10

Current Challenges of AI Standardisation in the Digitising Industry

Ovidiu Vermesan[1], Marcello Coppola[2], Reiner John[3], Cristina De Luca[4], Roy Bahr[1], and Giulio Urlini[5]

[1]SINTEF AS, Norway
[2]STMicroelectronics, France
[3]AVL List GmbH, Austria
[4]Silicon Austria Labs GmbH, Austria
[5]STMicroelectronics, Italy

Abstract

The digital transformation of industrial sectors is highly dynamic, and standardisation plays an essential role in achieving the objectives set for this transformation. In this context, AI standardisation efforts and industry AI efforts are intertwined. Industrial AI applications rely on standardisation to build and sustain trust in industrial AI. Conversely, standardisation relies on industrial AI applications to play an important role in forming emerging AI standards. Although the challenges involved differ from those of similar processes in the consumer market, AI standardisation a lever for the industry's digitalisation journey. This article provides an overview of the role of AI standardisation in digitising industry and the related objectives, while presenting several requirements and challenges impacting standardisation. Furthermore, it summarises the AI standards landscape and activities within Standards Development Organisations (SDOs), outlines industrial stakeholders' approaches, and provides recommendations for an AI standardisation roadmap (in which the industry should focus on AI standards work). Its concluding remarks relate to AI standardisation activities, priorities in industrial environments, and certification efforts to conceptualise new approaches to conformance and acceptance criteria.

Keywords: Artificial intelligence, standardisation, machine learning (ML), interoperability, trustworthiness, digitising industry, AI certification, ecosystem of excellence, AI standardisation roadmap, verification, validation and testing (VV&T), industrial internet of things (IIoT), autonomous systems, safety, and security.

10.1 Introduction

The development of AI technologies and applications for industrial environments requires standards that create common building blocks establishing foundations for product differentiation, technological innovation, and frameworks for industrial stakeholders in enabling reliable, responsible, safe, and secure AI solutions.

In North America, organisations such as NIST [2] have actively supported the development of AI standards and stated, "AI standards that articulate requirements, specifications, guidelines, or characteristics can help to ensure that AI technologies and systems meet critical objectives for functionality, interoperability, and trustworthiness - and that they perform accurately, reliably, and safely."

NIST developed a roadmap on AI standards to guide the development of technical standards and related tools to support reliable, robust, and trustworthy systems that use AI technologies. NIST focus areas for standards development are outlined in Figure 10.1. While progressing in developing the roadmap, the industry responded with submissions, some of which emphasised the importance of the standards being created by ISO/IEC JTC 1/SC 42. The roadmap [22]:

- Identifies areas of strategic focus for standardisation (Figure 10.1).
- Outlines the importance of co-ordination concerning standards-setting.
- Calls for strategic engagement with international parties to 'advance AI.

In Europe, the overall strategy on AI proposes an ecosystem of excellence and trust for AI [12]. The concept of an ecosystem of excellence in Europe refers to measures supporting research, fostering collaboration between the Member States, and increasing investment into AI development and deployment [15]. The trust ecosystem is based on EU values and fundamental rights and foresees robust requirements that would give citizens the confidence to embrace AI-based solutions while encouraging businesses to develop them [14]. The European approach for AI "aims to promote Europe's innovation capacity in AI while supporting the development and uptake of ethical

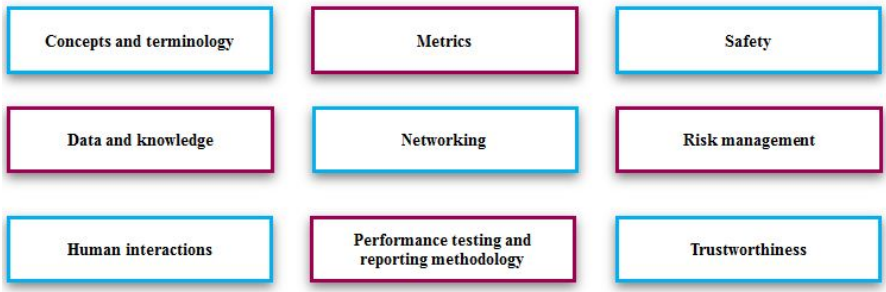

Figure 10.1 NIST focus areas for standards development.

and trustworthy AI across the EU economy. AI should work for people and be a force for good in society" [12], [13].

This article presents several issues related to AI standardisation drawn from the experience gained in the ECSEL AI4DI [1], ArchitectECA2030 [24] and AI4CSM [25] projects that addresses the challenges of digitising industry, automation of vehicles and the integration of AI-based components, techniques, methods, and applications to various industrial sectors. The project provides new reference architecture concepts, methodologies, new silicon-born-AI components supporting the development of AI- born embedded systems and integrating AI-born industrial systems, design languages, application generators, design automation and respective standardisation to accelerate the transfer of these technologies into industrial applications.

10.2 International Principles

The Organization for Economic Co-operation and Development (OECD) [4] provided a set of principles and encouraged governments to "promote the development of multi-stakeholder, consensus-driven global technical standards for interoperable and trustworthy AI" [4]. The principles proposed by OECD incorporate actionable measures to promote a framework for the "responsible stewardship of trustworthy AI", including design, development, and deployment of AI internationally. OECD's high-level value-based principles are summarised below:

- Inclusive growth, sustainable development, and well-being.
- Human-centred values and fairness require that AI-based systems are designed to respect the rule of law, defined values and diversity, and include appropriate safeguards, allowing human intervention where necessary.

- Transparency and responsible AI-based systems ensure that users understand AI-based outcomes and can challenge them.
- Robustness, security, and safety are embedded in AI-based systems throughout their life cycles by continually assessing and managing potential risks related to AI systems, including privacy, digital security, safety, and bias. AI actors should assure traceability concerning datasets, processes and decisions made during the AI system lifecycle to facilitate an analysis of the AI system's outcomes and responses to inquiry, suitable to the context and consistent with state-of-the-art.
- Accountability applies to organisations and individuals developing, deploying, or operating AI systems for the proper functioning of these systems in line with the above principles, based on their roles, the context, and consistent with the state-of-art.

The implementation of these principles is reflected in the developments of AI technology, regulations/legislation, and standards. The development of AI standards is done through SDOs that function mainly on a consensus basis.

The World Economic Forum (WEF) has strengthened the activities around the governance of AI, focusing on developing high-level principles-based guidance, frameworks, and workbooks to support decision-making and based on these activities, create partnerships with different national governments. These partnerships represent an additional valuable role in developing international standards to support the design, development, deployment, and evaluation of responsible AI systems, including within industrial sectors. The forum supports the organisation in implementing the practices and measures suggested in a Model Framework [5] and sharing experiences to inspire other organisations adopting AI to do so in a similarly responsible manner.

10.3 Role of AI Standardisation in Digitising Industry

AI, alongside IIoT, edge computing and intelligent connectivity, has become a core technology across various industries and one of the driving forces in digital transformation, and AI standardisation plays an essential role in shaping its future. AI standards are critical for building trust and confidence in AI technologies.

Standardisation activities ensure industry collaboration on the development of new AI standards, best practices, use cases and terminologies for scaling AI and enabling industries to achieve their full potential.

AI standardisation initiatives bring to industrial stakeholders common vocabularies, agreements on taxonomies and definitions, and new

pre-normative activities to address autonomous and semi-autonomous industrial systems.

AI standards form the basis for AI technologies and provide reference points for assessing AI systems' computational approaches and characteristics and studying technologies used by those systems, such as ML algorithms and reasoning, as well as their properties and features.

By analysing existing specialised industrial AI systems, stakeholders involved in standardisation processes can understand and identify the AI systems' underlying computational approaches, architectures, and characteristics.

Using representative use cases collected across application domains as a reference for emerging standards ensures that the standardisation process will reflect the contexts in which AI is being used and thus help to define AI architectural approaches.

Standardisation is expected to be a prominent driving force in the adoption and integration of AI in industrial applications. It is also expected to play a supportive role in mitigating some of the concerns and challenges brought by AI deployments in industrial environments. Moreover, the most essential requirements for AI standardisation can be naturally derived from these challenges.

10.4 Challenges Associated with AI Deployments in Industrial Environments

The challenges of AI deployments in industrial environments are associated with complexity, data acquisition and storage, training, testing, compliance requirements, high cost of failures/changes, and other variables used in the optimisation processes.

The sensors and IIoT-based systems that collect data capture many parameters from various processes, and inevitably also capture noisy information. As such, extensive storage, and computing resources for analytics capabilities are required.

To properly train AI-based systems, adequately large amounts of representative data, including information on expected and unexpected failures and other events, must be collected. This is a challenging task, as the data is available in different systems or platforms, provided in different formats and, in many cases, too scarce to be used for training purposes.

Testing AI-based systems on real-world production lines, manufacturing warehouses and other industrial facilities requires extensive time

and resources. For AI applications with a low technology readiness level (TRL), simulation environments are used for training and testing before deployment.

AI-based systems require adaptions in industrial manufacturing processes, and the cost of changes and failures at large-scale industrial facilities is very high.

Testing industrial AI applications is often required for specific deployment contexts in various industrial environments. Testing and certification bodies must depend on and increasingly trust more simulation or virtual testing to perform a conformity assessment (in addition to field testing) of industrial AI applications. AI verification, validation and testing (VV&T) approaches become essential for the safety demonstration of AI features in industrial applications.

Furthermore, since industrial environments must adhere to industry compliance requirements, changes to industrial processes often trigger extensive re-assessment of compliance, which implies a need for comprehensive VV&T of the AI-based systems and automation affected by the changes.

Manufacturing facilities and industrial systems are highly complex, often providing hundreds of parameters and inputs to AI and ML optimising algorithms. This is an enormous challenge for managing the complex AI solution space, both in terms of inference and training and learning.

Considering these challenges, the trustworthiness of organisations, products and services is critical in AI-based industrial environments. Moreover, this need for trust means that new standards for design, manufacturing and business practices must be implemented so that industrial environments can evolve and promote industry innovation and deliver reliable, responsible, safe, and secure industrial AI solutions.

Finally, the requirements and challenges of AI deployments in industrial environments must be captured in the AI standards as part of a pathway to certification for AI-based systems, products, and services. In this way, any gaps that arise between technical and ethical risks and between standardisation and certification efforts can be identified and closed.

10.5 AI Standardisation Needs in Industrial Automation

AI standardisation has a different focus in industrial applications than in consumer AI applications in terms of data quality and privacy, information content and the impact of AI on various stakeholders; therefore, it also has different needs.

In industrial AI, standardisation needs are identified and driven by use cases that are representative for various industrial sectors.

The challenge with AI standardisation lies in harmonising standardisation efforts across industrial sectors and applications to create a common set of AI requirements and standardisation needs. In this context, a differentiation needs to be made explicit between horizontal (related to generic issues across several industrial areas) and vertical (related to more specific issues relevant to a given sector or application area) standardisation tasks.

To facilitate these efforts, it makes sense to categorise the many complex AI topics based on their relevance and use, as illustrated in Figure 10.2. The AI topics are placed in a three-layer structure with generic topics on the top, horizontal topics in the middle and relevant AI4DI application areas at the bottom [1]. The generic topics form the basis for discussions on AI and include terminology, classification, methods, datasets and generic use-cases. The horizontal topics are common across industries and must be considered for the development of guidelines, standards, regulations, and certification to support AI-based systems governance in industrial environments. Ethical aspects and associated topics such as fairness, transparency, accountability, explainability, and control are part of the horizontal topics. AI is relevant for almost all industry sectors, and the application areas are very diverse, such as automotive, semiconductor, industrial machinery, food and beverages and transportation. The relevant industrial application area topics are found in the AI systems, components of AI systems and services, and manufacturing and support processes.

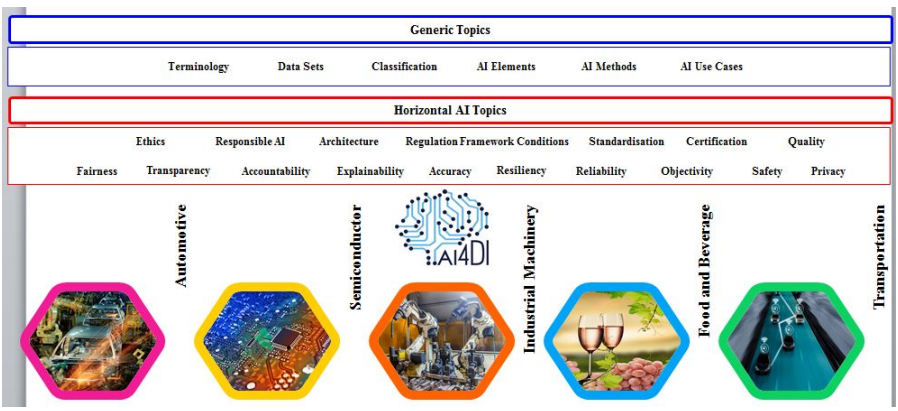

Figure 10.2 Three-layer AI topics structure: generic, horizontal, and relevant industrial application areas.

Finally, to address standardisation gaps and future standardisation activities, an interdisciplinary exchange between expert groups is needed. This exchange should focus on the role of AI in industrial environments, e.g., in the context of IIoT, functional and operational security and safety, given the complexity of AI technologies and applications.

10.6 Standardisation of Security and Safety in AI Systems

As industrial AI and ML become more and more integrated in critical systems, responsible for supporting or making decisions that can impact the security and safety of people, assets, and the environment, new challenges associated with the standardisation of security and safety in AI systems need to be addressed.

Existing safety standards in various industrial sectors are not compatible with AI methods, such as machine learning and computer vision. As such, they do not include criteria for the security and safety of AI systems or means of verification for compliance. Thus, either existing standards need to be adapted or new safety standards must emerge, or both.

Safety and security are intertwined when it comes to autonomous systems and IIoT devices, with differential approaches to address attacks against AI-based systems and services. The end-to-end and by-design principles applied to IIoT systems need to be applied to AI technologies and applications. The by-design model may be most appropriate for addressing additional concerns related to AI, such as security, safety, privacy, and inclusion.

One main challenge is to guarantee that the capabilities of AI systems, such as autonomous industrial systems and driverless vehicles, are tested before being used and monitored during operation. Physical and virtual safety validation ensures the correct and safe operation of a system in an environment. It plays a critical role in AI-based autonomous systems.

Security concerns include the protection of information within AI-based systems from unauthorised tampering, especially considering the different types of users (e.g., persons, systems, software agents, machines, IIoT devices) and levels of permission they hold. The security of AI-based systems, models, and algorithms is characterised by confidentiality, integrity, non-repudiation, accountability, and authenticity. When breached, the authenticity of data used in ML can cause significant deviations in an industrial system's outputs. In this way, accountability and responsibility are challenging to achieve for complex industrial AI-based systems if the dependencies between the system's components are not adequately identified.

AI systems could expose different kinds of security weaknesses throughout their use. However, providing security guidelines and standards-based end-to-end security, including addressing the quality of data and trained ML models, could improve the trustworthiness of AI solutions.

Safety in industrial environments is related to the use of AI-based systems and associated risks. Significant safety risks in industrial environments include ML system accidents, which can be defined as unintended or harmful behaviour that may emerge from the inadequate or faulty design or implementation of AI-based systems. Safety is also tightly linked to robustness, since robustness guarantees the proper operation of an AI-based system in each industrial context/environment.

The complexity of AI autonomous safety-critical systems often averts the use of formal verification, and real-world testing can be too complicated and lengthy during development. Simulation-based techniques are developed that consider the system under test as a black box operating in a simulated environment. Safety validation missions include the following:

- Find disturbances in the environment that cause the system to fail (falsification) by discovering previously unknown failure modes and determining regions where the system can operate safely.
- Locate the most-likely failure, based on a probabilistic model of the disturbances.
- Assess the probability of system failures.

Autonomous systems deployed in industrial environments or autonomous vehicles require inherent safety by design that starts with the design specifications, implementation strategy, and virtual validation for providing fail-operational properties and minimising residual risk by increasing the safety margin. Fail-operational safety and redundancy are achieved using redundant sensors and AI-based algorithms for safety-critical functions [23].

AI safety standards are critical for industrial processes, safety-critical applications, and new AI-based applications involving autonomy. AI-based autonomous systems are also evolving throughout their life cycles, learning new behaviours, and introducing unknown safety risks that need to be addressed with standard safety measures.

As a concluding remark, the first step in addressing this challenge is to review the legal and regulatory frameworks for security and safety-critical tasks in the industrial sectors. This will help to assess how AI will impact existing standards, as well as identify gaps. It is expected that most safety standards can be extended to cover AI methods fully or partially, until they

become too complex and difficult to use. At that point, new AI security and safety standards will need to be developed. Certification procedures will also need to be adapted. Therefore, it is important that existing and new standards are developed with the involvement of a large group of stakeholders to understand AI technology as well as industrial-specific use cases and integration at the systems level of industrial environments.

10.7 The Global AI Standards Landscape and Standardisation Activities

The development of AI standards in industrial environments requires coordinated efforts led by the industry and implemented by international standards bodies to support the global governance and alignment of AI development in the industrial sectors.

The international standards bodies have the institutional capacity to manage expert consensus and then publish AI standards across industrial sectors.

Standards shape the development and deployment of AI systems through product/service requirements and specifications for reliability, explainability, robustness, and fail-safe, fail-operational design. They influence the broader setting in which AI is researched, developed, and deployed through process and product requirements/specifications. The creation, dissemination, and enforcement of AI standards can build trust among industrial stakeholders, researchers, companies, and users.

AI standards are developed by international standards bodies which have the experience to monitor and enforce standards globally or other organisations that develop standards sponsored by different stakeholders. Examples of such development are the AI open-source software standards (e.g., software libraries TensorFlow, PyTorch, AI datasets, models, etc.,) developed by industry consortia, organisational sponsors, and individual contributors, which convert to standards across the industry over time [8]. Open-source AI enhances transparency by opening the AI black boxes and accelerating the deployment of new AI technologies, but it can bring unknown risks or negative consequences for industrial sectors.

Figure 10.3 illustrates an industrial AI standards system framework that includes the elements required and partly addressed in the existing standards and future standardisation activities.

AI national strategies confirm that several countries draft national standards and use the activities at the national level to leverage with the

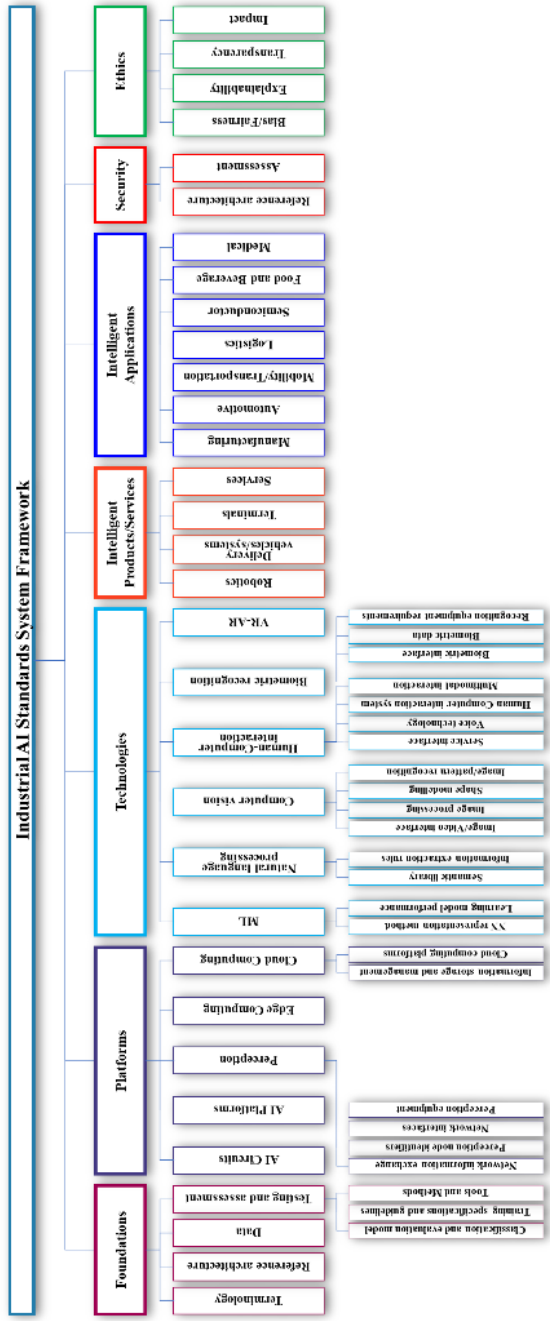

Figure 10.3 Industrial AI standards system framework.

involvement in international technical standards. Considering the market structure in the AI industry, the national standardisation bodies are encouraged and motivated to ensure that the international standards align as closely as possible to the national standards.

The following paragraphs give an overview of the AI standardisation activities covered by international standards bodies.

10.7.1 CEN-CENELEC

CEN and CENELEC continuously analyse whether relevant standards are already being produced at the international level and if European standards covering specific European needs, must be produced.

In the area of AI, CEN -CENELEC Focus Group on Artificial Intelligence has published the "Road Map on Artificial Intelligence (AI)" [10], [11] that provided an overview of existing standardisation activities in IEEE, ETSI, ISO/IEC, ITU-T and CEN-CENELEC.

The Focus Group on Artificial Intelligence addresses AI standardisation in Europe through a bottom-up approach (e.g., ISO/IEC JTC 1 SC 42) and a top-down approach (concentrating on a long-term plan for European standardisation). The Focus Group identified the following seven themes that are addressed for European standardisation:

- Mapping of current European and international standardisation initiatives on AI and identifying specific standardisation needs
- Promoting further European participation in the ISO and IEC TCs
- Formulating recommendations on the best way to address AI Ethics in the European context
- Identifying the CEN and CENELEC TCs that AI will impact
- Monitoring potential changes in European legislation
- Liaising with the European High-Level Expert Group on AI and identify synergies
- Acting as the focal point for the CEN and CENELEC TCs

10.7.2 ETSI

The ETSI community focuses on AI as a "tool" in architectural models, enhancing information/data models, redesigning operational processes, increasing solution interoperability, and data management for new ICT standards [9].

The ETSI Industry Specification Group (ISG) on Securing Artificial Intelligence (SAI) focuses on three areas: AI to enhance security, mitigate

against attacks that leverage AI, and secure AI itself from attack. ISG SAI cooperates with ENISA and have join activities. ISG SAI outputs are focusing on the following topics:

- The problem statement that guides the work of the group.
- AI threat ontology to align terminology.
- Data supply chain addressing data issues and risks for training AI.
- Mitigation strategy, with guidance to mitigate the impact of AI threats.
- Security testing of AI.
- Hardware in securing artificial intelligence.

Several other ETSI ISGs are working in the domain of ML for defining the specification of functionalities that are used in technology. A list of these ISGs is provided below:

- ISG on Experiential Networked Intelligence (ISG ENI) develops standards that use AI mechanisms to manage and orchestrate the network. The work supports making the deployment of future 5G networks more intelligent and efficient.
- ISG ZSM (Zero-touch network and Service Management) defines the ML enablers in end-to-end service and network management.
- ISG F5G on Fixed 5G defines the application of AI in the evolution towards "fibre to everything" of the fixed network.
- ISG CIM (Context Information Management) publishes specifications for a data interchange format (ETSI CIM GS 009 V1.2.1 NGSI-LD API) and a flexible information model (ETSI CIM GS 006 V1.1.1), which support the exchange of information from, e.g., knowledge graphs and can facilitate modelling of the real world, including relationships between entities.
- ISG ENI (Experiential Networked Intelligence) defines ML functionality that can be used/reused throughout the network, cloud, and end devices.

10.7.3 IEC

IEC addresses the AI through the standardisation evaluation group SEG 10, "Ethics in Autonomous and Artificial Intelligence Applications" which identifies ethical issues and societal concerns related to IEC technical activities and develops guidelines on ethical aspects related to autonomous and/or AI applications [16]. IEC's SEG 10 is consisting of two working groups:

- Autonomous and AI Applications Societal and Ethical Foundations (WG 1)

- Autonomous and AI Applications Specific Ethical Requirements (WG 2).

SEG 10 outputs are focusing on the following topics:

- Identify relevant ethical issues and societal concerns to IEC technical activities.
- Formulate appropriate recommendations to Standardization Management Board (SMB).
- Develop guidelines applicable for IEC committees on ethical aspects related to autonomous and/or AI applications.
- Assure work consistency across IEC committees and foster cooperation with JTC 1/SC 42.
- Analyse any change needed in the IEC use case template to address ethical issues and societal concerns.

10.7.4 ISO

ISO/IEC JTC 1, a joint technical committee formed between IEC and ISO on IT issues, addresses the activities related to AI terminology.

The principles and rules for drafting documents used by ISO and JTC1 [21] imply specific classifications and styles of normative language that include:

- A requirement, defined as an objectively verifiable criterion that must be met without deviation to claim conformance to the containing standards.
- A recommendation, that suggests a possible choice or course of action without excluding others.
- A permission, which conveys consent or liberty to do something. JTC 1 issued a series of International Standards on AI terminology:
    - ISO/IEC 2382-28:1995, Information technology – Vocabulary – Part 28: Artificial intelligence – Basic concepts and expert systems.
    - ISO/IEC 2382-29:1999, Information technology – Vocabulary – Part 29: Artificial intelligence – Speech recognition and synthesis.
    - ISO/IEC 2382-31:1997, Information technology – Vocabulary – Part 31: Artificial intelligence – Machine learning.
    - ISO/IEC 2382-34:1999, Information technology – Vocabulary – Part 34: Artificial intelligence – Neural networks.

All these parts are merged into the common JTC 1 standard for IT vocabulary: ISO/IEC 2382:2015 [17].

Standardisation in AI is covered by ISO/IEC JTC 1/SC 42-Artificial Intelligence, which focuses on JTC 1's standardisation program on AI and provides guidance to JTC 1, IEC, and ISO committees developing AI applications. ISO/IEC JTC 1/SC 42 topics within the work programme include:

- SC 42/WG 1 - Foundational AI standards.
  - ISO/IEC 22989: Artificial Intelligence Concepts and Terminology.
  - ISO/IEC 23053: Framework for Artificial Intelligence Systems Using Machine Learning.
- SC 42/WG 2 – Big data ecosystem.
  - ISO/IEC 20547-1: Information technology - Big data reference architecture – Part 1: Framework and application process.
  - ISO/IEC 20547-3: Information technology - Big data reference architecture - Part 3: Reference architecture.
  - ISO/IEC 24688: Information technology – Artificial Intelligence – Process management framework for big data analytics.
- SC 42/WG 3 – AI Trustworthiness.
  - ISO/IEC 24027: Information technology - Artificial Intelligence (AI) - Bias in AI systems and AI aided decision making.
  - ISO/IEC 24028: Information technology - Artificial Intelligence (AI) - Overview of trustworthiness in Artificial Intelligence.
  - ISO/IEC 24029: Information technology - Artificial Intelligence (AI) - Assessment of the robustness of neural networks.
  - ISO/IEC 23894 – Information technology - Artificial intelligence – Risk management.
  - ISO/IEC 24368: Information technology - Artificial Intelligence (AI) - Overview of Ethical and Societal Concerns.
- SC 42/WG 4 – AI Use cases and applications.
  - ISO/IEC 24030: Information technology - Artificial Intelligence (AI) – Use cases.
- SC 42/WG 5 – Computational approaches and computational characteristics of AI systems.
  - ISO/IEC 24372: Information technology - Artificial Intelligence (AI) - Overview of computational approaches for AI systems.
- SC 42/JWG 1 - Governance implications of AI.

    - ISO/IEC 38507 - Information technology - Governance of IT – Governance implications of the use of artificial intelligence by organisations.
- ISO/IEC JTC 1/SC 40 IT Service Management and IT Governance.
- SC 40/WG 1 has started work on ISO/IEC 38508 Governance of data — Guidelines for data classification.
- In addition to the above projects, several study topics are assigned to the various working groups that also include topics that cross multiple areas such as ethics, societal concerns and lifecycle that are being considered across the work programme.

The list with standards and/or projects under the direct responsibility of ISO/IEC JTC 1/SC 42 secretariat is given below:

- ISO/IEC WD TS 4213 - Information technology - Artificial Intelligence — Assessment of machine learning classification performance.
- ISO/IEC WD 5259-1 - Data quality for analytics and ML - Part 1: Overview, terminology, and examples.
- ISO/IEC AWI 5259-2 - Data quality for analytics and ML - Part 2: Part 2: Data quality measures.
- ISO/IEC WD 5259-3 - Data quality for analytics and ML - Part 3: Data quality management requirements and guidelines.
- ISO/IEC WD 5259-4 - Data quality for analytics and ML - Part 4: Data quality process framework.
- ISO/IEC WD 5338 - Information technology - Artificial intelligence — AI system life cycle processes.
- ISO/IEC WD 5339 - Information Technology - Artificial Intelligence — Guidelines for AI applications.
- ISO/IEC WD 5392 - Information technology - Artificial intelligence - Reference architecture of knowledge engineering.
- ISO/IEC AWI TR 5469 - Artificial intelligence - Functional safety and AI systems.
- ISO/IEC AWI TS 6254 - Information technology - Artificial intelligence — Objectives and methods for explainability of ML models and AI systems.
- ISO/IEC 20546:2019 - Information technology - Big data - Overview and vocabulary.
- ISO/IEC TR 20547-1:2020 - Information technology - Big data reference architecture — Part 1: Framework and application process.
- ISO/IEC TR 20547-2:2018 - Information technology - Big data reference architecture — Part 2: Use cases and derived requirements.

- ISO/IEC 20547-3:2020 - Information technology - Big data reference architecture — Part 3: Reference architecture.
- ISO/IEC TR 20547-5:2018 - Information technology - Big data reference architecture — Part 5: Standards roadmap.
- ISO/IEC CD 22989.2 - Artificial intelligence - Concepts and terminology.
- ISO/IEC CD 23053.2 - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).
- ISO/IEC CD 23894 - Information Technology - Artificial Intelligence - Risk Management.
- ISO/IEC DTR 24027 - Information technology - Artificial Intelligence (AI) - Bias in AI systems and AI aided decision making.
- ISO/IEC TR 24028:2020 - Information technology - Artificial intelligence - Overview of trustworthiness in artificial intelligence.
- ISO/IEC TR 24029-1 - Artificial Intelligence (AI) - Assessment of the robustness of neural networks - Part 1: Overview.
- ISO/IEC AWI 24029-2 - Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods.
- ISO/IEC PRF TR 24030 - Information technology - Artificial Intelligence (AI) - Use cases.
- ISO/IEC AWI TR 24368 - Information technology - Artificial intelligence - Overview of ethical and societal concerns.
- ISO/IEC DTR 24372 - Information technology - Artificial intelligence (AI) - Overview of computational approaches for AI systems.
- ISO/IEC CD 24668 - Information technology - Artificial intelligence - Process management framework for big data analytics.
- ISO/IEC AWI 25059 - Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI-based systems
- ISO/IEC DIS 38507 - Information technology — Governance of IT - Governance implications of the use of artificial intelligence by organizations.
- ISO/IEC AWI 42001 - Information Technology - Artificial intelligence - Management system.

ISO/IEC JTC 1/SC 42 has built more than 30 active liaisons with ISO and IEC committees, SDOs and industry organisations to promote cooperation and creating the industry ecosystem around AI.

10.7.5 IEEE

IEEE Standards Association (SA) has focused on the use and impact of autonomous and intelligent systems (A/IS) as they become pervasive. There is a necessity to establish societal and policy guidelines for such systems to remain human-centric, serving humanity's values and ethical principles. In this context, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems was started with a project addressing the "Ethically Aligned Design for Business: A call to action for businesses using AI" [18].

IEEE's AI standards series P7000TM address ethical considerations covering issues regarding autonomous and intelligent systems, including transparency, privacy, algorithmic bias, children's data, employee data, creating an algorithmic agent for individuals, creating an ethical robotic ontological framework, dealing with robotic nudging, creating a uniform fail-safe standard for A/IS, defining well-being metrics relating to A/IS, assessing news sources to keep them accountable and objective in reporting, creating machine-readable privacy terms for all individuals and updating facial recognition systems and databases to avoid bias. A list of the IEEE standardisation projects is presented below:

- IEEE P7000 - Model Process for Addressing Ethical Concerns During System Design.
- IEEE P7001 - Transparency of Autonomous Systems (defining levels of transparency for measurement).
- IEEE P7002 - Data Privacy Process.
- IEEE P7003 - Methodologies to address algorithmic bias in the development of AI systems.
- IEEE P7004 - Certification framework for child/student data governance.
- IEEE P7005 - Certification framework for employer data governance procedures based on GDPR.
- IEEE P7006 - Personalized AI agent specification.
- IEEE P7007 - Ontologies at different levels of abstraction for ethical design.
- IEEE P7008 - Ethically Driven AI Nudging methodologies.
- IEEE P7009 - Fail-Safe design of autonomous and semi-autonomous systems.
- IEEE P7010 - Well-being metrics for ethical AI.
- IEEE P7011 - Process of Identifying and Rating the Trustworthiness of News Sources.

- IEEE P7012 - Machine Readable Personal Privacy Terms.
- IEEE P7013 - Benchmarking Accuracy of Facial Recognition systems.
- IEEE ECPAIS - Certification for products and services in transparency, accountability, and algorithmic bias in systems.

Different other IEEE technical standardisation projects address various aspects of ML and different AI techniques:

- IEEE P2807 - Framework of Knowledge Graphs.
- IEEE P2807.1 - Standard for Technical Requirements and Evaluating Knowledge Graphs.
- IEEE P2830, Standard for Technical Framework and Requirements of Shared Machine Learning.
- IEEE P2841 - Framework and Process for Deep Learning Evaluation.
- IEEE P3652.1 - Guide for Architectural Framework and Application of Federated Machine Learning.

IEEE SA started developing an Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS), and the development is open to paid member organisations and individuals. ECPAIS seeks to develop three separate processes for certifications related to transparency, accountability, and algorithmic bias.

10.7.6 IETF

The activities related to AI are addressed by the IETF working group on "Autonomic Networking Integrated Model and Approach" [19]. With the development of the networks, it is necessary to introduce artificial intelligence technology to achieve self-adjustment, self-optimisation, and self-recovery of the network by collecting massive network state and machine learning data.

The work in IETF defined the architecture of Network Artificial Intelligence (NAI), including the key components and the critical protocol extension requirements.

IETF working group on "Autonomic Networking Integrated Model and Approach" develops a system of autonomic functions that carry out the intentions of the network operator without the need for detailed low-level management of individual devices.

Autonomic networking refers to the self-managing characteristics (configuration, protection, healing, and optimisation) of distributed network elements, adapting to unpredictable changes while hiding intrinsic complexity from operators and users.

Autonomic Networking, which usually involves closed-loop control, applies to the complete network (functions) lifecycle (e.g., installation, commissioning, operating, etc.). An autonomic function that works in a distributed way across various network elements is a candidate for protocol design. Such functions should allow central guidance and reporting and co-existence with non-autonomic methods of management.

The working group aims to enable the progressive introduction of autonomic functions into operational networks and reusable autonomic network infrastructure to reduce operating expenses.

10.7.7 ITU-T

ITU-T Focus Group on Machine Learning addresses the activities related to AI for future networks, including 5G. The working group has generated several documents covering methods for evaluating the intelligence level of future networks, data handling to enable machine learning in future networks, use cases of ML in future networks and unified architecture for ML in 5G.

A list of ITU-T documents related to AI is presented below:

- Recommendation ITU T Y.3172 - Architectural framework for machine learning in future networks including IMT-2020.
- Recommendations ITU-T Y.3173 - Framework for evaluating intelligence levels of future networks including IMT-2020.
- Y.3174 - Framework for data handling to enable machine learning in future networks including IMT-2020.
- Y.3176 - Machine learning marketplace integration in future networks including IMT-2020.
- Y.3170 - Requirements for machine learning-based quality of service assurance for the IMT-2020 network.
- Y.3175 - Functional architecture of machine learning-based quality of service assurance for the IMT-2020 network.
- Y.3531 - Cloud computing - Functional requirements for machine learning as a service.
- Y.ML-IMT2020-NA-RAFR - Architecture framework of AI-based network automation for resource adaptation and failure recovery in future networks including IMT-2020.
- Y.ML-IMT2020-serv-prov - Architecture framework of user-oriented network service provisioning for future networks including IMT-2020.

ITU-T plans to release a document on “Artificial Intelligence Standard Roadmap” [20] to assist in developing AI standards in the IT fields by

providing information about existing and under developing standards in key SDOs. In addition, it describes the overviews of AI itself and AI-related technical areas from a standards perspective, AI-related activities in SDOs, and gap analysis.

10.8 AI Certification

Certification is the process of issuing a certificate to indicate conformance with a standard, a set of guidelines or some similar norms.

Certification must have value to be accepted, successfully deployed, approved and promoted by industry.

A certification framework for AI-based systems in industrial environments can have value and provide support for the assessment and benchmark of AI-based products, services, models, algorithms for key requirements.

Producers can choose to have their AI-based products certified because they believe it will make the product more competitive.

Producers themselves may declare that their AI-based products conform to specified standards and issue accordingly a certificate referred to as self-certification or first-party certification. In other cases, a person, or an organisation with interest as a product user may require that products be submitted for certification by an independent body; this is referred to as requested third-party certification. Third-party certification is, therefore, when a body, independent of both the producer and the user, carries out the certification process.

The situation is slightly different in industrial sectors. Industrial stakeholders will not invest resources in a certification that does not achieve a goal. In other words, for certification of AI-based systems, for example, to be successful, its effect must match the stated purpose of the industrial sector.

In other cases, manufacturers of safety-critical systems may need AI-based systems certification because this is a regulatory requirement. Many industries have a regulatory authority that oversees all projects. The industry's regulations may specify that an independent third party demonstrate the conformity of a product. In this case, certification is mandatory, as opposed to the above-mentioned requested certification. This is referred to as a mandatory third-party certification.

The vast majority of AI4DI project partners agree that the standardisation goal must be to improve the efficiency of manufacturing processes and the quality of the resulting products to stay highly competitive in the global market. Furthermore, the quality embodies not only compliance with

functional requirements but also non-functional requirements. An AI-based product, system or process that failed the safety or ethical certification has not achieved its goal.

Based on the above and regardless of whether the certification is requested or mandatory, first-, second, or third-party, a common AI certification framework for AI-based systems in industrial sectors is needed. Furthermore, this AI certification framework should have the following two roles:

- To function as a quality and efficiency assessment framework during development.
- To serve as a conformity assessment framework during certification.

The AI certification framework's purpose should be to automate the procedures that support development and certification by offering standardised inspection, testing, calibration, verification and validation tools and methods. This AI certification framework would allow for many inferences using the AI algorithm under test on standardised input datasets. The results would be valuable inputs for designers and developers as well as certifiers.

In addition, the AI certification framework should have a comprehensive set of best-fit use cases for experimentation relevant to most industrial sectors (with minor adjustments) and specialised for one or several sectors.

Moreover, the AI certification must ensure that certified processes and products are more efficient and have improved quality. For instance, in the case of prediction AI systems, there must be an assurance that the prediction is as accurate as it is claimed to be.

Furthermore, virtual validation will be an essential tool, especially in autonomous systems where regulatory controls impose further qualifications for AI-based systems.

The standardised tools, AI methods, datasets, use-cases must ensure repeatability of the assessment results carried out by the same body and reproducibility of the results from assessment by different bodies.

The extent and scope of certification efforts largely depend on the AI system in question. Therefore, the AI certification framework should also include a classification scheme, allowing AI systems to be classified in desired dimensions. One such classification scheme is illustrated in Figure 10.4 and used as reference in several ECSEL JU projects such as AI4DI, ArchitectECA2030 and AI4CSM [1][24][25].

The criteria for evaluating AI systems reflect their suitability and can be uni- or multidimensional, technical, legal, or ethical, depending on the application and the application domain.

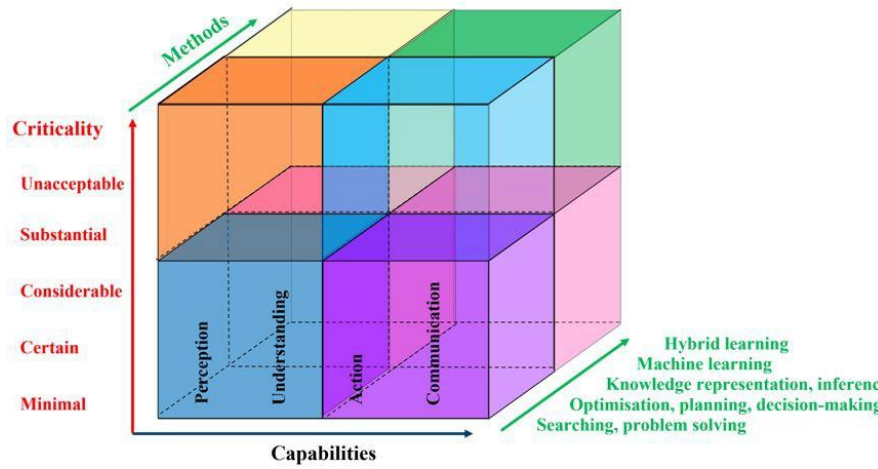

Figure 10.4 Classification scheme along with criticality, AI methods and capabilities.

One typical dimension is the potential for harm, which is commonly agreed to play a critical role in the acceptance of AI. The potential for damage can vary from minimal to unacceptable and is often related to the degree of autonomy. Other aspects, such as privacy and integrity, can be reflected through this critical dimension.

Given the wide range of capabilities of AI (from perception and understanding to communication and action), capabilities is another dimension, as the more capable the system, the greater is the risk for harm. AI methods are the third dimension, ranging from simple searching and optimisation to machine and hybrid learning. AI methods are used to achieve various AI capabilities. The more sophisticated the methods, the greater the risk.

Industrial sectors may embrace AI standardisation and certification at their own pace. But even if the ultimate goal is not to obtain a certificate, starting the design with certification in mind and using this framework towards efficient processes and high-quality AI-based products, systems, processes means the standardisation has achieved its goal.

10.9 Recommendations for an AI Standardisation Roadmap for Industrial Environments

The AI standards developments for industrial environments need to address responsible AI through standards development activities and voluntary use.

For applications in the industrial sector, AI researchers and projects that address the development of AI technologies and applications need to be involved in ongoing standardisation processes and create links with standards committees to contribute to and track outcomes. Identifying gaps in the AI standardisation landscape can benefit the development of pre-normative activities and standards with views from independent experts that provide and transfer their findings and standardisation proposal to international standards bodies under existing procedures.

In industrial environments, it is recommended that the standardisation and regulatory work concerning AI technologies and applications is progressed through multi-stakeholder discussions, allowing approaches to risk management to be tested to provide fit-for-purpose, scalability, and foster innovation.

The AI standards in industrial sectors are used to increase knowledge of reliability, trustworthiness, safety, security, and responsibility among AI developers and support the adoption of AI in different manufacturing processes.

Regulatory interventions in industrial sectors require to be proportionate to the possible and recognised harm(s) posed by AI in specific settings of the industrial sectors and identified areas of heightened vulnerability.

Different forms of certification models for AI are proposed, which involves industrial stakeholders developing the outlines of what could be recognised as responsible AI [3][6][7]. This is challenging as many large companies developed their principles for AI, which display elements of both more common values and more specific guidance elements through complementary resources.

The AI-based applications in industrial environments involve industry stakeholders and ecosystems that need best practices, standardised solutions, industry-grade benchmarking and reference data sets for training and learning. Further research is needed on industrial AI standards from technical and industrial perspectives. Technical standards desiderata can inform new standardisation efforts, and industrial strategies can develop paths for AI standards to spread in practice in different industrial sectors.

To evaluate the performance of AI-based algorithms, guidelines and reference datasets must be developed that can be used by various industrial actors in implementing AI solutions. The datasets depend on the industrial application area, and special requirements are placed on them together with guidelines that evaluate the datasets quantity/quality for training, validation, and testing.

AI and ML allow for vulnerabilities and misconfigurations, and as the manufacturing facilities are using more AI-based solutions, the more concerned they are about security risks. Open-source code is susceptible to attackers who can inject malicious code or has vulnerabilities or vulnerable dependencies.

Protecting the information in industrial environments is a crucial pillar for the performance and competitiveness of each manufacturing facility with data protection standards applied to AI systems, including training data.

All AI-based systems must integrate security by design built-in and developed around core data security principles, including encryption, logging, monitoring, authentication, and access controls. These policies must be applied even stricter considering the heterogeneous nature of AI- based solutions, including HW/SW, models, algorithms, IIoT devices and systems using open-source algorithms, commercial "black box" AI systems, or built-in AI models.

The results and outcomes from research and innovation projects with the involvement of the AI community should be aligned and provide input to the standards under development to further accelerate the advancements in AI for digitising industry. European AI projects and initiatives should dedicate efforts to understanding and engaging in standardisation processes through liaisons or partnerships with specific third-party organisations.

It is recommended that efforts be made to propose standardised AI virtual testing environments for industrial applications. These actions should include the development of standards for AI virtual testing facilities, for interoperability between AI-based digital twins and standardised AI virtual testing environments and standards for AI physical simulations/modelling (sensors, actuators, etc.).

Within industrial organisations, closer cooperation between product development units with experience in standards, industrial processes, and AI research teams can increase the efficient use of AI standards, identify the gaps, and enhance or create new standards.

Adopting AI standards under development and the involvement in activities for shaping future standards can further support the collaboration between AI research groups and the industry.

AI researchers should engage in ongoing standardisation processes. Projects addressing industrial AI should consider becoming liaisons with standards committees to contribute to and track developments. Different standards may benefit from independent development initially and then be transferred to an international standards body under existing procedures. The

involvement in AI standardisation activities support the work to create a roadmap for global AI standardisation and identify the gaps and the needs for further standardisation efforts. A roadmap is a tool for individual researchers, organisations, industrial consortia, or larger groups to evaluate the existing activities and initiate standardisation efforts in more AI-based technology and applications with priorities coming from both industry and the AI research community.

The acceleration of the digital transformation of the industry requires further research on AI standards from both technical and industrial enterprise perspectives. Technical AI standards requirements can generate new standardisation efforts, and industrial enterprise strategies can develop paths across industries in practice.

10.10 Conclusion

Building and sustaining trust in industrial AI requires developing ecosystems of industrial stakeholders that work together to define the functional and non-functional requirements for AI-based hardware, software, models, and systems; and to provide and promote reference designs and use cases employed across various industrial sectors.

In different industrial sectors, market incentives drive companies to develop product and service standards in relation to the use of AI technologies. Standards are a foundation for coordination and ensure that AI-based products and services produced across an industrial sector or different sectors are interoperable.

Standards constitute a common language and practice of communication among industry stakeholders that build guardrails that help support positive AI research and development outcomes.

The requirements for AI in industrial environments have a different focus and weight compared to those of AI in consumer and general business applications. Reliability, maintainability, explainability, safety, and security privacy are in many cases the primary concerns. Privacy, inclusion, and fairness are the specific issues addressed.

Industrial companies working with AI solutions are taking measures to protect personal information and personally identifiable information connected with deployments in the manufacturing processes.

This article presented the AI standardisation role and needs in industrial environments, derived from requirements and challenges defined and agreed upon by industrial stakeholders, provided an overview of ongoing AI

standardisation efforts, and offered recommendations for an AI standardisation roadmap for industrial environments.

The aim of this article is to encourage support for standardisation efforts in the form of improved and new representative use cases from various industry sectors and possibly spark new research topics related to AI standardisation.

Acknowledgements

Part of the work presented in this chapter was supported by the European Commission within the European Union's Horizon 2020 research and innovation programme funding, ECSEL Joint Undertaking project AI4DI under Grant Agreement No. 826060, ECSEL Joint Undertaking project ArchitectECA2030 under Grant Agreement No. 877539, ECSEL and ECSEL Joint Undertaking project AI4CSM under Grant Agreement No. 101007326.

References

[1] AI4DI (2019). Artificial Intelligence for Digitising Industry. Available at: https://ai4di.eu/

[2] NIST (2019). US Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools. Washington: NIST (US Department of Commerce), 8. Available online at: https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf

[3] Somani, A. (2019). "AI needs a certification process, not legislation". Available online at: https://venturebeat.com/2019/06/09/ai-needs-a-certification-process-not-legislation/

[4] Organization for Economic Co-operation and Development (2019). Principles on Artificial Intelligence. Paris: OECD. Available online at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[5] World Economic Forum (WEF). Model Artificial Intelligence Governance Framework and Assessment Guide. Available online at: https://www.weforum.org/projects/model-ai-governance-framework

[6] Finkel, A. (2018). "What will it take for us to trust AI?". Available online at: weforum.org

[7] Banavar, G. (2016). "What it will take for us to trust AI?". Available online at: https://hbr.org/2016/11/what-it-will-take-for-us-to-trust-ai

[8] Theben, A., Gunderson, L., López Forés, L., Misuraca, G., Lupiáñez Villanueva, F. (2021). Challenges and limits of an open-source approach to Artificial Intelligence, study for the Special Committee on Artificial Intelligence in a Digital Age (AIDA), Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg. Available online at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662908/IPOL_STU(2021)662908_EN.pdf

[9] ETSI White Paper No. #34 (2020). Artificial Intelligence and future directions for ETSI, First Edition, ISBN No. 979-10-92620-30-1. Available online at: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp34_Artificial_Intellignce_and_future_directions_for_ETSI.pdf

[10] CEN-CENELEC (2020). CEN-CENELEC response to the EC White Paper on AI. Available online at: https://ftp.cencenelec.eu/EN/News/PolicyOpinions/2020/CEN-CLC_AI_FG_White-Paper-Response_Final-Version_June-2020.pdf

[11] CEN-CENELEC (2020). CEN-CENELEC Focus Group Report: RoadMap on Artificial Intelligence (AI). Available online at: https://ftp.cenc enelec.eu/EN/EuropeanStandardization/Sectors/AI/CEN-CLC_FGR_RoadMapAI.pdf

[12] European Commission (2020). On Artificial Intelligence - A European approach to excellence and trust. White Paper. Available online at: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

[13] European Commission - High-Level Expert Group on Artificial Intelligence (2019). Ethics guidelines for trustworthy AI. Available online at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[14] European Commission (2021). Commission staff working document. Impact assessment. Accompanying the proposal for a regulation of the European Parliament and of the Council. Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative act. Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0084

[15] European Commission (2021). Rolling Plan for ICT standardization. Available online at: https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/rolling-plan-2021

[16] IEC. SEG 10 Ethics in Autonomous and Artificial Intelligence Applications. Available online at: https://www.iec.ch/ords/f?p=103:186:310164989157292::::FSP_ORG_ID,FSP_LANG_ID:22827,34

[17] ISO/IEC 2382:2015, Information technology – Vocabulary. Available online at: https://www.iso.org/standard/63598.html
[18] IEEE (2019). Ethically Aligned Design –: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems (A/IS). Version II. Available online at: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf
[19] IETF. Autonomic Networking Integrated Model and Approach. Available online at: https://datatracker.ietf.org/wg/anima/about/
[20] ITU-T. Focus Group on Machine Learning for Future Networks including 5G. Available online at: https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx
[21] ISO/IEC Directives, Part 2: "Principles and rules for drafting and structuring of ISO and IEC documents". https://www.iso.org/sites/directives/current/part2/index.xhtml
[22] NIST, 2019. 'U.S. Leadership in AI'. Available online at: https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf
[23] Vermesan O., et. al., "Automotive Intelligence Embedded in Electric Connected Autonomous and Shared Vehicles Technology for Sustainable Green Mobility. Frontiers in Future TransportationVol.2 2021. ISSN=2673-5210. https://www.frontiersin.org/article/10.3389/ffutr.2021.688482
[24] ArchitectECA2030 (2020). Trustable Architectures with Acceptable Residual Risk for the Electric, Connected and Automated Cars. Available at: https://autoc3rt.automotive.oth-aw.de/
[25] AI4CSM (2021). Automotive Intelligence for Connected Shared Mobility. Available at: https://ai4csm.automotive.oth-aw.de/