CEN-CENELEC Focus Group Report:

Road Map on Artificial Intelligence (AI)

Executive Summary

This report is the official Road Map analysis from the CEN-CENELEC Focus Group on AI. It builds on a strong consensus of over 80 experts.

The Focus Group has established an overall framework for European AI standardization, by developing a high-level vision (chapter 1.2). This vision is applicable for the whole AI ecosystem and aims at supporting the European AI industry and mitigate risks for European citizens.

The Road Map creates an overview of existing standardization activities in IEEE, ETSI, ISO/IEC, ITU-T and CEN-CENELEC (chapter 1.3 and Annex B). As part of this landscape analysis a total of 29 use cases were submitted from CEN-CENELEC TCs to the Focus Group (chapter 3.4).

Through a series of face-2-face and web meetings, the Focus Group identified 13 themes among which the following seven have been addressed for European standardization (chapter 3):

- Accountability
- Quality
- Data for AI
- Security and privacy
- Ethics
- Engineering of AI systems
- Safety of AI systems

Some of these themes are already covered in the work of other SDOs, e.g. ISO/IEC JTC 1/ SC 42 (see Annex B). But when mapping the current international activities to the themes, identified by the Focus Group, gaps do occur between the needed European standardization activities and the current work by other SDOs. An overview of potential work items is given in Annex D.

The Focus Group suggests that CEN-CENELEC adopts the following conclusions and actions:

1. The European handling of AI standardization requires a dedicated CEN-CENELEC group to be set up for the long term. A JTC (similar to the JTC for Cybersecurity) might be the most appropriate structure.
2. An initial proposal for a scope for such a JTC should be prepared by the AI Focus Group before the end of 2020. As soon as a JTC is operational, the AI Focus Group can conclude its work. It is anticipated that a number of AI Focus Group members will also play a role in a JTC.
3. The JTC should also act as a contact point for the European Commission as well as for other SDOs active in Europe in the field of AI standardization.

1. Introduction

European standards are market-driven and facilitate the smooth implementation of European policies and legislation. CEN and CENELEC have a close and well-established dialogue with the European Commission on these strategic issues.

Having a coordinated and consistent set of standards, created with the consensus of all interested parties, and adopted across the European Market instead of multiple conflicting national standards, helps significantly to ensure common levels of safety, security, and sustainability.

More than 24.000 existing European standards play a fundamental role in making the single market more efficient. By providing this support, standardization makes it easier to sell products and services across Europe and beyond, therefore improving safety, protecting consumers, reducing red tape and fostering innovation.

CEN and CENELEC have analysed whether relevant AI standards are already being produced at international level and if European standards covering specific European needs should also be produced. The work of the High-Level Expert Group on Artificial Intelligence (AI HLEG)[1] has also been taken into account, including the newly published Assessment List for Trustworthy Artificial Intelligence (ALTAI)[2]. Via the work of this Focus Group, CEN and CENELEC have included over 80 experts representing companies, consumers, trade unionists, researchers, conformity assessment bodies, member states and other societal stakeholders. The result is this European Road Map for AI standardization, which contains recommendations for the road ahead.

1.1 European standardization

European standards developed by CEN and CENELEC have an important role to play in ensuring that the European Union, the World's largest Internal Market, meets today's and tomorrow's technological and societal challenges.

Through a consensus-based and inclusive system, CEN and CENELEC are key players to support the harmonisation of Europe's internal market. One European standard is adopted identically in 34 countries, thus streamlining access to the single market, and reducing red tape. Through this unique aspect of the European standardization system, European standards support Europe's technological change while promoting European interests in international standards, thanks to their strong collaboration with and within ISO and IEC, the international standardization organisations.

The high level of convergence between the European and international standards is facilitated by the ongoing technical cooperation between CEN and ISO (Vienna agreement), and between CENELEC and IEC (Frankfurt agreement). Wherever appropriate, priority is given to international standardization, promoting the benefits of international standards to international trade and markets harmonisation, preventing the duplication of efforts, and reducing time when preparing standards. In this context, CEN and CENELEC support the international standardization and cooperation activities on Artificial Intelligence. In this field, as in ICT in general, CEN and CENELEC aim to work as the interface between international standardization and the European market's needs (business, policy, and regulatory contexts).

The European Regulation on standardization recognises the value that societal stakeholders can bring in the development of European standards and sets rules about their participation. However, these rules do not apply to ISO and IEC. It is therefore important that societal stakeholders' views are taken into account when adopting ISO and IEC deliverables in a European context.

[1] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[2] https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Recent developments in Europe, such as in the European Commission strategy on AI (COM(2018) 237), the Rolling Plan on ICT standardization, the MSP/DEI working group's recommendations and the EC's High-Level Expert Group on AI (AI HLEG), have led CEN and CENELEC to create the CEN-CENELEC Focus Group on Artificial Intelligence. The Focus Group does not develop standards but identifies specific European requirements for AI and has been tasked by the Technical Boards of CEN and CENELEC to develop a European Road Map for Artificial Intelligence standardization.

1.2 High-level vision for AI in Europe

In order to define the overall framework for European AI standardization, the Focus Group has developed a high-level vision. This vision is applicable for the whole AI ecosystem and aims at supporting the European AI industry and mitigate risks for European citizens.

The physical world is governed by mechanisms that have evolved over centuries (treaties, constitutions, policies, regulations, practices, values, etc.). The digital world, including flows of data, information, behaviour, and knowledge, should be governed by the same mechanisms.

AI standards can help economic and societal actors translate, complement, and support such mechanisms in the digital world. CEN and CENELEC should:

Ensure that AI is beneficial for citizens and society through standards that:

- **Respect fundamental values and human rights recognised in Europe**
- **Ensure appropriate governance of AI throughout the system lifecycle**
- **Ensure trustworthy (robust, safe, secure, etc.) AI**
- **Strengthen European competitiveness and benefits for society from AI**

CEN-CENELEC should use this vision when developing standards for implementation in Europe, or when adapting international standards to serve European needs.

Text box 1: High-level vision for AI in Europe

1.3 Landscape analysis

This chapter provides an overview of the international standardization landscape in the AI field. An overview of the different definitions of AI is provided in Annex A. The Focus Group has analysed the activities in the Standards Development Organisations (SDOs) that have a formal recognition by international treaties, regulation, etc. The SDOs examined in this report are IEEE, ETSI, ISO/IEC, ITU-T and CEN-CENELEC.

The Focus Group has based its analysis on the comprehensive work provided by StandICT.eu on activities of SDOs related to AI. StandICT.eu has identified the number of active work items, which are listed in the table below, related to activities per SDO [1]. The work items are separated in standardization and pre-standardization work. Standardization work items include documents like technical specifications and international standards which provide requirements and guidelines. Pre-standardization work items include documents like technical reports, road maps, and guides. While the StandICT report captures an accurate snapshot at a point of time, ongoing efforts will be required to track the rapidly growing range of standards in this domain.

	IEEE	ETSI	ISO/IEC	ITU-T	CEN-CENELEC
Standardization	14	1	6	-	-
Pre-standardization	1	2	8	2	1
Total	15	3	14	2	1

Table 1: Active work items related to activities per SDO (as of 2019).

There is extensive involvement from EU and European countries, especially in ISO/IEC JTC 1/SC 42, with 20 countries participating and holding a total of 18 leadership positions such as convenorships and editors. Annex B presents a table of the activities of the SDOs (as given of 2019) in a matrix based on themes. The main themes and the related SDOs are identified as:

- AI usage (ISO/IEC, ETSI, ITU-T, IEEE)
- Trustworthiness (ISO/IEC, IEEE)
- Transparency (IEEE)
- Ethics (ISO/IEC, IEEE)
- Foundational standards (ISO/IEC, IEEE)
- Security (ISO/IEC, ETSI)

Governance of AI and trustworthy AI are well-established activities in e.g. IEEE and ISO/IEC. Several standards are being developed now (see Annex B for an overview), and European experts are strongly represented in this work.

Vision statement	Pre-standardization activity	Standardization activity
Respect fundamental values and human rights recognised in Europe	Road Map Report (CEN-CLC FG on AI) Response to the EC White Paper on AI (CEN-CLC FG on AI) IEC SEG 10 categorisation work for AI systems and applications Some activities primarily in ISO/IEC	Some activities primarily in ISO/IEC
Ensure appropriate governance of AI throughout the system lifecycle	Several activities primarily in IEEE and ISO/IEC	Several activities primarily in IEEE and ISO/IEC
Ensure trustworthy (robust, safe, secure, etc.) AI	Several activities primarily in IEEE and ISO/IEC	Several activities primarily in IEEE and ISO/IEC
Strengthen European competitiveness and benefits for society from AI in a global context	Road Map Report (CEN-CLC FG on AI) Response to the EC White Paper on AI (CEN-CLC FG on AI) Active European participation (e.g. leadership, contribution, commenting) in many IEEE, ISO/IEC, IEC, ITU AI related work items.	Active European participation (e.g. leadership, contribution, commenting) in many IEEE, ISO/IEC, IEC, ITU AI related work items

Table 2: Mapping of the high-level vision of the Focus Group related to SDO activities (see Annex B)

1.3.1 The road ahead

Only six AI-related standards have been published to this date. These standards are mainly standards developed by ISO/IEC JTC 1/WG 9, which is now a part of the ISO/IEC JTC 1/SC 42 activities (WG 2 – Data). As stated in Annex B, a wide range of SDO activities are ongoing, mainly within the themes AI usage, trustworthiness, data quality, big data, transparency, ethics, foundational standards, and security. Since the AI standardization is a work in progress, there is an opportunity for European experts to engage and contribute to this work.

The fundamental European values and human rights and European competitiveness and benefits for society are not explicitly included in the standardization activities of the SDOs (see table 2). Some elements might be included in existing work by the European experts participating in SDO working groups, ad hocs, etc., but still the Focus Group has identified a need for a more dedicated approach to these themes.

The landscape analysis above shows the relevance of a Joint Technical Committee (JTC) under CEN-CENELEC. This Technical Committee should mirror, but not duplicate, the work of other SDOs (mainly ISO/IEC), and thereby facilitate the ongoing technical cooperation. The JTC will provide a framework for the optimal use of European resources and expertise available for standardization work; and a mechanism for information exchange between international and European Standardization Organisations (ESOs) to increase the transparency of ongoing work at international and European levels.

Potential working groups under this CEN-CENELEC Joint Technical Committee include but are not limited to:

Theme	**Current work of other SDOs which should be taken into account and in some cases mirrored**
Trustworthiness	ISO/IEC JTC 1/SC 42/WG3
Governance of AI	ISO/IEC JTC 1/SC 42/JWG1
Ethical and societal concerns	ISO/IEC JTC1/SC 42/WG3, IEC SEG 10
Safety of AI systems	ISO/IEC JTC1/SC 42/WG3
Chair Advisory Group	This group will not develop standards, so mirroring is not relevant. Inspired by CEN/CLC/JTC 13 this group should review the FG AI Road Map and asses how tasks could be taken up in the upcoming JTC work. Also, the group could maintain an overview of relevant standardization activities and do outreach to important stakeholders like the European Commission.

Table 3: Potential Working Groups (WGs) under a CEN-CENELEC Joint Technical Committee

Further proposals for European working groups and standardization items can be found in Annex D.

2. European context

2.1 Shaping Europe's digital future

The communication from the Commission on a European strategy for data (COM (2020) 66) sets a vision for EU's share of the digital economy. One of the main components of this vision is a 'European Data Space'[3], a real single market for data, compliant to EU standards and values, which can be trusted by all stakeholders participating. The AI HLEG has also identified a concept of trusted data spaces[4] in its policy and investment recommendations. Both concepts look to go beyond pure data spaces. As for most applications of AI, there is a need for data, making these two topics increasingly difficult to separate in the future. To make these data spaces work in practice, a standardization effort will be required, both on the side of AI and on the side of data spaces.

In its preliminary work, the Focus Group has discussed a potential conceptual framework named digital sphere. Digital spheres can be used to define privacy (private sphere), ownership and sovereignty but also to justify interoperability needs as digital spheres would have to interact with each other. Those three concepts (digital sphere/European data spaces/trusted data spaces) look very similar and at least very complementary. However, such concepts need proper definition and ontology (connection with other concepts) for their proper acceptance and implementation and must be considered in the general scope of digitalisation and not only specifically for AI.

The Focus Group identified the following recommended actions:

- Identify at the European Commission level what the implication of European Technological Sovereignty needs for European Standardization Organisations is.
- Set standardization directions at the European Commission level in order to address technological sovereignty and to lead in the adoption and standardization process of the future digital technologies.

2.2 Standardization and regulation working together

The European Commission has produced numerous Directives and Acts (GDPR, Product Safety Directive, Machinery Directive, European Accessibility Act, Unfair Trade Practices Act, Revised Payments Services Directive (PSD2), Directive on security of network and information systems (NIS)...), and all of them will be concerned by the development of AI technology. In this context, the European Commission is considering AI as strategic and is evaluating policy options including the review of applicable legislation. Potential adaptations or changes in European legislation might trigger standardization mandates for SDOs, the preparation of which would benefit from joint discussion with the experts involved in the SDOs.

In February 2020, the European Commission issued a White Paper on Artificial Intelligence (10 COM (2020) 65) presenting policy options. CEN-CENELEC replied to the White Paper by highlighting the role of standards in the European approach to excellence and trust in AI.

The White Paper on Artificial Intelligence is accompanied by a report assessing the implications of the emerging digital technologies on the existing safety and liability frameworks. This report aims to identify and examine the broader implications for and potential gaps in the liability and safety frameworks for AI, the IoT and robotics. The assessment of the European Union product safety legislation analyses whether the current legislative framework contains the relevant elements to ensure that emerging technologies, and AI systems in particular, integrate safety and security-by-design.

European standardization (Harmonised Standards) is also an essential element of the European Union legislation (e.g. safety, security, privacy related regulation).

[3] European data space – a genuine single market for data, open to data from across the world – where personal as well as non-personal data, including sensitive business data, are secure and businesses also have easy access to an almost infinite amount of high-quality industrial data.

[4] See also Data Trusts: A New Tool for Data Governance, by Nesta and Element AI (2019), available at: https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

The needed European standards and guidelines should mainly address trustworthiness based on the European fundamental values and human rights with breakdowns into its different characteristics: accountability, transparency, robustness, fairness, privacy, ethical and lawful use of AI.

It is acknowledged that at the moment most of the technical standardization work will be achieved by international SDOs like ISO, IEC, ETSI, ITU-T, IEEE, and others. The Focus Group notes the termination of the Category A Liaison between ISO/IEC JTC 1/SC 42 and IEEE, based on a formal request from the IEEE management[5]. SDOs must ensure efficient and active liaisons to inform each other, share expertise, and collaborate in this important area of IT standardization. It is the hope of the Focus Group that pathways to future coordination can be pursued. European organisations should actively contribute to relevant activities to ensure that the European perspectives are included in the standards. The EU should envision to launch R&D activities in support of highly relevant standardization work.

ISO and IEC standards could be adopted by CEN and CENELEC as needed, especially when they fit as European standards that support European legislation. For areas that are not appropriately covered by the international standardization work, CEN and CENELEC should start their own activities in coordination with ETSI. Indeed, international standards being developed might not take into account sufficiently or protect adequately the European values, principles, or specificities, thus requiring specific regional developments in Europe. The question of what the European specificities are, have been largely discussed by the Focus Group and will cover European societal concerns and sovereignty issues. A mechanism and/or coordination group to monitor, review, report, and advise on the AI related international standardization activities (including ISO/IEC JTC 1/SC 42) in the light of the European values, principles, and rules, should be considered.

The setting up of CEN/CLC/JTC 13 on “Cybersecurity and Data protection” is a good example of what could be organised in future at CEN-CENELEC level covering “AI and Data”. The framework of such a JTC mandate would then be:

> “Development of standards for AI and data covering all aspects of the evolving information society. Included in the scope is the identification and possible adoption of documents already published or under development by ISO/IEC JTC 1 and other SDOs and international bodies such as ISO, IEC, ITU-T, IEEE, and industrial fora. Where not being developed by other SDO's, the development of CEN/CENELEC AI and data publications for dealing with AI and data such as conceptual frameworks, management systems, techniques, guidelines, and products and services, including those in support of the EU Digital Single Market.”

Furthermore, the Focus Group has identified the following recommended actions:

- Coordination between international SDOs like e.g. ISO, IEC, ETSI, ITU-T and IEEE should be pursued. European organisations should actively contribute to relevant activities to ensure that the European perspectives are included in the international standards.
- For areas that are not appropriately covered by the international standardization work, CEN and CENELEC should start their own activities in coordination with ETSI.
- A coordination group to monitor, review, report, and advise on the AI related international standardization activities based on the European fundamental values and human rights, should be considered. This coordination group should also continue and develop the close relation between the European Commission and CEN-CENELEC that have been established during the work of the Focus Group.
- The framework and mandate of CEN/CLC/JTC 13 on “Cybersecurity and Data protection” is a good example of what could be organised in the future at CEN-CENELEC level covering “AI and Data”.
- The EU should envision to launch R&D activities in support of highly relevant standardization work.

[5] 4 March 2020, letter from Sam Sciacca, Senior Director, IEEE to ISO/IEC JTC 1/SC 42.

3. Proposed standardization activities

3.1 Introduction

This chapter contains standardization activities in the AI area that the Focus Group deems important to conduct soon. Standardization topics with high priorities are included in subchapter 3.2 "Priority standardization activities". Subchapter 3.3 lists further potential standardization activities and subchapter 3.4 copes with AI use cases.

3.2 Priority standardization activities

3.2.1 Accountability

ISO/IEC 38500:2015 [2] defines the term accountability as the state of being answerable for actions, decisions, and performance. Accountability encompasses the fulfilment of liability requirements with regard to regulations or contractual commitments (ex post), but also – in the case of AI based products or services – the provisioning of ex ante evidence for the responsible development and offering of AI based products, the responsible provisioning of AI based services, and the responsible use of such systems or services. Standards can help organisations to:

1. Demonstrate that products or services they offer or provide are trustworthy, accurate, reliable, resilient, objective, secure, explainable, safe and ensures accountability.
2. Demonstrate that they have trustworthy means of updating the system as appropriate.[6]
3. Employ good risk management strategies by providing guidance on processes and responsibilities to address accountability within the organisation.

The options above are complementary elements in establishing a relation of trust between the producer or provider of AI and its customer. Accountability should also be considered for stakeholders, which are not direct customers or consumers of AI-based products and services. A wide range of other stakeholders may be impacted by decisions made by AI. This should include bystanders harmed by an AI embodied in a self-driving car or drone, or a citizen being denied access to services. Societal level accountability is therefore needed to gain and retain the trust of these non-customer stakeholders.

3.2.2 Quality

The quality of products and services is determined by the ability to satisfy customers and the intended and unintended impact on relevant interested parties. It includes not only their intended function and performance, but also their perceived value and benefit to the customer and other stakeholders. Quality must be ensured by having appropriate general organisational processes and specific requirements and evaluation methods for the various lifecycle steps of the products and services. As the use of AI introduced new development approaches (e.g. data driven) and imposes new issues, AI specific quality standards are need both for processes, products, and services. ISO/IEC JTC1 is starting work on an AI management system standard and on quality standards for AI development and data.

ISO 9001 [3] or ISO/IEC 27001 [4] are prominent examples of a series of management system standards (MSS), which provide requirements on the management processes of an organisation. Audits can be conducted based on the management standards. ISO recognises the following forms of audits:

1st party – audit by the same organisation
2nd party – audit by a customer or supplier in a relationship with the organisation
3rd party – audit by an independent organisation

[6] This is a key area: for example UNECE has just issued a Recommendation on "Software Updates and Software Updates Management Systems" for vehicles (https://undocs.org/ECE/TRANS/WP.29/2020/80).

A clear distinction needs to be drawn in the work of TCs and SCs between standards containing specifications for a product, management system, personnel, etc., and documents setting down the operating procedures for a sector-specific conformity assessment.

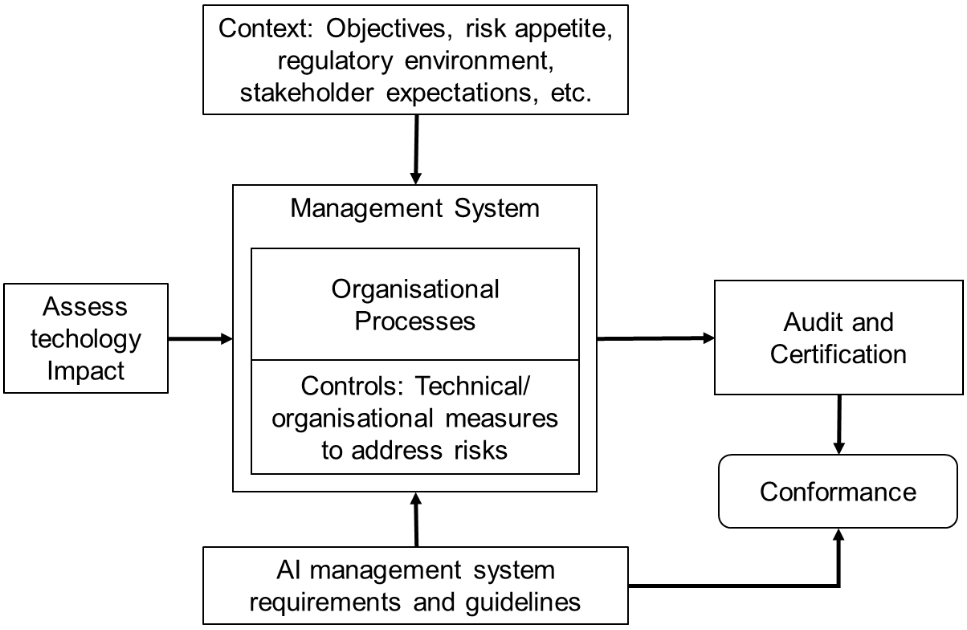

Context: Objectives, risk appetite, regulatory environment, stakeholder expectations, etc.
Management System
Organisational Processes
Controls: Technical/ organisational measures to address risks
Assess techology Impact
Audit and Certification
Conformance
AI management system requirements and guidelines

Figure 1: Conformity assessment based to provide evidence for accountability for AI.

Figure 1 shows the principle process an organisation can use for conformity assessment in regard to the development, provisioning or use of AI based products or services, based on MSS for AI. The following factors are to be considered:

1. The context of an organisation comprises its objectives, its risks (ISO/IEC 38500 [2]), the relevant regulatory environment, stakeholder expectations, etc.
2. Organisations should do an impact assessment before engaging into novel technologies to understand consequences to their stakeholders.
3. Technical and organisational requirements on the management level (or, using the ISO term, the management system) of an organisation are provided by MSS. Conformance assessment is related to the effective implementation of such requirements. Additionally, many MSS also provide guidelines on technical and organisational measures (controls) which help organisation to fulfil the requirements of MSS.

Figure 2 provides an overview of the components of a management system.

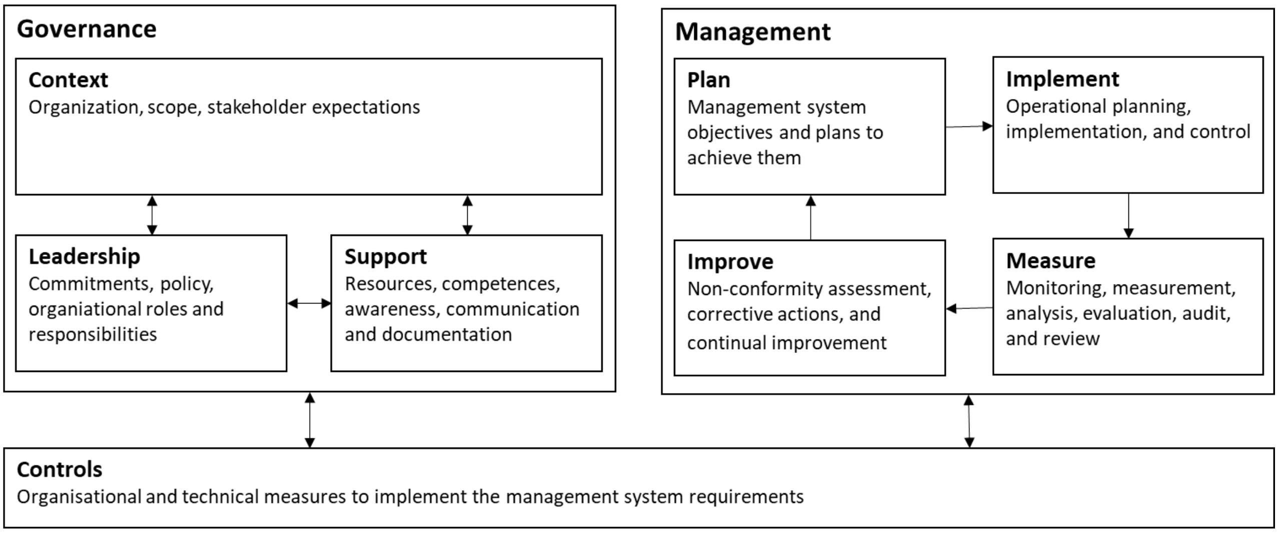

Governance
Context
Organization, scope, stakeholder expectations
Leadership
Commitments, policy, organiational roles and responsibilities
Support
Resources, competences, awareness, communication and documentation
Management
Plan
Management system objectives and plans to achieve them
Implement
Operational planning, implementation, and control
Improve
Non-conformity assessment, corrective actions, and continual improvement
Measure
Monitoring, measurement, analysis, evaluation, audit, and review
Controls
Organisational and technical measures to implement the management system requirements

Figure 2: Design and implementation of a management system ([5], modified).

Annex L of the Consolidated ISO Supplement [6] provides a mandatory codified structure and core text for all MSS. The purpose is to make it possible that the same organisational management system can easily integrate requirements from multiple MSS, for instance, on quality management, IT security management, as well as specific requirements related to AI [7].

The Focus Group identified the following recommended actions:

- ISO/IEC JTC 1/SC 42 has initiated work on an MSS for AI. It is recommended that CEN/CENELEC follows the development of the project within this SC and encourages European National Bodies within SC 42 to engage and contribute to it.
- Upon publication, CEN/CENELEC should evaluate the adoption of the standard as a European Norm.

3.2.3 Data for AI

It is necessary to identify what is in the scope or out of the scope for standardization of data. Data standards make the training, test and operating data needed for machine learning applications more visible and more usable to all authorized users. There are various current activities in data standardization (see Annex B). Data standards must preserve privacy, ensure accessibility, interoperability, and assist potential users in making informed decisions about the applicability of the data to their purpose, along with preventing misuse.

AI and data are interconnected and yet still most often addressed separately. Data spaces are different than interoperable, protected, and accessible data while being utilized by AI applications. Such an approach is necessary to develop a European trusted AI and digital ecosystem. Therefore, a global conceptual framework that takes into account not only data space, but division of labour (between storage, interoperability/portability, and AI/algorithm), algorithm and computing requirements, and the need for protection is needed.

ISO/IEC JTC1 has published and is working on various data standards including reference architecture, process frameworks and quality.

3.2.4 Security and privacy

With regard to security, standardization should study how to prevent the information system from wrong or forbidden use, modification, or abuse. Security also includes procedures where human stays in control. The Focus Group finds this issue important and relevant for standardization, and notes that further consideration is needed.

3.2.5 Ethics

For a detailed discussion see chapter 5, addressing AI ethics in the European context.

3.2.6 Engineering of AI systems

Companies that deploy AI and machine-learning in software systems experience challenges related to many facets of the system, including quality of data used for training, design methods and processes, the performance of models, deployment, testing and compliance [8]. Software engineering researchers [9], as well as guidelines such as the AI HLEG ethics guidelines for trustworthy AI [10], national standards such as the CAN/CIOSC 101:2019 Ethical design and use of automated decision systems [11], and national specifications such as the Danish PAS 2500-1:2020 on Transparency in AI and PAS 2500-2:2020 on the use of AI for decision support in public case management [12,13] make clear that the challenges are not only found in the development and test phases of software systems: New challenges are also appearing in the idea phase where the legality and ethical concerns of the use of AI must be considered, and the usage phase, where the AI systems may need to be monitored, adapted or even interrupted by other systems or humans. Researchers in law studies also point to the need for standards for placing responsibility for failures at (or removing it from) the vendor of IT systems in order to be able to compensate for or provide insurances against failures of IT systems embedding AI [14].

Sartor and Lagioia [15] have analysed for the European Parliament how the EU GDPR can coexist with applying AI on personal data, which is feasible on a principal but which is unclear on the technical level.

In order to achieve accountability, trustworthiness and transparency for AI-based software systems as described in earlier chapters, it is therefore necessary to reconsider the software engineering processes and likely develop standards for such, through the entire life cycle of software products – from inception of the idea using AI, over development, monitoring and disposal.

AI and, primarily, machine-learning is based on statistical methods and is intractable [16] for many used algorithms. ML gains its effectiveness because it does not require a formal programme specification for the learned model. But such a programme specification is indispensable for rigorous validation and verification, e.g., analysis or testing, and makes, in turn, the correctness verification of ML, i.e., algorithm and learned model, problematic [17] or infeasible. Zhang et al. [17] identify a research gap since many ML categories and tasks, as well as programme properties, have not been addressed by researchers so far. Some of the desired programme qualities of AI are difficult to specify, e.g., ethical principles [10], and demand the collaboration of interdisciplinary teams for developing such specifications or general guidelines.

3.2.7 Safety of AI systems in general and for specific sectors

Safety of AI systems in general was identified as a potential standardization activity. This work includes:

- Safety requirements for AI systems
- Safety requirements for industrial applications and consumer applications in the AI area
- Safety requirements for embedded AI (hardware, software, cloud)
- Specification of physical impact and long- term psychological impact
- Safety relevant requirements for quality of Data preparation and training

To ensure a wide use of safety standards, it is essential that it is relevant for manufacturers and other economic operators (from the AI value chain) and that it suits their needs (bottom-up approach). Also, the standards should support and provide a presumption of conformity with relevant legal requirements in existing and future EU safety legislation (top-down approach). Finally, safety standards should be aligned / compatible with international safety standards dealing with AI systems / software (global approach).

Existing European safety standards supporting EU legislation should be reviewed to be sure that protection against hazards related to the use of products implementing AI is adequate (e.g. interaction between humans and AI algorithms/systems). Special consideration has to be given to AI systems that continue learning during their usage, i.e. after their implementation/development has been completed. In this case, functionality of such an AI system is modified and may impact its safety related behaviour after the AI system is brought to the market. How to handle such AI systems from a safety point of view and how to ensure compliance to safety regulation is still open and needs further research and related standardization.

In this context, the focus of European standardization activities should be the development of a methodology and tools allowing for the verification, supervision, traceability, and recurring conformity assessment of AI systems with EU safety legislation. Also, an overview of existing safety standards is needed in order to identify the possibility to include AI aspects. The project ISO/IEC TR 5469 on functional safety under ISO/IEC JTC 1/SC 42 should be monitored. The outcome of the CEN-CENELEC Stakeholder Workshop on AI and Health in October 2020 should also be taken into account.

3.3 Further potential standardization activities

The following standardization activities were identified through several workshops in the Focus Group. All items are identified as important by the Focus Group, but during the workshop they were not identified as first priority for the European standardization work.

3.3.1 Foundational standards for AI

The work of ISO/IEC JTC 1/SC 42 on foundational standards should be taken into account before introducing any European work items.

3.3.2 Architecture

The AI architecture includes different layers, such as AI application, AI capability platform, AI framework, etc. However, standardization of AI architecture is still non-existent. The overall AI architecture lacks a unified concept and standardized definitions. This situation is not conducive to the overall development of an AI ecosystem.

It is recommended that the AI architecture should be standardized. For example, define the concept of AI architecture, describe the layers contained in the AI architecture, the hierarchical relationships (interfaces) between these layers, as well as the basic functions, input and output information of each layer in the AI architecture.

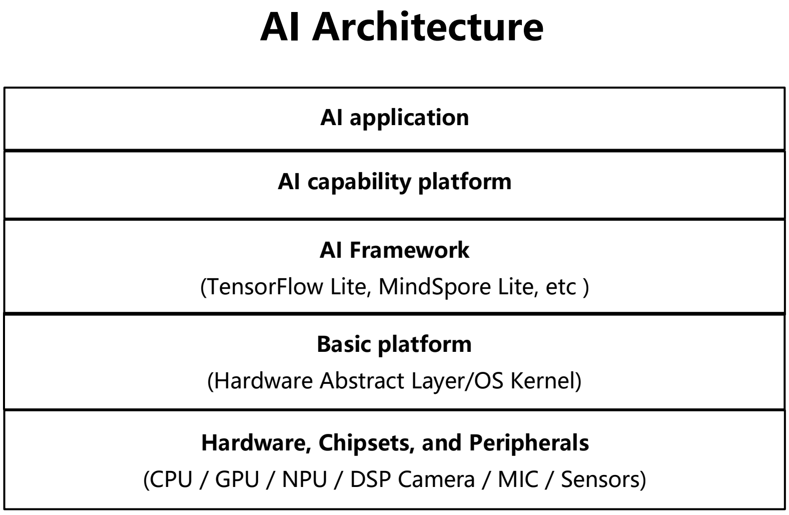

AI Architecture
AI application
AI capability platform
AI Framework
(TensorFlow Lite, MindSpore Lite, etc)
Basic platform
(Hardware Abstract Layer/OS Kernel)
Hardware, Chipsets, and Peripherals
(CPU / GPU / NPU / DSP Camera / MIC / Sensors)

Figure 3: AI Architecture

3.3.3 Other potential standardization activities

The following standardization activities were identified by the Focus Group, but not discussed in detail:

- AI-specific interoperability and portability
- Usability, inclusiveness, accessibility of AI systems
- Sustainability of AI systems, supporting the UN sustainable development goals
- AI for standardization, using AI to develop and review standards

The Focus Group recommends that all these potential standardization activities will be addressed in the future work within CEN-CENELEC.

3.4 AI use cases

A key task for the Focus Group was to identify the CEN and CENELEC Technical Committees that will be impacted by AI. That is why the Focus Group prepared a Use Case Submission Form (see Annex C), that was sent to all CEN and CENELEC Technical Committees in July 2019 with request to provide AI use cases relevant for them. A total of 29 use cases were submitted from the TCs to the Focus Group, listed in the table below. These examples of use cases are a non-exhaustive list and there are many other use cases besides the ones shown in this document. The following CEN-CENELEC TCs submitted use cases:

- JTC5 - Space
- TC61- Safety of household and similar electrical appliances[7]
- TC64 - Electrical installations and protection against electric shock
- TC134 - Resilient, textile and laminate floor coverings
- TC248 - Textiles
- TC307 - Blockchain and distributed ledger technologies
- TC332 - Laboratory equipment
- TC348 - Facility management

Overview of use cases
001. Smart energy grid
002. Smart Textile products for health, medical and wellness applications
003. AI additive management of the thread wheel machining in railway maintenance machine applications
004. Healthcare Embedded systems. Operation of safety storage cabinets and fume cupboards in laboratories
005. Garments and ensembles of garments that provide protection against heat and flame, with integrated smart textiles
006. Coherence in Standards. AI can scan a standard proposal on its coherence
007. Limited number of test methods required in standards. A target is to reduce the number of test methods
008. Transfer learning across time, sensors, and space in EO Data Exploitation
009. Robotic domestic appliances. Safety and security of autonomous domestic appliances
010. Value-based Service in Manufacturing
011. Self-organising Adaptive Logistics. Dynamic autonomous self-organisation logistics systems
012. Order-Controlled Production in Manufacturing. Automatic distribution of production jobs across supplier networks
013. Operator Support in Production. Human-technology-human interaction for assisting humans
014. Adaptable Factory. (Semi-)Automatic change of a production system's capacities
015. Platform for Agriculture. Personalised information to farmers on crop management
016. Customer relation management, Smart Information Systems, and ethics
017. SIS in Human brain research
018. Business use of IoT for surveillance. IoT-based software for monitoring and tracking customers/employees
019. Smart energy grid. This Smart grid case study explores SIS ethics
020. Insurance, Smart Information Systems (SIS) and ethics
021. Use of AI and big data systems in cyber security. Improved cyber security
022. "Drukteradar" from the municipality of Amsterdam ("crowdedness monitor")
023. Technologies that mimic people in social care. Ethical aspects of using holograms/avatars in health and social care
024. Ethics of using smart city AI and big data. SIS in urban environments
025. Predictive risk intelligence in supply chain management, insurance, finance, sustainability, medicine
026. (26/27/28) General Summary of Possible AI Use-Cases in Real Estate / Facility Management. Optimization of assets/facilities
029. Blockchain

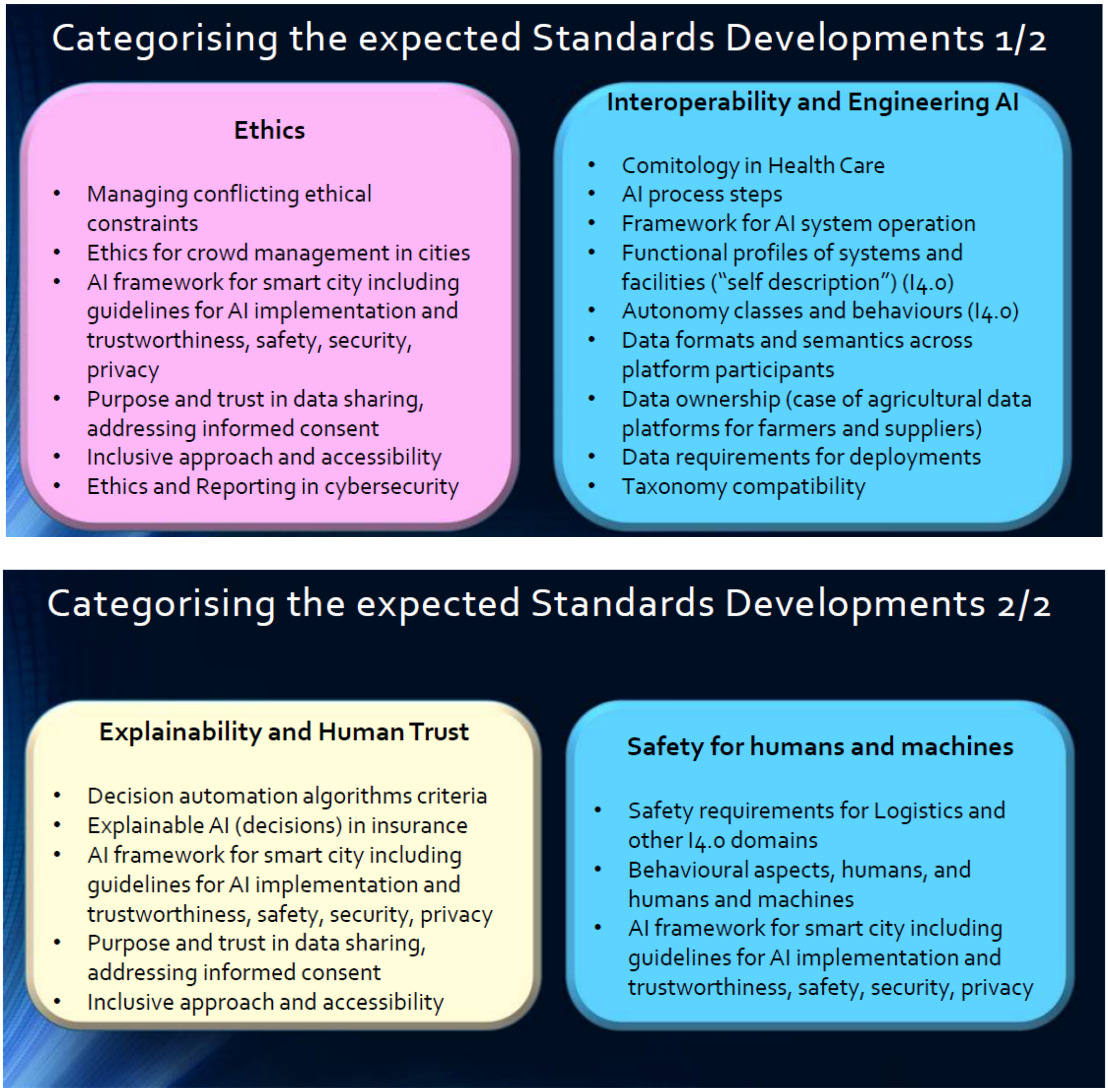
Table 4: Overview of use cases

The collected use cases vary very much in approach, maturity, and level of detail. Some are business statements, some are policy recommendations, and some have detailed reference. Standards requirements are sometimes missing, sometimes generic, and sometimes precise. The Focus Group notes that this work is a case of Commons-based Peer Production (CBPP), hence a positive mindset needs to be taken when reading them or trying to use them as input for identifying potential standardization items.

Nevertheless, the quality, depth and breadth of the 80+ use cases collected in the global exercise of ISO/IEC JTC 1/SC 42 is superior, and since many of these use cases are submitted by European entities, this work has also been taken into consideration by the Focus Group.

[7] Use cases were submitted by one member of TC 61 and do not reflect the opinion of TC 61 as a whole.

The below two figures sum up the CEN-CENELEC use cases into four categories.



Categorising the expected Standards Developments 1/2

Ethics

- Managing conflicting ethical constraints
- Ethics for crowd management in cities
- AI framework for smart city including guidelines for AI implementation and trustworthiness, safety, security, privacy
- Purpose and trust in data sharing, addressing informed consent
- Inclusive approach and accessibility
- Ethics and Reporting in cybersecurity

Interoperability and Engineering AI

- Comitology in Health Care
- AI process steps
- Framework for AI system operation
- Functional profiles of systems and facilities ("self description") (I4.0)
- Autonomy classes and behaviours (I4.0)
- Data formats and semantics across platform participants
- Data ownership (case of agricultural data platforms for farmers and suppliers)
- Data requirements for deployments
- Taxonomy compatibility

Categorising the expected Standards Developments 2/2

Explainability and Human Trust

- Decision automation algorithms criteria
- Explainable AI (decisions) in insurance
- AI framework for smart city including guidelines for AI implementation and trustworthiness, safety, security, privacy
- Purpose and trust in data sharing, addressing informed consent
- Inclusive approach and accessibility

Safety for humans and machines

- Safety requirements for Logistics and other I4.0 domains
- Behavioural aspects, humans, and humans and machines
- AI framework for smart city including guidelines for AI implementation and trustworthiness, safety, security, privacy

Figure 4: Use cases divided into four categories

4. R&D needs for AI from a standardization perspective

4.1 Rationale for AI Research leading to AI standards

AI technology is developing across the globe at a rapid pace and Europe needs to develop and strengthen its own position and role within this environment. Experience has been gained in Europe from two success stories in the early digital transformation of markets:

- Mobile communications
- Digital television

In these two cases, the resulting position for Europe is now world-class. Mobile communication is standardized at 3GPP worldwide, and television at DVB for Europe[8]. For mobile communications and for television, Europe had faced the risk to be left behind by proprietary technology. From the start, Europe had to address cross-border coordination[9], further enhanced by cooperative research projects[10]. This led Europe to global standardization leadership.

Europe has two instruments to support economic growth through AI:

- *Policy and regulation*, ensuring that AI in Europe is serving the European values and vision
- *Research and technology development*, funded by the framework programme: Horizon Europe for 2021-2025, with a research budget above 100 billion euro.

Chapter 4 addresses research and technology development in the AI area. Subchapter 4.2 identifies priority research activities which may impact on AI standardization. It also proposes mechanisms intended for use in the Horizon Europe Programme, ensuring that the whole programme, and not just those projects specialised in developing new AI methods, can have an impact on AI standardization.

Today's research on AI of any type, in any sector, usually uses data intensively, with associated AI to process it. Future Horizon Europe research projects are most likely to fall into at least one of two categories: those relevant for AI users and those relevant for AI developers. A critical mass can be achieved for this purpose at European level, and such dynamics could be helped by the mechanisms from the submission stage of research project proposals.

4.2 Recommendations for priority research activities

4.2.1 Themes for research, from a standards development perspective

The list of AI standardization items that need to be addressed includes, but are not limited to:

- AI system architecture
- Symbolic approaches to AI
- AI Data: data governance and handling
- AI and ML algorithms
- Engineering of systems using AI
- Privacy and Trustworthiness of AI
- Transparency and accountability
- Explainability of AI in particular for sub-symbolic AI/black box
- Safety and security

[8] DVB specification feed ETSI and CENELEC, under the careful management of the Joint Technical Committee on Broadcast

[9] For users to be able to roam from one country to the next, but also with spectrum coordination to maximise efficiency and minimise interference risks in border regions.

[10] As the EU research project FRAMES to shape 3G, and EU projects METIS 1 and 2 to shape 4G and 5G.

With such AI standard challenges to address AI research, including input from social sciences, standardization is best fed when casting light on:

AI Data

- The quality of data is very important for the use of algorithms and ML. It is necessary to recommend a data quality model such as ISO/IEC 25012 that defines 15 characteristics of data to be considered: accuracy, completeness, consistency, credibility, currentness, accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability, availability, portability, recoverability, recoverability
- Generic multiple task datasets

AI and Machine Learning algorithms

- Key computational problems description and ontology
- Specification of ML enabled components

Symbolic approaches to AI

- Explainability, combination of machine learning and symbolic approaches

Privacy and Security in AI

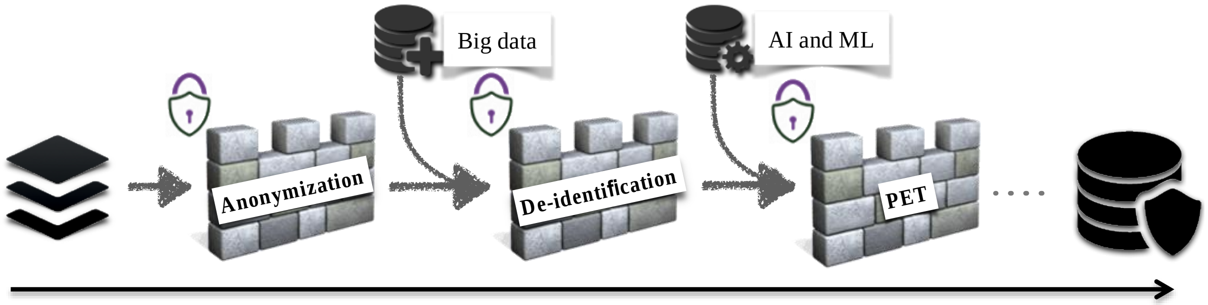


Figure 5: Evolution of privacy-preserving techniques over time extracted from [31]

Figure 5 describes the evolution of privacy protection techniques over time. The criticality of data, data analysis, and data learning approaches has required more complex privacy-preserving techniques. Federated Machine Learning FML, a distributed machine learning approach, has gained much popularity in recent years. FML distributes AI model training, e.g. to end clients (here users devices or organisations localized digital spheres), making it possible to train a global model from private local models, ensuring the privacy and security of the data. Integrating FML with traditional privacy protection techniques can further enhance the privacy and security of the AI systems.[11]

Table 5 illustrates some kinds of attacks on AI systems, mitigation strategies proposed in research community and the standards that integrated some of these research outcomes.

[11] Anonymization. Data anonymization is a type of information sanitation for privacy and data protection. The process is followed by removing personal identifiers from datasets, such that they remain anonymous. The two main functions as the basic tools for unionization are generalization (replacing the data with less precise information) and suppression (removing the identifiers from the data or replacing them with tags), that are applied to data before publishing it.
De-identification. The process that is used to prevent an individual's identity from being connected with other information in a dataset. The k-anonymity is one popular technique of this family of de-identification.
Privacy-enhancing Techniques (PET). A set of methods for protecting personal data by minimizing the possession of personal data without losing the functionality of an information system.

Attacks	Mitigation strategies	Standardization projects
Bias in data	• Reject option classification • Adversarial debiasing Metrics to identify bias: • Difference in means • Equal opportunity	• ISO/IEC PD TR 24028 • ISO/IEC NP TR 24027
Data poisoning	• Anomaly detection • Data sanitizations • Accuracy check of new joint data	• ISO/IEC PD TR 24028
Model extraction	• PATE • Misleading adversary	• ISO/IEC PD TR 24028 • ISO/IEC CD 20547-4
Evasion	• Adversarial training • Deepfool	• ISO/IEC PD TR 24028

Table 5: Popular attacks against AI systems with related mitigation strategies and standardization projects extracted from [31]

4.2.2 AI and GDPR

There have been research initiatives to develop the tools allowing for transparency and explainability of AI algorithms [18]. Also, new privacy protection techniques have been explored [19]. This is a fast-moving area, for example the new FML approach offers a privacy-by-design solution where the data privacy is preserved and the data is available for building robust AI models. FML essentially assumes that data is not available on central servers and is private, confidential, and could reside on the digital sphere. As opposed to traditional AI approach, FML intrinsically enhances data privacy as the data is never collected and transferred to other locations to be used by AI systems. However, the question of AI compliance with GDPR remains open [20]. Technical standards could typically be used to achieve such compliance. And the development of such standards could be driven by research. In this context, the following research objectives are identified:

- Extract from GDPR legal data protection requirements that AI needs to satisfy[12]
- Translate legal requirements into technical and measurable ones
- Identify and analyse relevant standards and emerging standardization efforts in order to map the technical requirements and existing good practices, and identify the gaps
- Identify existing research outcomes that could fill in the gaps and new research targets for the remaining gaps

While GDPR addresses issues common to all types of public, specific attention should be given to data protection at the workplace. More and more activities are based on workers' data: access control, geolocation, corporate social networks, connected tools, etc.

4.2.3 Definition of trustworthiness and related standardization activities

A trustworthy AI has been identified with three components by the High-Level Expert Group on Artificial Intelligence (AI HLEG): lawful AI, ethical AI, and robust AI. Robustness is being looked at by ISO/IEC JTC 1/SC 42 under activity 24029, project editors are European.

Various trustworthiness aspects and terminology are discussed under ISO/IEC JTC 1/WG 13, ISO/IEC JTC 1/SC 42 and IEEE (see Annex B).

4.2.4 AI Transparency, Explainability and Accountability

AI systems are intrinsically complex. Fully trustworthy systems can be addressed by research as a long-term goal, with standardization opportunities at every step towards this goal. Pragmatic steps can be taken to address concerns arising from use by consumers, workers, enterprise, and government. Their level of trustworthiness will determine the societal and economic acceptability of specific AI systems

[12] Relevant guidance can be found in EC WP29 document "Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679": https://ec.europa.eu/newsroom/document.cfm?doc_id=47742

prior to their deployment. Such trustworthiness level builds on three main criteria: (a) transparency[13], (b) explainability[14], and (c) accountability[15]. Self-assessment by the system of how it performs, at each step of its execution may be considered.

4.2.5 Robustness

Based on the definition proposed by IEEE (others 1990), robustness[16] is "the degree to which a system or component can function correctly in the presence of invalid inputs or stressful environmental conditions". AI systems can reinforce injustice and discrimination while making decisions, or they may be polluted by poisoned data which can turn the system into an abusive tool for adversarial goals. In addition, the trained AI model can also be the target of an adversarial attack such that one can extract the properties of the model. Another popular adversarial attack against the AI robustness is called evasion that is to manipulate the system boundaries of classifiers towards misclassifications (see also table 5 in chapter 4.2.1).

Typical examples for robustness goals are:

- meet one or more thresholds over a set of statistical metrics that need to hold on some evaluation data
- invariance of performance against certain types of data perturbations
- invariance of performance against systematic changes of input data (e.g. drift, change of operating scenario)
- stability of training results under small variations of the training set, given the stochastic nature of many machine learning algorithms
- consistent system output for similar input data, e.g. resistance to so-called adversarial examples

It should also be noted that the European Union Directive 2016/1148 "Concerning measures for a high common level of security of network and information systems", already brings in legislation, which covers "the ability of network and information systems to resist, at a given level of confidence, any action that compromises the availability, authenticity, integrity or confidentiality of stored or transmitted or processed data or the related services offered by, or accessible via, those network and information systems". The interpretation of this directive suggests a direct correlation with this chapter and therefore any consideration given to the text of this chapter should be done so in conjunction with the directive.

Depth versus Robustness: a trade-off?

Traditional machine learning models with few parameters (shallow models) are better suited to meet robustness goals than complex (deep) models. However, machine learning using Deep Neural Networks (DNN) is attributed to most of the success of AI in recent years. Deep models can be effectively trained while yielding superior generalisation. The complexity of deep models poses a risk in terms of robustness.

ISO/IEC JTC 1/SC 42 has started to work on the "assessment of the robustness of neural networks". First results suggest that the state-of-the-art in statistical and empirical methods may be a reasonable candidate standard for the assessment of robustness. Formal methods for the verification of deep neural networks or for the assessment of their robustness are still at an early stage.

[13] Transparency can be increased by providing tools for analyzing AI systems: Analysis of machine learning models and visualization of the multiple layers of AI/ML processes (structure, dynamic and function) will increase trust. As such, an AI system can be seen as a complex network.

[14] Explainability can be increased by providing reporting mechanism on the process and reasoning of the AI system as well as requiring description components in the AI.

[15] Accountability is increased by using concepts of self-regulatory AI: Self-regulated by AI refers to learning that is contains elements of (a) Self-observation i.e. monitoring tasks performed;

[16] Robustness is often used as a general term for describing properties that are required for the acceptance of new high-stakes AI applications. Robustness is a term that is being used in many recent research papers on machine learning. However, the meaning of robustness is case-specific, given that certain industries, organisations or applications will require different robustness goals and specific metrics to determine if a goal is met.

Standardized methods for the assessment of the robustness of neural networks would open the way for DNN deployment in safety critical domains, thanks to the availability of appropriate tests for robustness. Minimizing the risk for producing a non-robust DNN solution in the development phase of a system is another motivation for robustness standards.

In summary, when looking at deep neural networks and their usage in industry, there are mainly two areas of research that have not produced sufficiently effective results yet; (1) assessment of the robustness properties of deep neural networks by formal methods and (2) architectures and training methods for robust solutions based on deep neural networks.

5. Addressing AI ethics in the European context

5.1 Introduction

The concern for ethical AI is a global phenomenon. With more and more actions and decisions being driven by AI it is vital for humanity that AI conforms to certain values and principles such as respect for human dignity, freedom, equality, solidarity, citizens' rights, and justice. For Europe, the High-Level Expert Group has proposed a set of such ethical principles that AI needs to reflect in development, training, and operation[17].

It would not be appropriate to standardize ethics in itself. As an example, a rule such as "it is preferable for AI to seriously harm 100 people compared to killing 1 person" (or vice versa) would be impossible to agree on even within one country, let alone across Europe. The standardization community is not the right place to discuss fundamental ethical dilemmas. More generally, the **judgement** of what is ethically acceptable and what is not acceptable, typically depending on application, purpose, jurisdiction etc., remains a **political decision**. Political decisions are rendered through legal tools (laws, directives, regulations, etc.) while standards are used to help common understanding and definitions of the issues.

However, the **description of ethical characteristics** of AI systems as well as the **description of ethical risks** of AI in a given scenario can be standardized. As will be explained in this chapter, this might support political decision making as well as market transparency, consumer protection and procurement processes.

5.2 Ethics and risk

It is important to note that AI systems must satisfy different requirements for a given characteristic (such as safety, trustworthiness, fairness, or transparency) in different scenarios. For example, when an AI system may cause serious consequences for personal safety, user privacy or property (e.g. in self-driving vehicles or smart healthcare), it must satisfy higher requirements than when it is used e.g. for optimising industrial processes, although even such purposes might pose ethical challenges. Therefore, minimum ethical levels to be reached are usually dependent on the field of application and usage scenario. In some cases, the same AI system can be deployed in contexts with different ethical sensitivities.

When understanding ethics to imply the principle of "Do no harm"[18], then harm should not only be interpreted as physical injury, damage to health or damage to property, but also include negative impacts on an individual or a group's dignity, autonomy, privacy or its civil and societal rights. While such impacts would include physical harm, they extend far beyond this and would therefore not be addressed by a risk assessment of AI systems focussed only on physical harm.

5.3 Need for continuous stakeholder engagement

Any approach by an organisation to address the ethical issues of its use of AI systems must:

1) Address the distinct ethical concerns of different stakeholders including value chain partners, workers, consumers, local communities, broader society, and future generations (impacted by environmental and societal impacts).
2) Accommodate ongoing engagement with stakeholders, both before and after the deployment of an AI application, to assess how they are affected by it and how negative risks can be treated.
3) Provide mechanisms to resolve differing views held by different stakeholders on the assessment and treatment of risks associated with a given use of AI. This mechanism must engage with appropriate representatives of different affected stakeholder types, providing them with
    a. transparent access to risk assessments and how they were arrived at,
    b. access to expertise in assessing the efficacy of any risk treatments,

[17] Chapter I; https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

[18] Note that there are broader understandings, e.g. "promoting happiness", especially in other regions.

    c. clear accountability on who is responsible for risk assessments and for the effectiveness of risk treatments.

4) Accommodate existing legislation by reviewing and mapping the ethical issues to any currently available legislation for instance European convention on Human Rights or General Data Protection Regulation (GDPR). Any map will show where AI is currently accommodated and highlight standardization gaps that need to be addressed.

5.4 Standardizing the description of ethical characteristics of AI systems

Much of the discussion on a standardized way of demonstrating AI ethics has revolved around "labelling". This has caused confusion since different audiences understand the term "labelling" in different ways. In particular, an "AI ethics label" is sometimes thought of as a simple sticker on an AI system that claims it is ethical, comparable e.g. to the Fair Trade label for social sustainability or the Blue Angel for environmental sustainability. AI ethics have too many dimensions and is too context-dependent to be captured in this simple yes/no manner. A simple "AI ethics label" would open the door to marketing-driven ethics washing and would fail to provide clarity for consumers.

A more useful approach might be a **standardized datasheet of the ethical characteristics** of an AI system (which may or may not be called a "label"). Inspiration can be taken from the well-established energy efficiency rating. It establishes levels A to G for the energy consumption of different categories of electrical devices and has proved to be useful and acceptable to consumers, industry, and regulators alike. A standardized datasheet for AI ethics could look somewhat similar although with a broader set of information[19]. The datasheet could e.g. reflect the degrees to which the AI system implements the seven key requirements for trustworthy AI that have been identified by the High-Level Expert Group[20,21,22].

It is debatable whether such a standardized datasheet should be called a "label" at all, and it might be preferable to avoid the term altogether. Some members of the Focus Group have proposed a hybrid approach, where a symbol on an AI system indicates that the above datasheet is available for the system, possibly together with a QR code linking to it.

The realisation of a standardized datasheet of ethical characteristics of an AI system poses several challenges that need to be addressed by the standardization community:

1. Products are already subject to various labelling and/or certification schemes – some of them voluntary, some of them compulsory (especially in regulated sectors as well as for characteristics such as safety). Adding a description of ethical characteristics does provide important additional information but it also adds complexity and might be seen as diluting existing schemes such as the CE marking or other industry focussed schemes as existing in medical device, automotive, aerospace and other industry sectors. Therefore, it needs to be analysed whether the proposed datasheet of ethical characteristics is best implemented as a new scheme that stands on its own, or whether it should be integrated with one of the existing schemes.

[19] Energy consumption is the one category that is displayed on the label regardless of the appliance. Depending on type of appliance (refrigerator, washing machine, etc.), additional information is stated on the label, e.g. washing or drying performance.

[20] Chapter II; https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

[21] Existing proposals from industry fora such as the IEEE and the Partnership for AI could also provide suggestions for the content of the data sheet, while ISO/IEC JTC1 work on AI terminology and concepts, AI bias, neural network robustness, AI risk management, AI governance, AI system engineering, and AI ethical and societal issues provide a source of further information and process standards against which datasheet-related standardization can be assembled. The proposed SC42 Management System Standard also offers a potential route to certification of datasheet generation and verification processes.

[22] https://aifs360.mybluemix.net/

2. A standardized way of showing ethical characteristics of an AI system assumes that categories such as "Transparency" or "Fairness" are made measurable. Work on "making ethics measurable" is ongoing in the academic community in a variety of directions, most recently e.g. around a concept of "augmented utilitarianism" [22] or a VCIO (values, criteria, indicators, observables) framework [24], all of which have strengths and weaknesses in different types of systems in different situations [21]. There has also been work on bias in ISO/IEC JTC 1/SC 42, but this needs to mature further before being sufficiently solid and detailed to be cast into a standard and will require extensive engagement with stakeholder representatives with an interest in AI ethics and trustworthiness and its impact on different sectors of society.

3. Users or operators might not be able to verify measurements of complex characteristics such as "transparency" and "fairness" on their own. This means that a high level of trust is required in such a rating scheme, in the accuracy of the measurements and in the standardized measures being used[23].

4. Not every AI system is meant to be used in ethically sensitive applications. In particular, there is at least anecdotal evidence that the majority of industrial use cases of AI might not pose significant ethical risks. Therefore, a standardized way of quickly identifying these non-sensitive use cases (and thus minimizing red tape for companies) needs to be developed. It could be based on various risk classification metrics that have already been proposed [10].

5. The specific scope, form and detail in the datasheet might need to be adapted to the needs of different stakeholders and types of systems. This adaptation process itself will also need to be standardized so that the integrity and consistency of datasheet information provided to different types of stakeholder can be audited and verified. Separate datasheets appropriate for technology developers or end-users might be needed.

6. The specific scope, form and detail in the datasheet might need to be adapted to the needs of different stakeholders and types of systems. This adaptation process itself will also need to be standardized so that the integrity and consistency of datasheet information provided to different types of stakeholder can be audited and verified.

7. Given the self-learning nature of many AI systems, the characteristics shown in the datasheet might need to be updated, and this needs to be addressed e.g. through processes surrounding the datasheet.

5.5 Perspective of stakeholders and global viability

If and when the above challenges have been resolved, then a standardized datasheet of ethical characteristics can be provided in a form that is verifiably adapted to meet the needs of a broad range of stakeholders:

- For scenarios with significant ethical risks, **regulators** can restrict the use of AI to systems that
  (1) come with a standardized short datasheet, and
  (2) satisfy minimum levels in each category.
  This is a clean way of keeping the **description in the hands of standardization**, and the **judgement in the hands of regulation**.

- **Consumers** can use the datasheet to compare AI products and services and make informed decisions what is acceptable to them and/or worth spending money on. Moreover, consumers can trust that they are protected by minimum regulatory requirements, and that any information asymmetries (e.g. through the evolution of an AI system) are properly handled.

[23] By contrast, for the established energy efficiency label, users or operators could more easily verify the accuracy of the expressed rating by making their own measurements of energy consumption or noise levels, for example.

- **Manufacturers** of AI systems can gain market advantage by demonstrating relevant ethical characteristics of their products and the transparency and accountability of their AI risk management and governance processes in a meaningful way that is recognized worldwide. This applies in both B2C and B2B settings.

- **Purchasers** (both in private and public sector procurement) can use the format of a datasheet to create clear minimum specifications, and they benefit from market transparency and risk management for AI procurement processes. As **deployers and users** of AI systems, they have a recognized way of demonstrating ethical behaviour.

A similar concept has already been discussed in IEC SEG 10 (Ethics in Artificial Intelligence and Autonomous Systems Applications) where there is strong participation from Asia and in particular China. There are clear indications that the overall approach outlined above has the potential to be adopted internationally rather than just within Europe. Again, the key factor here is the separation of the description of an AI system from the judgement of its acceptability in a given context.

5.6 Role of a process perspective on AI ethics

The ethical characteristics of an AI system cannot be fully understood without taking into account the various processes during development, training, and operation of the system. A process perspective is helpful in several different ways:

- Processes play an important role for evaluation/verification: When analysing an AI system in order to fill in its short datasheet as outlined above, the levels reached in each category can only be partially determined from characteristics of the system itself. As an example, the top-level A for “Fairness” might, amongst others, indicate that all training data sets meet certain statistical tests. However, in addition, it might indicate that the AI manufacturer has conducted and documented a certain **process**, e.g. an external stakeholder dialogue, to agree what aspects of fairness are relevant for this system and how they could be demonstrated. Similarly, robust version control of algorithms might constitute a process criterion.

- Processes tie AI ethics into existing management system and risk management frameworks[24]: Just like security or sustainability, AI ethics need to be supported by appropriate management system and/or risk management frameworks. These frameworks typically take a process perspective, and they are often designed to certify the manufacturer or operator as an organisation rather than a specific product[25]. By requiring external consultation and/or auditing, these processes can be designed to contain safeguards against a lack of compliance and transparency, e.g. in cases where ethical dilemmas should not just be discussed and addressed internally.

A standardized datasheet for the ethical characteristics of AI products and services may provide a means to inform consumers about the ethical considerations undertaken in its development. In fact, such a datasheet would document evidence of compliance to standardized risk management, transparency, and accountability processes as well as specific results from these processes for the AI application in question in a standardized form.

There are information technology governance standardization projects that are equally applicable to AI as they are to other technologies such as ISO/IEC 38507 “Information technology - Governance of IT - Governance implications of the use of artificial intelligence by organisations”. It may be necessary to augment governance standards with AI considerations or to create new standards in this regard, as already being undertaken by the ISO/IEC JTC1 joint working group between SC40 and SC42 (JTC 1 SC 42/JWG 1).

[24] See Annex B for an overview of ongoing work in SC 42

[25] Certifying organisations can also to some extent help to avoid the difficulties of defining the boundaries of an AI system.

5.7 Remarks on specific aspects of AI ethics for Europe

While it is preferable to leverage international standards as the basis for making ethical characteristics of AI systems measurable, it is acknowledged that regions may develop frameworks to allow for region-specific needs. This approach has been used successfully for concerns other than ethics e.g. limiting radio emissions. Here, an international standard specifies the measurement method, but individual countries specify the limits[26].

In addition to differences in value judgements at the European level, some further examples for the relevance of a specific European view have been discussed in the Focus Group, including:

1. Multiple working languages and the role of translation. Naïve machine translation, based on corpora that reflect historical prejudices and biases, will frequently get genders wrong or reflect biases [25]. We note that at least one automatic translation system has recently moved some way towards correcting such bias. But the fundamental problem remains, and the same system, for example, translates the Romanian "profesor" as English "professor", and the equivalent female "profesoara" as "teacher".

2. A greater willingness to regulate individuals and companies for the good of society: largely environmental (e.g. forbidding individuals to burn coal in London in 1956), but also in areas like gender discrimination. Consider a company that wishes to hire IT people, who in many countries are 90% male. If it is advertising on a pay-per-view basis, then it is more cost-effective to target the advertisements at men only. It is preferable that a human being would not make that decision, but an AI system optimising efficiency will do so unless prevented [26]. Similar issues in indirect gender discrimination in the insurance market occur [27]. AI systems can support dynamic pricing for many more items and services, and these can create many invisible biases [28].

3. In many practical ways, European countries differ from other parts of the world. Europe tends to have national databases of speed limits, and current navigation systems, which have no visual sensors, warn human drivers of speed limits by mapping location to this database. USA has no such equivalent database, so driving assistance systems have to read speed limit signs, which opens them up to adversarial attacks[27]. Equally, the revulsion against facial recognition in USA is caused by a combination of what seems to be racially biased technology and racially biased armed police forces.

[26] International standards for how to measure radio frequency emissions levels from, and the susceptibility of, ICT equipment are set by the IEC (committee IEC/CISPR). Such standardized test methods are subsequently adopted in many regions globally as the basis for supporting region-specific radio interference regulations e,g, by CENELEC/TC 210 in Europe. However, TC 210, under mandate from the EC to publish standards in support of the EMC Directive, while adopting the IEC/CISPR standards for test methods, also publishes standards that specify acceptable levels of emissions & susceptibility that differ from other regions. As a comparison, the US FCC does not specify any requirement for equipment susceptibility.

[27] https://www.mcafee.com/blogs/other-blogs/mcafee-labs/model-hacking-adas-to-pave-safer-roads-for-autonomous-vehicles/

6. Issues around conformity assessment and certification of AI

6.1 Introduction to conformity assessment

Formalized conformity assessment is usually performed based on specific schemes, which are a set of rules and procedures that describes the objects of conformity assessment, identify the specific requirements, and provide the methodology for performing conformity assessment. Standards are often the base for such schemes as they can provide requirements and assessment methodologies that are agreed by a wide number and variety of stakeholders. The more international standards are used in schemes, the more likely the schemes will be globally accepted. Regulation often refers directly to standards or conformity assessment schemes.

6.2 AI specific issues around conformity assessment

Conformity Assessment of development, deployment and usage of AI based solutions can foster the trust of the various stakeholders, based on specific trustworthiness topics or a general approach.

For several trustworthiness topics like safety, security, privacy, and bias already various EU legislation exist, defining general requirements like fundamental rights, consumer law and product safety. Furthermore, also more specific regulations exist like the GDPR and Machinery Directive. They are not technology specific and normally also apply when AI solutions are used for products that fall in the specific category. However, they may have shortcomings related to specific risks due to the use of AI, and the existing regulation may not apply to certain AI based solutions. Updates and extensions of existing regulation are therefore considered, which also may require new conformity assessment schemes and related standards.

An AI management system standard, as currently considered by ISO/IEC JTC 1/SC 42, could assure that organisations take AI specific issues and concerns of all stakeholders into account when developing, deploying, and using AI based solutions. A management system standard defines requirements for the processes of the organisation to assure, for example, that the needs and expectations of all interested parties are taken into account and that related risks and opportunities are assessed. However, they do not define specific requirements, limits, and measures for products to achieve, for example, the required safety. Conformity to such an AI management system standard only provides a general trust that the organisation develops and uses artificial intelligence responsibly. It also needs to be considered that virtually every organisation is using AI in some tool or service, and that the proposed AI management system standard therefore could be relevant for all these organisations that are using AI.

For providing product specific conformity assessment related to the different trustworthiness topics, specific requirements and related verification, validation, and testing (VV&T), methods have to be defined. Safety, security, privacy, and general quality standards with related requirements are already covered by various standards, and they might be extended to cover AI specific issues. Bias/fairness and transparency are new topics for which completely new standards might be needed.

Especially for verification, validation, and testing (VV&T) certain AI technologies pose new challenges. The concrete behaviour of neural networks is hard to explain and may show unexpected behaviour for certain input data. Safety instrumented functions require a high level of reliability and robustness. If it cannot be verified that an AI solution achieves the required level of robustness it cannot be used in such a system. While standardization activities on VV&T of neural networks are starting in ISO/IEC JTC 1 /SC 42, it is still an area of active research, and there are limitations to conformity assessment methods (verification, validation, testing, and audit).

For AI systems that learn during their usage, the challenge is that the behaviour of the system changes continuously and would require a new conformity assessment each time. It is an open question if such kind of “real-time” in service conformity assessment is possible as the assessment usually requires a specific test equipment.

Furthermore, re-assessment of just the AI component is usually not enough as changes in its behaviour may impact the whole system behaviour, and conformity assessment is related to the whole system including for example hardware components. An alternative approach is that the system itself ensures that modifications of its functionality due to self-learning have no negative impact on assessment topics like safety or fairness.

It should also be noted that the need for conformity assessment depends on whether the AI component is part of the critical functionality. In safety-critical systems, safety instrumented and intended functions of the systems are often separated. If AI is used only for the intended functions it should not be an issue as the safety instrumented function ensures correct behaviour from a safety point of view. However, in complex systems, such as autonomous vehicles, this traditional separation between safety instrumented and intended function is no longer the case and an intended function might be safety relevant. The vision capability of an autonomous vehicle should be considered; this either does or does not detect (with some certainty) a pedestrian, and already machine learning is invoked here.

6.3 Conformity assessment issues for complex AI functions embedded in critical systems

More and more organisations are already relying on simulation for design and conformity assessment of their products or services because it could be the only way to efficiently perform some optimisation trade-offs.

As general conformity assessment principles, leading in some cases to certification, are well defined, complex AI systems may pose new issues regarding their specifications and their conformity assessment schemes.

Indeed, AI systems act upon data coming from their open operating environment which complexity multiplies exponentially the scenarios where proper functioning will have to be assessed. Consequently, most of the conformity assessment may be based on simulation in addition to field testing.

The scheme for conformity assessment using simulation would then rely on digital twins of the final AI system. Such an approach would require to clearly specify the operating domain and to model the operating environment, which may even be standardized (e.g. a self-driving car operating environment). A series of scenarios (set of events) would then be run using both the digital twin of the AI systems and the operating domain model. This is raising the question of the validity of the operating domain model. Several issues will therefore have to be addressed:

- For systems that need certification (e.g. transportation safety regulated systems), does the certification body need its own operating domain model?
- Are operating domain or product specific?
- How are operating domains for AI systems specified? How are they assessed?
- Should operating domains be standardized in some cases?
- Is an operating domain model part of the specification for safety regulation?
- Do industry and testing third-party operating domain models need certification when used for safety validation?
- Are there any interoperability issues between digital twins and operating domain models?

Furthermore, questions concerning the validity of the dataset used for training/learning, whether initial or continuous, may be addressed through simulation as well. Obviously, this whole set of issues may need to be addressed by AI experts, digital twin experts, conformity assessment experts, safety regulatory experts in a joint working group. The general question would be: What are the implications of the use of simulation for conformity assessment (from testing to certification) in addition of field testing?

6.4 Criticality of AI (risks, high risk sectors, etc.)

Like other technologies AI is diverse. Made up of many unique components, a critical aspect of an AI system is how the system is deployed (where, why, what are the system's objectives).

How are different AI examples compared? This is very difficult as they could have very little or even nothing in common and vary in their impact to users or environments. For instance, comparing an AI algorithm that has a focus on geospatial positioning, to keep a self-driving car on the road against a music service that recommends the latest release of a music track to a user would be like comparing the throwing technique of an athlete specialising in javelin with a sprinter. They both have common characteristics, but the requirements and deployment are different.

Criticality of AI can be interpreted as those areas that have the potential to have significant impact on human lives or the environment, whether it be directly or indirectly. This is about ensuring the reduction of risk with an emphasis on safety and security, be it physical or related to the digital domain. Although criticality is not limited to these areas and can include bias and privacy.

Therefore, it is important to look more closely at the AI related specifics before it is possible to make any judgment. This would suggest that some degree of categorisation could be needed, to help focus on the important characteristics of any given use case.

There are many different ways to perform categorisation, starting with a bottom-up approach identifying many of the different attributes (data, process, algorithm, product) or starting higher up and dividing by domain or application (financial, automotive, supply chain). The specifics of AI, at a low level, extend across multiple higher-level domains, so by combining AI components, they become more unique and less likely to fit inside any specific model. In a similar way to 'data policy' and the 'mosaic affect' AI components, correctly developed in isolation, according to an accepted standard may have different and unexpected results when deployed in combination with each other.

To make progress, standardization of AI technology or related situations could be reviewed based on their perceived criticality. Noting GDPR article 35 "Data protection impact assessment", which has a similar goal, the principles defined there could provide a similar framework for use in AI.

End users/stakeholders want products that they can trust, that are safe and that limit risk to them, physically and in the digital domain. In the absence of formal regulation, for a manufacturer to be open and transparent from the outset, manufacturers will assist in delivering end-user confidence. Whilst work is done towards a greater understanding and appreciation of the impact of AI on individuals, it is possible to set out a high-level process to assist in reaching a satisfactory level or trust and safety in AI.

1. Identify work on AI where standardization makes sense, for instance bias. This would establish the core elements to be built on where manufacturers, producers and providers can come together and voluntarily adopt a specific level or direction.
2. Anticipate societal concerns/requirements that could result in the creation of standards to address these. Followed by self-certification/declaration.
3. Where AI is used in areas that could adversely affect human life or the environment, specifically occupational safety and health in the case of professional use (critical or high-risk sectors), third-party certification could be appropriate, as in the case of current medical devices.

Determining the criticality level of an AI system is particularly important for conformity assessment, as it can influence the expected requirements and even the assessment procedure. Indeed, when conformity assessment is carried out by the manufacturer itself, it is generally that the regulator considers that the item being evaluated presents a low risk and low complexity. When the level of criticality is higher, the regulator requires the intervention of a third-party compliance assessment body. The latter must be impartial, independent of the organisation being assessed (Art R17.3 of Decision 768/2008/EC) and may not engage in any activity that would call into question its impartiality, typically a conflict of interest (Art R21.2.c of Decision 768/2008/EC).

Given the European Commission White Paper[28] scope of "risk", which goes beyond pure safety issues and includes items such as privacy, ethics, etc., an "assessment of risk/criticality guide" might be necessary. This guide may define different levels of risk, each leading to a certain conformity assessment scheme. Given that it may include the specific European approach of risk/criticality, it is most likely that the establishment of such a guide might be a European specificity.

The question on how Europe should deal with risks on sovereignty has been discussed in detail within the Focus Group, but no consensus was reached. The Focus Group recommends that the discussion on this topic will be continue in the Focus Group or a dedicated JTC on AI.

6.5 European specificity

In view of its history of urbanisation, Europe has tended to regulate behaviour (by individuals or companies/groups) that affects society more than some other countries do. There are early examples, but pollution regulation (Clean Air Acts and similar) is a clear example. AI is capable of damaging society by reinforcing stereotypes and historic biases, e.g. by not showing job advertisements to certain classes of people [29]. Since Europe is a multilingual society, the Focus Group notes similar reinforcement in machine translation [30].

[28] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

7. Conclusion

The AI Focus Group has identified and analysed a broad range of aspects of AI and outlined their relevance to Europe. This report provides guidance and recommendations in a number of areas but also shows that there is not yet a “finished” description of how to handle AI standardization at a European level. As a context for the more detailed recommendations in this report, the Focus Group therefore suggests that CEN-CENELEC adopts the following conclusions and actions:

1. The European handling of AI standardization requires a dedicated CEN-CENELEC group to be set up for the long term. A JTC (similar to the JTC for Cybersecurity) might be the most appropriate structure.
2. An initial proposal for a scope for such a JTC should be prepared by the AI Focus Group before the end of 2020. As soon as a JTC is operational, the AI Focus Group can conclude its work. It is anticipated that a number of AI Focus Group members will also play a role in a JTC.
3. The JTC should also act as a contact point for the European Commission as well as for other SDOs active in Europe in the field of AI standardization.

References

[1] Ziegler, W. (2020) ‘A Landscape Analysis of Standardization in the Field of Artificial Intelligence’. Journal of ICT, Vol. 8_2, 151–184. River Publishers

[2] ISO/IEC 38500:2015 Information technology — Governance of IT for the organisation

[3] ISO 9001:2015 Quality management systems — Requirements

[4] ISO/IEC 27001:2017 Information technology — Security techniques — Information security management systems — Requirements

[5] Institute of Risk Management (2018) ‘A Risk Practitioners Guide to ISO 31000:2018’. IRM Report. Link: https://safety4tradies.co.nz/wp-content/uploads/2019/06/IRM-Report-ISO-31000-2018-v3.pdf

[6] ISO (2018) ’The Integrated Use of Management System Standards (IUMSS)’. ISO Handbook

[7] ISO/IEC Directives, Part 1 (2020) ‘Procedures for the technical work’. Sixteenth Edition. Link: https://www.iso.org/sites/directives/current/part1/index.xhtml

[8] Bosch J, Crnkovic I, Olsson HH. (2020) ‘Engineering AI Systems: A Research Agenda’. ArXiv preprint arXiv:2001.07522

[9] Kästner C, Kang E. (2020) ‘Teaching Software Engineering for AI-Enabled Systems’. ArXiv preprint arXiv:2001.06691

[10] High-Level Expert Group on Artificial Intelligence - AI HLEG (2020) ‘Ethics Guidelines for Trustworthy AI’. Link: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[11] CAN/CIOSC 101 (2019) ‘Ethical design and use of automated decision systems’. Link: https://url12.mailanyone.net/v1/?m=1k5N75-0002n6-53&i=57e1b682&c=BI-PMu3GRWaBuUqJBQGekOfooMbKik2_N3NqpMv96uLUlyP7vfMFLMC095Hq5GkiBQMXc6zEQv_NAg9S5jn6HfwyGHm16XOCBAyhIadxR213x_D-9x9sen3tMiFvEmfuocg4YStObhLlpQuyRF_KO1BvoXC_8P-_RSlVLf9LdzQh_oBT-44zFo_HCguYgL2qH_ch8KNeNxw-SSS-ah16MVI_GahB6EojcfZ923qPClWFR7UM-i9SbgVZ_Ee6IiWQ

[12] PAS 2500-1:2020. Kunstig intelligens – Del 1: Gennemsigtighed, Danish Standards

[13] PAS 2500-2:2020. Kunstig intelligens – Del 2: Beslutningsstøttende anvendelse i offentlig sagsbehandling, Danish Standards

[14] Rompaey, L. (2020) ‘Discretionary Robots Conceptual Challenges in the Legal Regulation of Machine Behaviour’. PhD thesis, Faculty of Law, University of Copenhagen

[15] Sartor, G. and Lagioia, F. (2020) ‘The impact of the General Data Protection Regulation (GDPR) on artificial intelligence’. European Parliamentary Research Service. Link: https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2020)641530

[16] Finley T, Joachims T (2008) ‘Training structural SVMs when exact inference is intractable’. Proceedings of the 25th international conference on Machine learning, 2008 Jul 5 (pp. 304-311)

[17] Zhang JM, Harman M, Ma L, Liu Y. (2020) ‘Machine learning testing: Survey, landscapes and horizons’. IEEE Transactions on Software Engineering

[18] Edwards, L. and Veale M. (2017) 'Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for'. Duke L. & Tech. Rev. 18

[19] Carlini, N. (2019) 'Recent Advances in Adversarial Machine Learning'

[20] Panel for the Future of Science and Technology of the European Parliamentary Research Service (2020) 'The impact of the General Data Protection Regulation (GDPR) on artificial intelligence' Scientific Foresight Unit (STOA), PE 641.530-June 2020

[21] Aliman, N.M. and Kester, L. (2019)'Requisite Variety in Ethical Utility Functions for AI Value Alignment'. IJCAI AI Safety Workshop 2019

[22] Aliman, N.M., Kester, L.(2019) 'Augmented Utilitarianism for AGI safety'. International Conference on Artificial General Intelligence. Springer.

[23] Aliman, N.M., Kester, L., Werkhoven, P., Yampolskiy, R. (2019)'Orthogonality-Based Disentanglement of Responsibilities for Ethical Intelligent Systems'. International Conference on Artificial General Intelligence. Springer.

[24] Report of the AIEI Group (2020) 'From Principles to Practice - An interdisciplinary framework to operationalise AI ethics'. Link: https://www.ai-ethics-impact.org/resource/blob/1961130/c6db9894ee73aefa489d6249f5ee2b9f/aieig---report---download-hb-data.pdf

[25] Caliskan,A. and Bryson,J.J. and Narayanan,A. (2017)'Semantics derived automatically from language corpora contain human-like biases', Science 356 pp. 183-186

[26] A. Datta, M.C. Tschantz, and A. Datta. Automated Experiments on Ad Privacy Settings. Proceedings on Privacy Enhancing Technologies, 2015:92-112, 2015

[27] McDonald,S. (2015) 'Indirect Gender Discrimination and the 'Test-Achats Ruling': An Examination of the UK Motor Insurance Market'. Presentation to Royal Economic Society, April 2015. https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=RES2015&paper_id=791

[28] Lu, D. (2020) 'Uber and Lyft pricing algorithms charge more in non-white areas'. Link: https://www.newscientist.com/article/2246202-uber-and-lyft-pricing-algorithms-charge-more-in-non-white-areas/#ixzz6PmrztgUH

[29] Speicher,T., Ali,M., Venkatadri,G., Nunes, R.,F., Arvanitakis,G., Benevenuto,F., Gummadi,K.P., Loiseau,P. & Mislove,A. (2018) 'Potential for Discrimination in Online Targeted Advertising' Proceedings of Machine Learning Research 81(,pp. 1-15.

[30] Caliskan,A., Bryson,J.J. & Narayanan,A. (2017) 'Semantics derived automatically from language corpora contain human-like biases'. Science 356, pp. 183-186

[31] Dilmaghani, Saharnaz, et al. 'Privacy and security of big data in AI systems: A research and standards perspective.' 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019.

Annex A

Overview of definitions of Artificial Intelligence (AI) – as of 2019

EC AI White Paper: “AI is a collection of technologies that combine data, algorithms and computing power.” Later refined by claiming that AI is the combination of the first two, i.e. *data* and *algorithms*.

ISO/IEC/IEEE: artificial intelligence (AI) is a branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement [ISO/IEC 2382:2015, Information technology — Vocabulary; ISO/IEC/IEEE International Standard - Systems and software engineering--Vocabulary," in *ISO/IEC/IEEE 24765:2017(E)* , vol., no., pp.1-541, 28 Aug. 2017]

ISO/IEC 2382
artificial intelligence
capability of a functional unit to perform functions that are generally associated with human intelligence such as reasoning and learning
Note 1 to entry: artificial intelligence; AI: term, abbreviation and definition standardized by ISO/IEC [ISO/IEC 2382-28:1995].
Note 2 to entry: 28.01.02 (2382)
[SOURCE: ISO-IEC-2382-28 * 1995 * * *]

artificial intelligence
interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning
Note 1 to entry: This is an improved version of the definition in ISO/IEC 2382-1:1993.
Note 2 to entry: artificial intelligence; AI: term, abbreviation and definition standardized by ISO/IEC [ISO/IEC 2382-28:1995].
Note 3 to entry: 28.01.01 (2328)
[SOURCE: ISO-IEC-2382-28 * 1995 * * *]

artificial intelligence
branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement
Note 1 to entry: artificial intelligence; Al: term, abbreviation and definition standardized by ISO/IEC [ISO/IEC 2382-1:1993].
Note 2 to entry: 01.06.12 (2382)
[SOURCE: ISO-IEC-2382-1 * 1993 * * *]

Dartmouth Artificial Intelligence Conference 1955: “the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving”

AI100 Stanford: “Artificial Intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.” [“Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence,” 2016, http://ai100.stanford.edu/2016-report.]

WEF: “Artificial intelligence – systems that act by sensing, interpreting data, learning, reasoning and deciding the best course of action.” [WEF Empowering AI Leadership, an oversight toolkit for boards of directors https://spark.adobe.com/page/RsXNkZANwMLEf/]

OECD AI Experts Group (AIGO): “An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments. It uses machine and/or human-based inputs to perceive real and/or virtual environments; abstract such perceptions into models (in an automated manner e.g. with ML or manually); and use model inference to formulate options for information or action. AI systems are designed to operate with varying levels of autonomy.” [OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en.]

OECD Network of Experts on AI (ONE AI), which had its first meeting on 26-27 February 2020 and will support the OECD AI Policy Observatory, will re-visit the question of defining AI. One initial suggestion coming out of the MIT Internet Policy Research Initiative is to pursue an approach that combines the following:

1. Allow the definition of “AI” to remain vague, along the lines of “automated systems that have some level of adaptive, decision-making or other capacity typically associated with human intelligence” without closer definition of “adaptive”, “decision-making” and “intelligent”.
2. Ensure that all substantive policy discussion relating to AI specifies and focuses on particular modalities of AI, e.g. “use of machine learning for automated decision-making in the financial sector”.

Sources:
IEEE Ethically Aligned Design Glossary https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_glossary.pdf
OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en.
WEF Empowering AI Leadership, an oversight toolkit for boards of directors https://spark.adobe.com/page/RsXNkZANwMLEf/
https://thenextweb.com/artificial-intelligence/2017/09/10/glossary-basic-artificial-intelligence-terms-concepts/
ISO online browsing platform https://iso.org/obp

Annex B

Overview of standardization activities related to SDOs - as of 2019

	IEEE	ISO/IEC	ETSI	ITU-T	CEN-CENELEC
Usage	P7013	SC42 WG4	ENI ISG ZSM ISG	ML5G	
Security		SC27 WG4	SAI ISG		
Foundational standards	P7009	SC42 WG1			
Trustworthiness	P7002 P7011	SC42 WG3 JTC 1/WG13			
Ethics	P7000 P7008 P7007	SC42 WG3 IEC SEG 10			
Personalised AI	P7006				
Transparency of autonomous systems	P7001 P7003				
Wellbeing metrics	P7010				
Transparency of data processing	P7004 P7005				
Privacy	P7012				
Big data		SC42 WG2 SC7 WG6 liason to SC42			
AI governance		SC42 JWG1			
Computational approaches		SC42 WG5			
AI for health				AI4H	
Conceptualization and specification of domain knowledge					

Table is based on "A Landscape Analysis of Standardization in the Field of Artificial Intelligence" by Wolfgang Ziegler, *Journal of ICT, Vol. 8_2,* 151–184. *River Publishers, April 2020,* with a few additions from Focus Group experts. For more information on the standardization activities above please refer to the website of each SDO:

IEEE: https://standards.ieee.org/project/
ISO: https://www.iso.org/standards.html
IEC: https://www.iec.ch/standardsdev/
ETSI: https://www.etsi.org/standards
ITU-T: https://www.itu.int/en/ITU-T/Pages/default.aspx
CEN-CENELEC: https://www.cencenelec.eu/Pages/default.aspx

Annex C

Use Case Submission Form to CEN-CENELEC Technical Committees

ID	(leave blank, for internal use)	
Use case name		
Application domain	(Select from pull-down menu)	
Deployment model	(Select from pull-down menu)	
Status	(Select from pull-down menu)	
Scope[29]		
Objective(s)[30]		
Narrative	Short description (not more than 150 words)	
	Complete description	
Stakeholders[31]		
Stakeholders' assets, values[32]		
System's threats & vulnerabilities[33]		
Standardization opportunities/ requirements		
Challenges and issues		
Societal Concerns[34]	Primary social concern	(Select from pull-down menu)
	Description	
SDGs[35] to be achieved (select primary goal)	(Select from pull-down menu)	

[29] The scope defines the intended area of applicability, limits, and audience.

[30] The intention of the system; what is to be accomplished?; who/what will benefit?.

[31] Stakeholder are those that can affect or be affected by the AI system in the scenario; e.g., organisations, customers, 3rd parties, end users, community, environment, negative influencers, bad actors, etc.

[32] Stakeholders' assets and values that are at stake with potential risk of being compromised by the AI system deployment – e.g., competitiveness, reputation, trustworthiness, fair treatment, safety, privacy, stability, etc.

[33] Threats and vulnerabilities can compromise the assets and values above - e.g., different sources of bias, incorrect AI system use, new security threats, challenges to accountability, new privacy threats (hidden patterns), etc.

[34] Societal concerns within this context are considerations that come into play when choosing a technology or recommendations on its usage / deployment that affect the outcome in a socially, ethically or business undesirable way. Examples are considerations regarding trustworthiness, privacy, accountability, robustness, bias, etc. For further inspiration see https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines

[35] The Sustainable Development Goals (SDGs), also known as the Global Goals, are a collection of 17 global goals set by the United Nations General Assembly. SDGs are a universal call to action to end poverty, protect the planet and ensure that all people enjoy peace and prosperity.
URL: http://www.undp.org/content/undp/en/home/sustainable-development-goals.htm

Annex D

Proposed standardization items

The Focus Group has identified a preliminary set of important standards per European needs. Some of those standards are already addressed by ISO/IEC or even by a dedicated group (e.g. the AI HLEG from the EU). Others are not addressed yet and some specific items may probably be addressed by a European standardization approach as they covered key societal European concerns (ethics, sovereignty, respect of the law, etc.).

From this set of standards of European interests/needs, recommendations are proposed covering R&D, pre-standardization, or direct standardization activities. As industrial, R&D, pre-standardization and standardization activities unfold, other gaps may be identified, and the following list of recommendations will have to be reviewed and updated on a regular basis.

Proposed standardization items:

Area	Name of item	Standardization body	Comments
Terminology/ Foundations	Description of scopes of AI standardization and regulation	New TC under CEN-CENELEC	In support of EU regulation
Terminology/ Foundations	Horizontal levels of automation/autonomy	To be determined	In support of EU regulation
Trust(worthiness)	AI and Data Management System	Already covered in SC42	Follow SC42 work
Trust(worthiness)	Quality and accuracy of training data	Partly covered in SC42	Follow SC42 work
Trust(worthiness)	Trusted Data Space	New TC under CEN-CENELEC	Coordinate with GAIA-X, Industrial Data Spaces etc.
Trust(worthiness)	Assessment list(s) for aspects of trustworthiness	New TC under CEN-CENELEC	Starting from AI HLEG work (ALTAI)
Trust(worthiness)	Ontology of Trustworthiness	ISO-IEC/JTC1/WG13	Follow JTC1 work and verify whether it is covering European needs
Trust(worthiness)	Explainability, verifiability	To be guided from a standardization perspective by CEN-CENLEC JTC	Pre-standardization R&D to be funded by EU
Trust(worthiness)	Robustness	Already covered in SC42	Follow SC42 work
Trust(worthiness)	Data quality management	Already covered in SC42	Follow SC42 work
Ethics	Summary description of the ethical properties of AI systems	New TC under CEN-CENELEC	Alternatively: Reference to IEC SEG 10
Ethics	Categorisation of ethical risk levels of AI application scenarios	New TC under CEN-CENELEC	
Security	(several items)	To be addressed in CEN-CENELEC JTC	

Safety	(several items – both safety of AI and AI for safety)	To be addressed in CEN-CENELEC JTC	
Safety	Implications of the use of simulation for conformity assessment (from testing to certification) in addition of field testing	To be guided from a standardisation perspective by CEN-CENLEC JTC	Pre-standardisation
Safety	Description of AI system operating domains	To be determined	AI systems must be assessed in a defined/standardized operating domain
Resilience & Sovereignty	Framework for regional digital sovereignty (including sovereignty ontology)	New TC under CEN-CENELEC	
Resilience & Sovereignty	Impact of sovereignty on standardisation	To be addressed in CEN-CENELEC JTC	
Resilience & Sovereignty	Framework for digital identity of data for AI	To be addressed in CEN-CENELEC JTC	
Respect of the law	Data Jurisdictions	To be addressed in CEN-CENELEC JTC	In support of EU regulation
Respect of the law	Data protection in AI from a European (GDPR) perspective	To be addressed in CEN-CENELEC JTC	In support of EU regulation
Respect of the law	Permission to use data - Methods to manage access and reuse data	To be addressed in CEN-CENELEC JTC	In support of EU regulation and ethics
Respect of the law	Permission to use data - Consent/revoke the use of personal data	To be addressed in CEN-CENELEC JTC	In support of EU regulation and ethics
Other	Volume and Velocity - impact on standards	To be guided from a standardisation perspective by CEN-CENLEC JTC	R&D needed