CEN-CENELEC response to the EC White Paper on AI

Executive Summary

This paper is the official response from CEN-CENELEC on the EC White Paper on AI. It builds on a strong consensus of over 70 experts joined together in the CEN-CENELEC Focus Group on AI.

The Focus Group has identified 12 important themes, which in total contains 34 recommendations for the European Commission. These 12 themes are:

- Fundamental considerations
- Definitions and terminology
- Defining the scope of the regulatory framework
- Addressing the risks and values
- Voluntary labelling of non high-risk AI systems
- Role of Digital Sovereignty
- Explainability (AI)
- Virtual/digital testing
- Safety and conformity assessment
- Connecting R&D with standardization
- Shaping Europe's Digital Future
- European data strategy

Besides these recommendations the Focus Group is preparing a European Road Map for AI standardization, which is expected to be finalised in September 2020.

1 Introduction

The EU has much to gain by shaping a framework for the use of Artificial Intelligence that is associated with trust and respect for human rights of individuals and the rule of law.

The Commission has invited stakeholders to comment on the White Paper on Artificial Intelligence COM (2020) 65. CEN and CENELEC would like to take this opportunity to comment on the White Paper and on Commission documents A European strategy on data COM (2020) 66 and Shaping Europe Digital Future COM (2020) 67, which were released the same day.

CEN and CENELEC recognize that the Commission continues its strong support of a human-centric approach where AI clearly must serve people and the society. CEN and CENELEC appraise that the White Paper focuses on the characteristics, capabilities and applications of current narrow AI technologies and does not speculate and raise fears about general or super AI.

CEN and CENELEC are currently analysing whether relevant standards are already being produced at international level and if European standards covering specific European needs should also be produced. CEN and CENELEC have included over 70 experts representing companies, consumers, researchers, conformity assessment bodies, member states and other societal stakeholders. This expert group is preparing a European Road Map for AI standardization, which is expected to be finalised in September 2020.

2 What are European standards and what do they do?

European standards are market-driven and facilitate the smooth implementation of European policies and legislation. The European Commission may ask European Standardization Organizations (ESOs) to produce harmonised standards. These standards, which can be harmonised (providing presumption of conformity with legal requirements), make up around a quarter of all European standards. CEN and CENELEC have a close and well-established dialogue with the European Commission on these strategic issues.

Having one single standard, created with the consensus of all interested parties and adopted across the European Market, instead of 34 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.

3 International AI standards development

IEC and ISO have already set up a joint committee (ISO/IEC JTC 1/SC 42) which carries out standardization activities for artificial intelligence. The current structure for the standardization work of this committee include working groups on: Foundational standards, trustworthiness, use cases and applications, data, computational approaches and computational characteristics of Artificial Intelligence and AI Applications, and governance implications of AI. Currently 45 countries participate actively in this work and there is extensive involvement from EU and European countries, with 20 countries participating and holding a total of 18 leadership positions such as convenorships and editors. The SC 42 technical committee is coordinated by an American management team. Besides SC 42, also other ISO and IEC committees have ongoing standardization activities for artificial

intelligence, such as ISO/IEC JTC1 SC 27 (Information security, cybersecurity and privacy protection), ISO/IEC JTC1 SC 7 (Software and systems engineering), ISO TC 199 (Safety of machinery), ISO TC 22 (Road vehicles), ISO TC 215 (Health informatics), IEC TC 65 (Industrial-process measurement, control and automation) and IEC SEG 10 (Ethics in autonomous and artificial intelligence applications).

4 Fundamental considerations

Artificial Intelligence (AI) covers many techniques and methods. More and more AI applications (or AI components of applications) deployed today or being planned for real-world use in the immediate future, are based on Machine Learning. Moreover, the field of AI is rapidly evolving.

The performances of Machine Learning systems based on "training data", i.e. examples on which to learn, are inherited from the strengths and weaknesses of the data used, as a representation of the world into which the AI is to be deployed. Unless the data encodes it, which is very hard to determine, there is no "common sense" as such in an AI system. Hence it is possible for humans and machines to make very different decisions, even given the same data.

Recommendations

4.1 Legislation/regulation should bear in mind that AI is rapidly evolving and should incorporate reviews during the lifecycle of AI.

4.2 When dealing with Machine Learning, regulators must pay attention to the relationship between the data and the real world to which the results of Machine Learning will be applied.

5 Definitions and terminology

There are many proposed definitions for AI, and more than 60 years after its inception, experts are still debating what it is and what it will be. ISO/IEC JTC1/SC42 is currently developing ISO/IEC 22989 "Artificial Intelligence Concepts and Terminology" which could be referenced when it is finally adopted for certain terms and definitions. Concepts and terminology specific to machine learning are described in ISO/IEC 23053 "Framework for Artificial Intelligence (AI) Using Machine Learning (ML)".

The White Paper uses various definitions and terms like AI, machine learning and deep learning. The simple approach used in clause 1 of the white paper which defines AI as a collection of technologies that combine data, algorithms and computing power is very generic. This approach, if literally taken, is very broad and may cover almost any piece of conventional software.

Other organizations have their own set of different terms and definitions, sometimes more restrictive, depending on their viewpoints, expectations and purpose.

When it comes to AI regulation, stakeholders need to understand whether it applies to an organization, product or service. Any legislation should be formulated around the specific properties of the system (e.g. machine learning; statistical inference) and application context.

Recommendations

5.1 Consider the upcoming ISO/IEC 22989 "Artificial Intelligence - Concepts and Terminology" and ISO/IEC 23053 "Framework for Artificial Intelligence (AI) Using Machine Learning (ML)" as reference for terms and definitions.

5.2 The EC needs to provide clear scopes when formulating European policy or legislation. This includes using relevant technical terms (e.g. autonomy vs. automation, AI vs. Automated Decision Making, etc.) more precisely. Reference to existing or future standards and other sources is recommended. Using a glossary in preparation of policy making or legislation etc. should also be considered.

6 Defining the scope of the regulatory framework

The White Paper states in clause 5.C "A key issue for the future specific regulatory framework on AI intelligence is to determine the scope of its application. The working assumption is that the regulatory framework would apply to products and services relying on AI. AI should therefore be clearly defined for the purposes of this White Paper, as well as any possible future policy-making initiative". A clear definition to which products and services the regulatory framework is relevant is indeed necessary in order to have legal certainty for all stakeholders.

Looking on the specific issues that shall be covered by the regulatory framework (i.e. risks of fundamental rights, including personal data and privacy protection and non-discrimination and risks for safety and the effective functioning of the liability regime), the related issues and types of legal requirements scoping the framework specifically on AI seems not appropriate. While AI is a trigger for the establishment of a future regulatory framework, the actual issues are not limited to AI technologies. Coverage of stand-alone software by existing legislation, change of functionality of products due to software update, allocation of responsibilities between different operators in the supply chain and changes to the concept of safety are not limited to AI, but apply in general to ICT based solutions. Lack of transparency or explainability can also be an issue for complex conventional systems.

Requirements on keeping of records and data, information provisioning, robustness and accuracy, human oversight and requirements for biometric identification for high-risk applications should be independent of the underlying technologies. Coverage of all relevant scenarios by the training and test data (the latter is missing in the White Paper) and the quality of the data applies to certain AI technologies like ML, but not to all. In rule-based systems the coverage of all relevant scenarios must be ensured by the sum of the defined rules and is equally important. Products and services are made of various technology components and only the combination of all these technologies results in the behaviour of the system and the interplay of the technologies has cover all relevant scenarios, not only the AI training data. Products and services for high-risk applications might even be realized without AI. Still the same requirements should apply in order to ensure the same level of trust. Instead of scoping the regulatory framework on AI it should be scoped by the technology and application areas that fall under the framework.

A risk-based approach is proposed for identification of relevant applications for the regulatory framework. Two cumulative criteria shall be used to identify high-risk applications. In a first step sectors where significant risk can be expected shall be identified and listed in the regulatory framework as high-risk sectors. In a second step applications in these sectors shall be evaluated concerning their risks. The use of a risk-based approach is fully supported. ISO/IEC JTC1/SC42 is currently working on a standard on AI Risk Management. However, the two cumulative criteria approach must be questioned. While in high-risk sectors more risks can be expected to occur, this does not mean that significant risks will not also occur in so-called "low-risk" sectors, especially with the new scope of risks as defined by the White Paper.

To sum-up, some "high-risk" sectors have already specific regulations due to their specific application areas, but a general approach that only certain sectors have high risks is not valid taking for example the various areas of risk like safety, privacy, security and equal treatment into account. Furthermore, it is required to clearly define the border between high-risk and low/medium risk applications. The low/no risk example of a flaw in the appointment scheduling system in a hospital provided in the White Paper shows the complexity of the issue. Such a flaw in the appointment system could prefer or prevent treatment of certain population groups (e.g. based on income, skin colour, etc.) violating the fundamental right on non-discrimination.

In future, new risks and challenges to the attribution of liability may emerge with the increasing complexity of AI Applications. For example, the accuracy of AI in controlling real-time systems, such as self-driving vehicles, may need to be investigated soon in order to bring legal certainty to AI development and use in this specific area. A particular legal challenge of AI software development and use may lay in the dependency on third-party software modules with increasing amount of heterogeneous code, where traceability tools might be needed. Standards can be used as tools to support the application of liability rules in terms of best practices and in the framework of determining the duty of care (in the context of tort law).

Recommendations

6.1 Scope the regulatory framework around the relevant application behaviour and not around AI, this includes the revision of e.g. Machinery Directive/General Product Safety Directive /RED/MDR and the supporting European standards.
6.2 Clearly define the criteria for risk-assessment of AI applications in the future regulatory framework.
6.3 Consider existing and evolving standards on risk management (e.g. ISO 31000, ISO 12100, ISO/IEC 23894 in prep.) for the risk-based approach.

7 Addressing the risks and values

As with any technology, an artificial intelligence system can be used in various ways, some of them leading to individual and societal risks. Values or rights approach should therefore be an addition to the application risk-based approach. Many of these rights are, in specified circumstances, legally enforceable in the EU so that compliance with their terms is legally obligatory. But still ethical and value-based reflections can help us understand how the development, deployment and use of AI applications may implicate fundamental rights and their underlying values. This will provide a more fine-grained guidance when seeking to identify what should be done rather than what (currently) can be done with technology[1].

The potentially conflicting and complementary aspects of positive and negative risks must be specified. Outlines of (a) how to safeguard values by a risk-based approach and (b) how to avoid risks of not developing or using AI are needed as a basis for the operationality of European AI.

The FG furthermore notices that the current focus of the White Paper is on products and services. The FG support the view that the conformity assessment for technologies utilizing AI functions in high-risk

[1] "ETHICS GUIDELINES FOR TRUSTWORTHY AI" by High-Level Expert Group on Artificial Intelligence, 8 April 2019

applications is needed, but it would like to point out that this view needs to be complemented by the way AI products and services are designed, developed, and used by European industries and public sector organizations.

It can be expected that in a few years every Internet based service will utilize AI in some way or another. But since Internet based services (a) will be frequently used in high-risks contexts, e.g., as part of critical infrastructures, and (b) undergo frequent and dynamic changes, updates, and revisions, therefore, appropriate service management measures are needed to ensure that requirements such as the ones mentioned in Section D of the White Paper are met on a continual bases.

Recommendations

7.1 Express clearly how values can be safeguarded by a risk-based approach. Reference existing standards and other sources is recommended when working with a risk-based approach.
7.2 Discuss risks of not developing/using AI (e.g. in health care for an aging population). Work towards a broad societal consensus about which risks are acceptable (e.g. the legal decision to make seat belts compulsory even though in rare circumstances they can be negative).
7.3 Work towards identifying and understanding conflicts between values and fundamental rights in specific cases (e.g. privacy vs. transparency), and ways of handling such conflicts, including research on how conflicts can be avoided in the first place (e.g. in anonymisation of medical data, using simulated data, etc.).
7.4 Clarify the possible working space for research with respect to safeguarding values and taking risks.
7.5 Promote discourse on prioritising values and fundamental rights depending on circumstances and context, and whether different values should be protected from risk in different ways.
7.6 Provide specific requirements for values or rights so it can be integrated in the risk-based approach, which takes proportionality into account.
7.7 In addition to risk to individuals, risk to society and environment should be considered.
7.8 The assessment of the suitability of organizational management structures about the development and use of AI is considered by the European Commission complementary to requirements on products and services utilizing AI functions. Standards setting requirements on such management structures, including those on;

- the management of risks, and, as part of risk management, the impact assessment of AI,
- the management of data related aspects such as quality, statistical bias, etc.,
- the documentation of testing and validation procedures, and
- the establishment of structures for human oversight,

should be understood as measures to ensure and to document regulatory conformance on organizational level.

8 Voluntary labelling of non high-risk AI Applications

The suggestion of a voluntary labelling regime sounds good in principle, but a meaningful labelling system will require careful thought, as the absence of clear legal rules can create confusion if the labelling scheme is not trustworthy. Labelling of an AI component could not be meaningful as the AI may be integrated into some application. It is the labelling of the AI applications which needs to be

looked at. Furthermore, depending on the specific application various aspects (safety, fairness, privacy, security) have different relevance which must be considered by such a labelling scheme.

Even though the idea is a voluntary label, it has all the ingredients to move towards a certification scheme. Especially in the B2B market, relationships are based on mutual contractual agreements where requirements concerning the behaviour of a system are defined. Therefore, labelling is more useful for the B2C market if it considers both the benefits and costs from the perspective of consumers and companies.

Recommendations

8.1 Consider existing labelling schemes in Europe and the correspondent standards as inspiration, e.g. EU Ecolabel, upcoming certification scheme for Cybersecurity (Cybersecurity Act) and national AI labelling schemes (e.g. Denmark, Malta, etc).

8.2 To build a trustworthy and reliable label, standards are needed. Standards which currently do not yet exist. To first promote the development and acceptance of these standards, before introducing a labelling scheme. After introduction of a labelling scheme it is important to keep evaluating its effects.

9 Role of Digital Sovereignty

The objective of digital sovereignty for Europe is currently dependent on AI toolkits provided by US companies. These toolkits have become de-facto standards due to being technologically fit for purpose and free to use. The toolkits are open source but too complex for most programmers to fully understand, especially with respect to their cloud integration. There is a need to ensure that these technologies are easy to use in the European context – that they can be used straightforwardly with European infrastructure. There is also a need to enable European requirements to be addressed in future evolution of these toolkits. One of the ways to maintain such interface is to use the future EU GAIA-X infrastructure[2].

Recommendations

9.1 Support the creation of "AI toolkit distributions", tailoring the AI toolkits to the European context, e.g. by using the future EU GAIA-X infrastructure

9.2 Press for the open governance of the various AI toolkits, enabling European input into the evolution of the toolkits

10 Explainability

The problem of explainability needs to be framed for different kinds of systems such as decision making or sensing systems. Assessment of explainability and explanations needs to be grounded in context, benchmarking, and the targeted recipients or stakeholders, such as developers, users, consumers, etc.

High level requirements for explainability need to be defined in international standards in concert with defining transparency and verifiability for AI applications and their relationship to explainability in various contexts and at various comprehension levels.

[2] https://www.bmwi.de/Redaktion/EN/Publikationen/Digitale-Welt/das-projekt-gaia-x-executive-summary.pdf?__blob=publicationFile&v=6

Recommendations

10.1 Launch calls for inter-disciplinary research on explainability, in order to make the technology ready for standardization.
10.2 Develop research-based metrics for explainability (to tie in with high level conceptual requirements), which can be developed into pre-standards like workshop agreements or technical reports.
10.3 Launch calls for research on verification and verification coverage techniques.

11 Virtual/digital testing

Testing of AI applications will occur more frequently than development of new AI applications, as testing is often required for specific deployment contexts. Testing and certification bodies will rely more and more on simulation or virtual testing to perform conformity assessment (in addition to field testing) of AI applications. That approach will be key for safety demonstration, but also in many other cases such as market surveillance.

For self-driving vehicles, for example, the development of a standardized virtual testing environment may be required, which in addition may lead to interoperability needs between the digital twins of the car and the environment model. Virtual testing facilities could be given access to SMEs and to new stakeholders. A connection between virtual testing environments and physical testing infrastructures will have to be established.

Even for applications that are not safety critical, there are significant risks to society if ethical aspects of systems cannot be validated, for instance to detect biased outcomes based on personal data. It is also noted that biased outcomes are not always detectable by an organization as they may not have information on protected characteristics, and AI components may be utilizing "hidden variables" that are not understood. For low-risk systems that may be implemented there are also challenges in the verification techniques used that lead to a reduction in perceived quality.

Recommendations

11.1 Fund project calls for standardized virtual testing environments; “testing facilities” programs launched on the next EU FP could also include this topic.
11.2 Request European standardization organizations to come up with an overall standard for virtual testing facilities, including standards for interoperability between digital twins and standardized virtual testing environment and standards for physical simulations/modelling (sensors, actuators, etc.).
11.3 Consider a focus on the quality of AI applications as a key differentiator for Europe industry, and fund project calls for research and industry collaboration in this area.
11.4 Fund a project call for a recommendation and / or technical solution that allows organizations to assess biased outcomes without increasing the organization’s access to personal data.

12 Safety and conformity assessment

Safety and conformity assessment are an essential activity for AI applications risk management to ensure trustworthiness. The next ISO IEC 61508 standard update may enable using AI in level 2 safety critical systems. So, there is a reinforced need to clarify the assessment issues for safety within such AI based systems. Along with the “Risk Management for AI” standard within ISO SC42 (future IS n°23894), there are other standardization bodies which are exploring how to address safety and conformity issues for AI applications. There is a strong need for alignment regarding terminology,

concepts, criteria, metrics and methods among horizontals (general purpose) and verticals (domain specific) standardization organizations.

Recommendations

12.1 Fund specific European events/sessions for sharing the efforts made by the different standardization bodies (horizontals and verticals) regarding assessment criteria, metrics and methods, and promote a set of general-purpose European standards for safety of AI applications.

12.2 Request European Standardization Organizations to provide sector-specific standardization e.g. automobile, health and aeronautics, conformity assessment schemes. And by doing so, include all stakeholders for applications in relevant sectors, for use of AI technology in a safety critical system.

13 Connecting R&D with standardization

Conformity assessment methods, metrics and practices often rely on standards, especially when they deal with safety and risk issues. However, it is difficult to develop standards in new technical areas such as AI, without preliminary achievements in research. Therefore, connecting the "Researchers" and the "Practitioners" to standardization through some R&D activities might speed up the process and focus on the much-needed EU framework of standards for AI components and applications, therefore contributing to the building of an AI trust Ecosystem. For AI components and applications, this may require research activities into technical topics such as explainability, transparency, robustness, verifiability, etc. Furthermore, rooting good practice of AI use in EU research projects from the start of research programmes such as Horizon Europe, will allow for those good practices to be promoted in every sector of the Research Programme, health, transport, and any other, leading to solid contribution to standards rooted in future-looking and sector-motivated use cases. This would help to secure specifications reflecting the European needs and know-how, in the European and Global contexts.

Recommendations

13.1 Expand and reinforce the Stand ICT programme in order to support the participation of academic and public researchers from any research domain into standardization of AI, where they either contribute ground-breaking use cases (advanced sectorial use cases of AI) or disruptive AI developments.

13.2 Fund academics and experts for standards contributions.

13.3 Actively use the CSA (coordination and support action) type of funding available in the Framework Programme (currently Horizon 2020, soon Horizon Europe), as well as IA (innovation actions) to ensure the dynamic development of standards for new AI core functionalities (AI research) or the use of proven AI in a variety of systems (AI engineering and sectorial use, including in non-AI research), aiming at an adequate standardization of AI in the EU while ensuring global standardization wherever possible for critical mass and scale. The funding should target the overall AI-based systems that address; explainability, transparency, verifiability, and similar features from the area of trustworthiness, or other features with impact on improved acceptance and societal benefit.

14 Shaping Europe's digital future

A European trusted digital environment is strategic and by essence sovereign, meaning that not everything is market driven. It is up to the European Commission, as a major stakeholder in the standardization domain, to set its directions.

The question of addressing sovereignty within standardization is therefore a new topic for SDOs, with many opened questions: What does it imply for an SDO to be European? What is the meaning of sovereignty within standardization? What would be the associated roadmap and directions?...

Recommendations

14.1 Identify at the European Commission level what are the implication of European sovereignty needs for European Standardization Organizations.

14.2 Set standardization directions at the European Commission level in order to address sovereignty properly and to lead in the adoption and standardization process of the future digital technologies.

15 European data strategy

The communication from the Commission on a European strategy for data (COM (2020) 66) sets a vision for the EU's share of the digital economy. In doing so, it refers numerous times to the building of Data Spaces. In the meantime, the AI HLEG has identified a concept of "Trusted data spaces" in its Policy and Investment recommendations. Both concepts look to go beyond pure Data Spaces.

In its preliminary work, the AI FG has discussed a potential conceptual framework named "Digital Sphere". Digital Spheres can be used to define privacy (private sphere), ownership and sovereignty but also justify interoperability needs as Digital Spheres would have to interact with each other. This "Digital sphere" concept is still being discussed and there is no consensus on its scope.

Those three concepts 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.

Recommendation

15.1 Request European Standardization Organizations to come up with terminology, concept definition and ontology for data space/trusted data space/digital sphere covering European needs