics Guidelines for Trustworthy AI [i.5]	Defines **seven key requirements** for Trustworthy AI derived from **four ethical principles**: Human Agency and Oversight; Technical Robustness and Safety; Privacy and Data Governance; Transparency; Diversity, Non-discrimination and Fairness; Societal and Environmental Well-being; Accountability.
OECD Principles on AI [i.3]	Articulates **five value-based principles** for responsible AI stewardship: Inclusive Growth, Sustainable Development and Well-being; Human-centred Values and Fairness; Transparency and Explainability; Robustness, Security and Safety; Accountability.
ISO/IEC TR 24028:2020 [i.1]	Surveys methods for establishing and assessing AI trustworthiness, covering: transparency, explainability, controllability, engineering risks, mitigation techniques, and qualities like availability, resiliency, reliability, safety, security, and privacy.
AI4People [i.52]	Proposes five ethical principles (Beneficence, Non-maleficence, Autonomy, Justice, Explicability) for ethical AI, focusing on opportunities and risks.
Requirements (SQuaRE) - Quality model for AI systems [i.6]	Extends the SQuaRE framework for AI systems, focusing on AI-specific characteristics: user controllability, functional adaptability, robustness, and societal/ethical risk mitigation.

B.3 Operationalization of Trustworthiness in the EU AI Act

The EU AI Act directly operationalizes the principles of Trustworthy AI, particularly for high-risk AI systems, by translating ethical and robust considerations into concrete legal obligations. The Act reflects the "Lawful," "Ethical," and "Robust" components as follows:

- **Lawful:** The EU AI Act itself constitutes a foundational legal framework. It mandates compliance with extant legislation (e.g. GDPR) through provisions such as data governance requirements (Art. 10) and comprehensive documentation obligations (Art. 11).
- **Ethical:**
  - **Human Agency and Oversight (AI HLEG):** Addressed by the Act's mandate for human oversight mechanisms (Art.14), ensuring continued human control and intervention capabilities.
  - **Privacy and Data Governance (AI HLEG):** Directly paralleled in the stringent requirements for data quality and robust data governance practices (Art. 10) to prevent discriminatory outcomes and safeguard privacy.
  - **Transparency (AI HLEG):** Ensured through obligations pertaining to transparent operation, clear instructions for use (Art. 13), and meticulous record-keeping (Art. 12).
  - **Diversity, Non-discrimination, and Fairness (AI HLEG):** Addressed through the emphasis on preventing bias within training datasets and model outputs (Art. 10), thereby promoting equitable outcomes.
  - **Societal and Environmental Well-being (AI HLEG)** & **Accountability (AI HLEG):** Supported by requirements for robust risk management systems (Art. 9) and post-market monitoring (Art. 61), which collectively aim to identify, assess, and mitigate broader societal impacts and assign responsibility.
- **Robust:**
  - **Technical Robustness and Safety (AI HLEG):** Explicitly encompassed by the Act's provisions concerning accuracy, robustness, and cybersecurity (Art. 15), mandating resilience against errors, faults, and malicious interventions to ensure safe operational performance.

Annex C:
Risk Mitigation by Documentation

As stated in the publication Best Practices in AI Documentation [i.31], to build trustworthy AI-based systems, it is necessary to consider a variety of risks associated with the availability of poor documentation about their structure and building methodology. Connected with this statement, some researchers claim that the potential of such AI-based systems can be largely overestimated, having virtually no documentation demonstrating an actual trustworthiness. Others have raised concerns regarding potential adverse consequences of such systems, including person harm, technical, and socio-ethical risks [i.60], [i.80], [i.81], [i.82] and [i.83]. These works paved the way to the awareness that low quality, or absence, of documentation can lead to **seven categories of risks**:

- Human harm due to AI errors.
- Misuse of AI tools.
- Risk of bias in AI and perpetuation of inequities.
- Lack of transparency.
- Privacy and security issues.
- Gaps in AI accountability.
- Obstacles to implementation in real-world scenarios.

These risks could result in harm to individuals, which results in the reduction of the level of **trust** in AI-based systems by the society at large. Therefore, the development, review, and deployment stages of an AI-enabled system should include risk assessment and management as core components for establishing trustworthiness.

Human harm due to AI errors:

- Why documentation is important to avoid human harm.
- Impact of low-quality documentation on stakeholders.
- Main stakeholders affected: AI Customer, AI Subject.
- Main quality aspects: technical robustness/safety

AI systems are sometimes linked to malfunctions that might ultimately lead to **safety issues** for their users despite ongoing advancements in data accessibility and machine learning. The effects of such issues with AI tools in sensitive domains (e.g. healthcare) include, among others (i) *false negatives* concerning missed classifications concerning life-threatening conditions; (ii) *excessive optimistic/pessimistic behaviour* because of erroneous false positives (i.e. healthy people mistakenly regarded as ill by the AI algorithm); and, (iii) *inappropriate interventions* due to imprecise classification (e.g. inaccurate prioritization of interventions in emergency rooms).

Hence, to avoid end-users' harm, AI engineers should document errors and adjustments during AI deployment to support transparency. Furthermore, AI solutions should be dynamic, as such, they should include features that continuously learn from new scenarios and from errors detected in actual use. Still, in order to detect issues as they arise there is the need for some human management and oversight, which consequently may lead to higher expenses and a loss in the early advantages of AI.

Misuse of AI tools

- The importance of having high quality documentation to improve the appropriate usage of AI systems and, at the same time, to increase trust among end users.
- Main stakeholders: AI Customer, AI Subject.
- Main concern (quality aspect): reliability.

There is always a danger of human mistakes and human misuse in the context AI systems' usage. As a matter of fact, even though AI systems are accurate and robust, the efficacy and **reliability** of such tools depend on how the end users will utilize them in practice. AI technologies are vulnerable to incorrect usage or human mistakes due to a variety of issues. They have commonly been created and developed by computer/data scientists with little input from end users, which can lead to complex and unnatural interactions that require the users to become accustomed to the new technology to learn how to use it.

To decrease human mistakes or improper usage of AI systems, an effective documentation strategy should be used. To improve the knowledge and abilities of AI users and thereby decrease human mistakes, a complete and effective documentation of AI systems should be established and broadly distributed throughout society at large.

Risk of bias in AI and perpetuation of inequities

- Data catalogues used to build an AI model may contain biases.
- Sometimes, biases cannot be avoided, the documentation may provide details about known biases, mitigation actions, and/or motivation about their presence.
- Main stakeholders: AI Customer, AI Subject, Relevant Authorities.
- Main quality aspect: reliability.

Although there are constant advancements in the research and treatments of *data biases within AI systems*, significant inequities and prejudice still exist throughout the majority of the world's countries, which inherently influence how AI technologies function. Sex and gender, age, ethnicity, wealth, education, and geography are the primary causes of these disparities.

Additionally, even though some of these injustices are institutional, because of factors like socioeconomic disparities and discrimination, personal biases still play a significant part. For instance, if the medical domain is considered as test-bed for this type of analysis, research surveys in the United States have shown that doctors do not treat Black patients' complaints of pain as seriously or as promptly as they do White patients' ones [i.64], [i.53], [i.57] and [i.61]. Gender-based bias is another illustration of a widespread prejudice that is prevalent in most nations throughout the world's healthcare systems, somewhat to varied degrees.

Therefore, there exists the fear that, if not appropriately developed, assessed, and controlled, future AI-enabled systems might entrench and even magnify the widespread imbalances and human biases that lead to general disparities.

Lack of transparency

- Low-quality, or absence, of documentation affects the overall transparency of the AI system.
- Main stakeholders: AI Customer, AI Subject, Relevant Authorities.
- Main quality aspect: transparency, explainability, accountability.

Despite ongoing developments in AI-powered solutions, people as well as professionals still see existing algorithms as intricate and obscure technologies that are challenging to completely understand, trust, and accept.

Lack of transparency is frequently cited as a significant problem with the creation and application of AI solutions. Such and issue particularly affect high-stakes fields like healthcare and finance. This may lead to a serious lack of trustworthiness in AI, particularly in delicate fields like medicine, finance, transportation that are concerned with the life of humans. Likewise, a low level of trustworthiness will undoubtedly affect how extensively stakeholders embrace new AI algorithms.

A crucial component of trustworthy AI is *traceability*, which refers to the comprehensive documentation of the complete AI development process and monitoring of how the AI model performs in actual use after deployment [i.68], [i.56] and [i.59]. Whereas traceability focuses on the **transparency** of the AI algorithm, **explainability** is crucial for ensuring transparency for each prediction and decision made by an AI system [i.74] and [i.79]. Thus, the **lack of explainability** makes it challenging to determine the cause of AI failures and establish **accountability** when things go wrong. Therefore, lack of transparency hinders stakeholders from applying AI solutions to their everyday jobs since, in order to employ a given AI solution, a user should be able to comprehend the underlying ideas that underlie each choice and/or prediction, even if the algorithm itself has the potential to increase its productivity [i.66].

Privacy and security issues

- Gaps in documentation may cause issues in the management of both the privacy and the security aspects of an AI systems.
- Main stakeholders: AI Customer, AI Subject, Relevant Authorities.
- Main quality aspect: security, privacy, confidentiality.

The creation of AI-based solutions raised significant hazards for a **lack of data privacy**, **confidentiality**, and **protection**, which might result in serious repercussions, including the release and use of private information that violates people's rights or the reusing of people data for purposes other than the ones for which the AI solution has been developed. These problems are connected to informed consent, which is the provision of sufficient information to users allowing them to make informed decisions, such as whether to share personal data.

With the advent of digital technology into daily lives and the formalization of informed consent in the Helsinki Declaration, informed consent has become an increasingly important and fundamental aspect of the users' experience [i.71]. Moreover, according to [i.72], informed consent is related to a number of ethical concerns, such as safeguarding against damage, upholding autonomy, protecting privacy, and preserving property rights over data tissue.

The amount of autonomy and the potential of collaborative stakeholders decision-making is nonetheless constrained by the introduction of obscure AI algorithms and confusing informed consent procedures [i.77]. Users are finding it ever more challenging to comprehend the decision-making process, the many uses for which their data may be put to, and the precise procedures for choosing not to share their data. The interest reader can find more details and several examples in the literature [i.54], [i.60], [i.62], [i.63], [i.65], [i.67] and [i.70].

Gaps in AI accountability

- High-quality documentation is essential to trace the accountability of information sources used to build the AI models.
- Researchers and groups working to address the legal implications of the introduction and use of AI algorithms in various facets of human life have given the term "algorithmic accountability" greater attention.
- Main stakeholders: AI Customer, AI Subject.
- Main quality aspect: Accountability.

The expression "algorithmic accountability" may seem to relate to the attempt to keep the algorithm itself responsible, but, it means the exact opposite. Indeed, it highlights that algorithms are developed using a combination of machine learning and human configuration and that errors in algorithms are caused by the people who develop, implement, or use the machines, particularly considering that AI systems cannot be held morally or legally accountable by themselves [i.73]. *Accountability* is crucial in AI for several fields since it will help the technology gain acceptance, credibility, and eventual adoption in the society [i.62] and [i.76].

AI developers and engineers typically operate within ethical guidelines, whereas the end users need to be accountable for their acts, according to regulatory obligations, as a necessary part of their professional activity [i.78]. Additionally, the ethical codes and accountability standards that several private corporations employ have frequently come under fire for being ambiguous and challenging to implement in reality [i.73]. As a result, the end users who are unable to explain their actions and choice process are at risk of losing the ability to practice their work. Whereas, in the same circumstances the repercussions for a technician are far less severe. Also, even if an AI developer is determined to be at fault because numerous different engineers and researchers collaborate on any single AI system, it can be challenging to place the responsibility for the error on a single individual.

Obstacles to implementation in real-world scenarios

- Low-quality, or absence, of documentation can impede the deployment of AI-based solutions.
- Poor documentation affects integration with existing systems, limiting practical applicability.
- Main stakeholders: AI Customer, AI Subject.
- Main quality aspects: technical robustness, reliability.

Over the past 10 years, several algorithms for AI have been created and suggested to be used in a variety of applications [i.58] and [i.69]. Nevertheless, the deployment, integration, and adoption of AI technologies are still paved reality with unique challenges, even when the technologies have gone through the validation process and have been found to be reliable and secure, morally upright and compliant [i.75], and interoperable [i.55].

Annex D:
Documentation Schemes and Gap Analysis to the EU AI Act

D.1 Data-Focused Documentation Approaches

D.1.1 Datasheet for Datasets

In 2018, Gebru et al. [i.19], proposed Datasheets for Datasets which was designed to document the creation and use of datasets, making them a valuable resource for the following group of stakeholders:

- Dataset creators.
- Dataset consumers.
- Policymakers.
- Consumer advocates.
- Investigative journalists.
- Individual whose data is included in datasets.
- Individuals impacted by models trained or evaluated using datasets.

This documentation approach spans the following key stages of the dataset life cycle:

- Motivation.
- Composition.
- Collection process.
- Processing/cleaning/labelling.
- Uses.
- Distribution.
- Maintenance.

It is produced using a questionnaire, with the aim of enhancing transparency and accountability in dataset handling. However, while Datasheets for Datasets provide in-depth documentation, they can be resource-intensive to create and maintain, especially for large and evolving datasets. Also, it focuses exclusively on documenting datasets, which limits its scope of application.

D.1.2 DescribeML

DescribeML was proposed by Giner-Miguelez et al. [i.20] for documenting the structure, data provenance and social concerns of ML datasets. It intends to meet the needs of the following stakeholder:

- Dataset creators.
- Dataset consumers.

This proposed approach spans the following stages of data creation:

- Gathering.
- Labelling.

- Design.

This documentation approach employs a Domain Specific Language in documenting datasets. While DescribeML emphasizes the ethical and social dimensions of data usage, it is also limited in its focus which makes it unapplicable in documenting technical performance aspects of an AI system.

D.1.3 Dataset Nutrition Label

The Dataset Nutrition Label framework was proposed by Holland et al. [i.32] to enhance data quality standards by providing a clear and standardized way to describe datasets. Inspired by nutritional labels on food, these labels offer detailed information about datasets, including their provenance, composition, and any potential biases. This framework is intended to help researchers and practitioners make more informed decisions about the datasets they use, ultimately leading to more reliable and ethical AI systems. The methodology is aimed at the following stakeholders:

- Data specialist.
- Dataset builders and publishers.

Prior to model development, Dataset Nutrition Label is used to document dataset 'ingredients' at the following stages of the ML development pipeline:

- Dataset collection.
- Dataset preprocessing.

It uses a web-based application as its documentation approach. Despite Dataset Nutrition Labels comprehensive documentation of dataset 'ingredients', it may be difficult to apply this documentation approach to build a label for sensitive or proprietary data as such data might be accessible only to those who created the dataset and not to the public.

D.1.4 Data Cards

Data cards [i.35] are introduced as a documentation tool to promote transparency and responsibility in AI dataset usage. They provide detailed descriptions of datasets, including their creation, intended use, and potential biases. The purpose is to help users understand the data's characteristics and limitations, ensuring more ethical and effective application of AI technologies. The methodology is aimed at the following stakeholders:

- Producers (dataset creators).
- Agents (stakeholders who read transparency report and have the authority to use or decide how to datasets will be used).
- End users.

This documentation approach documents key information about ML dataset across the dataset's life cycle, employing Google Docs as its documentation template. While using Google Docs facilitates collaboration among multiple stakeholders, it limits the way input could be provided and may also cause template fragmentation as multiple changes are made to an individual field.

D.1.5 Dataset Development Life Cycle Documentation Framework

This paper [i.34] explores methods to improve accountability in Machine Learning (ML) datasets by drawing parallels with practices from software engineering and infrastructure. The authors propose frameworks and guidelines to document the provenance, characteristics, and usage of datasets, emphasizing the importance of version control, issue tracking, and Continuous Integration/Deployment (CI/CD) pipelines. The goal is to enhance transparency, reproducibility, and accountability in ML dataset management. For convenience, their proposed approach as the Dataset Development Life Cycle Documentation Framework (a term introduced by the present document to capture their documentation-based methodology) is referred to here. The methodology is aimed at the following stakeholders:

- Domain experts.
- Data creators/labeller.

- Data scientists.
- Adversarial testers.

This documentation approach is applied at each stage of the dataset life cycle:

- Requirement analysis.
- Design.
- Implementation.
- Testing.
- Maintenance.

It is created using an information sheet. Although it offers detailed documentation for each stage of the dataset development life cycle, its focus is limited, similar to other data-focused documentation methods, and it does not apply to the entire ML development life cycle.

D.2 Model-And-Method-Focused Documentation Approaches

D.2.1 Model Cards

Researchers at Google® published Model Cards [i.40] for Model Reporting which focuses on documenting the characteristics of trained models, including their performance, intended use cases, and any relevant attributes for which performance may vary. This documentation approach serves a diverse group of stakeholders:

- ML and AI practitioners.
- Model developers.
- Software developers.
- Policy makers.
- Organizations.
- ML-knowledgeable individuals.
- Impacted Individuals.

Model cards ensure that key information about a model is documented across the following stages of the AI-system life cycle (see clause 6.4):

- Development.
- Deployment.

It employs an information sheet as its documentation technique. While Model cards provide a detailed documentation of ML models, it fails to provide documentation coverage for the broader context of data provenance and life cycle management as comprehensively as Datasheets for Datasets.

D.2.2 Method Card

In 2022, Method Cards was proposed by Adkins et al. [i.21] to support robust auditing and evaluation of ML systems through the documentation of both ML models and non-ML components like data acquisition and human-in-the-loop interfaces. These cards are primarily intended for expert stakeholders such as:

- Model developers (engineers).
- External model reviewers (auditors).

Their documentation process spans various stages of ML development like:

- Training.
- Testing.
- Debugging.

It is produced using information sheets. Method Cards can be highly technical and may not be beneficial to non-expert stakeholders.

D.3 System-Focused Documentation Approaches

D.3.1 FactSheets

In 2019, Arnold et al. [i.22], introduced a documentation approach called FactSheets for documenting AI services. An AI service according to [i.22] can be defined as an amalgam of many models trained on many datasets. This documentation approach targets the needs of multiple stakeholders:

- AI Service suppliers.
- AI Service consumers (developers).
- Standard bodies.
- Civil society.
- Professional organizations.

FactSheets cover the entire AI-service life cycle, specifically:

- Service development.
- Testing.
- Deployment.
- Maintenance.

FactSheets use an information sheet as its documentation technique and offer a broader perspective by providing documentation coverage for both, models and datasets, within a service. Furthermore, they play a vital role in providing a structured documentation framework that facilitates transparency and helps in regulatory compliance. Although FactSheets inform consumers about AI service intent and construction, they cannot prevent unintended or malicious uses of AI services.

D.3.2 System Cards

In 2022 researchers at Meta AI researched into the importance of system-level transparency in ML systems [i.33]. They proposed the System Card as a documentation approach to document and communicate various aspects of ML systems, including data, models, and decision-making processes. The aim is to enhance user trust and understanding by providing clear and accessible information about how ML systems work and their potential impacts. The methodology is aimed at the following stakeholders:

- Model developers.
- Reviewers.
- Users of ML systems.

System Cards documentation spans across the entire AI-system life cycle. However, it is more focused at providing insight into the system architecture of an ML-based system. In ~~as much~~ systems-level transparency, creating Systems cards may be tedious as it relies heavily on manual work, including crafting system diagrams and user interfaces, which requires substantial expertise to simplify technical information effectively.

D.4 Domain Specific Documentation Approaches

D.4.1 Model Facts Label

Model Facts Label [i.38] were proposed in 2020 to specifically document a sepsis prediction model for clinical settings, highlighting model name, performance and uses. This documentation approach is designed by an interdisciplinary team of:

- Developers.
- Clinicians.
- Regulatory experts.

However, the target stakeholders are:

- Clinicians.

The documentation, according to Model Facts Labels, is created, when a system with integrated ML Model is brought into operation in a clinical environment. It employs an information sheet as a documentation technique, to ensure that critical model information is accurately conveyed to the end-users in the healthcare domains. As Model Facts Label are highly specialized and tailored for clinical use, their narrow focus on a specific type of model limit their generalizability to other domains. There also remain many unanswered questions about their design and how to ensure they are accessible, intelligible, and assessable to clinicians.

D.4.2 Risk Cards

In 2023, Risk Cards [i.36] were proposed by Derczynski et al., to focus on structure assessment and the documentation of risks associated with language model applications. They address the need of:

- Inspection Organizations such as Auditors.
- AI trainers.
- Researchers.
- Policy makers.
- End users.

This documentation is carried out during the development and deployment phases of language models, using an information sheet. Risk Cards are instrumental in identifying and mitigating potential risks, enhancing the transparency of language model usage. Also, they rely on manual evaluation for detailed risk assessment, but this process is costly and may hinder adoption, especially by low-resource teams and organizations.

D.4.3 Datasheet for Subjective and Objective Quality Assessment Datasets

Barman et al. [i.37] also proposed a datasheet template to document the Quality of Experience (QoE) for 2D video streaming, addressing both subjective and objective assessments. The primary stakeholders are:

- Dataset creators.
- End users.

The documentation is facilitated through multiple formats such as Google Sheets and PDFs across the dataset life cycle. This approach ensures that QoE parameters are transparently reported, aiding in the evaluation and improvement of video streaming services. Nonetheless, its applicability is limited to multimedia contexts.

In Table D.1, a list of other existing documentation approaches is listed that were not covered in this clause.

Table D.1

Documentation Approaches	Focus
Data Statements [i.44]	Data
Data Card and Model Card for NLP [i.45]	Model and Data
Dataset Development Lifecycle Documentation Framework [i.34] and [i.46]	Data
CrowdWorkSheets [i.46]	Data
Value Cards [i.49]	Model and Method
Consumer Labels for ML Models [i.50]	Model and Method
Reward Reports for Reinforcement Learning [i.41] and [i.47]	System
Robustness Gym [i.48]	System
ABOUT ML [i.48]	System

D.4.4 Assurance Cases to document the reasoning behind other documented artifacts

Assurance Cases [i.30] are a framework to provide a structured argumentation of why a selection of evidences are considered appropriate to imply that a system is good enough to be used. It can address any requirement and is especially suitable for addressing non-functional requirements that are difficult to operationalize. Currently they are frequently used in the automotive domain to provide sufficient evidence for safety claims, but the framework is applicable to any soft requirements, like fairness or even ethics in general.

A main claim, for example a given system is fair, is decomposed into sub-claims that are either also based on the fulfilment of hierarchically structured sub-claims or that can be directly induced from evidence. Each decomposition of a claim is made explicit by an argument or reasoning step that explains the idea behind a decomposition. Furthermore, all relevant assumptions for concluding that the sub-claims imply the claim are made explicit and connected to the argument. To ease the understanding of an argument, contextual information can be attached to it as well.

In its details, the framework can be used as a pragmatic approach to come to a well-documented argument about when and under which assumptions a system is deemed good enough to be used in any terms of interest. An Assurance Case can address:

- Auditors/Reviewers
- Public authorities
- Compliance manager

By modelling the argumentation about why the evidence confirms the achievement of the objectives as an Assurance Case, the argumentation for decisions can be documented and disclosed for an external review or audit. By employing the approach before development and by automating the tests and documenting the results, this process yields the potential to provide a long-term protection against unwanted changes, for example through further training or errors when changing the code. Additionally, if similar applications have to be audited again and again, for example, in the context of banking audits, with the help of Assurance Cases, best practices can develop over time to help making well-reasoned decisions in the context of AI based applications.

D.5 Gap Analysis to EU AI Act

A comprehensive gap analysis of widely recognized AI documentation approaches with respect to the documentation requirements outlined in the EU AI Act (refer to clause 7.3.8) is discussed in the present clause. The purpose of this analysis was to assess the extent to which each documentation approach addresses the specific documentation needs prescribed by the EU AI Act, with a particular focus on coverage gaps.

Selection of Documentation Approaches:

The twelve documentation approaches were selected based on their prominence and usage within the AI community, as well as their relevance to AI governance and accountability. These approaches include well-established frameworks like Datasheets, DescribeML, Model Card, Factsheets, and others, ensuring a diverse representation of documentation practices across the AI landscape.

Mapping Information Elements:

The core of the methodology involved mapping the information elements stipulated in the EU AI Act to documentation template of each of the twelve documentation approaches. To achieve this, the relevant documentation templates associated with each approach were compiled. These templates were either sourced directly from academic literature or retrieved from publicly available GitHub repositories (if applicable). Where templates were unavailable, the official documentation, provided by the creators of the respective approaches, was referenced.

Documentation Coverage Evaluation:

Once the templates were gathered, they were systematically analysed by evaluating the inclusion or omission of each specific information element defined by the EU AI Act. The evaluation focused on three main categories of documentation requirements:

1) Data-related documentation.

2) System and model-related documentation.

3) Control-related documentation.

For each documentation approach, a binary indicator system in the analysis table was used:

- An "X" was used to denote that the approach either fully or partially addresses the corresponding information element.

- A "-" was used to indicate that the element was not addressed by the approach at all.

This binary classification allows to clearly differentiate between covered and entirely uncovered requirements, providing a straightforward overview of how well each approach aligns with the EU AI Act.

Table D.2: Assessment of State-of-the-Art Documentation Approaches in Relation to the Information Elements Defined by the EU AI Act

	Information Elements	Datasheet for Datasets	Dataset Nutrition Label	Data Cards	DescribeML	Model Cards	Method Card	Factsheet	System Card	Dataset Development Life Cycle Documentation Framework	Model Facts Label	Risk Cards	QoE Datasheet
Datasets	Provenance	x	x	x	x	x	x	x	x	x	x	-	x
	Scope	x	x	x	x	x	x	x	x	x	x	-	x
	Characteristics	x	x	x	x	x	x	x	x	x	x	-	x
	Collection	x	x	x	x	-	-	x	-	x	-	-	x
	Preprocessing	x	-	x	x	x	x	x	x	x	-	-	x
	Validation procedures	-	-	x	x	-	x	x	-	x	-	-	-
	Impact assessment	x	-	x	x	-	-	x	-	x	-	-	x

Table D.3: Assessment of State-of-the-Art Documentation Approaches in Relation to the Information Elements Defined by the EU AI Act

	Information Elements	Datasheet for Datasets	Dataset Nutrition Label	Data Cards	DescribeML	Model Cards	Method Card	Factsheet	System Card	Dataset Development Life Cycle Documentation Framework	Model Facts Label	Risk Cards	QoE Datasheet
AI system	Intended purpose	-	-	-	-	x	x	x	x	-	x	-	-
	Risks	-	-	-	-	x	x	x	x	-	x	x	-
	Version history	-	-	-	-	x	x	-	x	-	x	-	-
	Interaction details	-	-	-	-	-	-	-	-	-	-	-	-
	Version and version update requirements	-	-	-	-	-	-	x	-	-	-	-	-
	Hardware	-	-	-	-	-	-	-	-	-	-	-	-
	User interface	-	-	-	-	-	-	-	-	-	-	-	-
	Instruction for use	-	-	-	-	x	x	x	x	-	x	-	-
	Development process	-	-	-	-	-	-	x	x	-	x	-	-
	Design specifications	-	-	-	-	x	x	x	x	-	x	-	-
	System architecture	-	-	-	-	-	x	x	x	-	x	-	-
	Life cycle changes	-	-	-	-	-	-	x	-	-	-	-	-

Table D.4: Assessment of State-of-the-Art Documentation Approaches in Relation to the Information Elements Defined by the EU AI Act

	Information Elements	Datasheet for Datasets	Dataset Nutrition Label	Data Cards	DescribeML	Model Cards	Method Card	Factsheet	System Card	Dataset Development Life Cycle Documentation Framework	Model Facts Label	Risk Cards	QoE Datasheet
Controls	Human oversight	-	-	-	-	-	-	x	-	-	-	-	-
	Monitoring and control	-	-	-	-	-	-	x	-	-	-	-	-
	Accuracy	-	-	-	-	x	-	x	x	x	x	-	-
	Robustness	-	-	-	-	x	x	x	x	x	x	-	-
	Cybersecurity measures	-	-		-	-	-	-	-	-	-	-	-
	Performance	-	-	x	-	x	x	x	x	x	x	-	-
	Risk management	x	-	x	-	x	x	x	x	x	x	x	x
	Post-market evaluation	-	-	-	-	-	-	x	-	-	-	-	-
	Testing	-	-	-	-	x	x	x	x	x	x	-	-
	Privacy	x	-	-	x	-	x	x	-	-	-	-	x

History

Document history		
V1.1.1	September 2025	Publication