The Human Side of Service Engineering, Vol. 182, 2025, 158–177
https://doi.org/10.54941/ahfe1006416

Protection of AI/ML End Users From 'Bad Actors': Challenges and Policy Responses

Christian M. Stiefmueller[1], Christine Leitner[1], and Stephen K. Kwan[2]

[1]Centre for Economics and Public Administration Ltd, London, UK/Vienna, Austria
[2]San José State University, San José, CA, 95192, USA

ABSTRACT

There is widespread agreement globally on the potentially huge benefits and risks associated with the adoption of AI/ML. In a previous article we have explored regulatory frameworks for AI/ML in major global jurisdictions through different lenses. We were interested in comparing how major global jurisdictions, such as the US, China and the EU, approached the challenge of reconciling the potentially disruptive impact of adopting this new technology with the responsibility of policymakers to take into account the interests, needs, and rights of their citizens. At the present time, there is no obvious global consensus on where the balance between the two should be struck. Most recently, an international declaration on the inclusive and sustainable use of AI was signed in Paris by sixty countries, including all EU member states and China, but not the US and the UK. In the absence of a political consensus among major jurisdictions and faced with often divergent regulatory approaches at the national and regional levels, efforts by national and international standard-setting organisations, such as ANSIâŁ™s federation of Standard Development Organisations (SDOs), CEN/CENELEC/ETSI, and ISO/IEC/ITU, are bearing the burden of the increasingly important tasks of establishing common AI standards at the technical level. Some of these organisations have created dedicated work streams that expressly seek to incorporate the perspective - and voice - of the citizen into the technology design, development and adoption process. This article will examine the most important of these initiatives and try to assess their possible contribution as well as the implicit limitations of their mandate and capacities.

Keywords: Artificial intelligence, Machine learning, Digital governance, Regulatory model, Standardisation, AI ecosystem, Beneficial AI, Stakeholder lens, AI Act, Digital trade

1. INTRODUCTION

This paper seeks to examine the social-political and regulatory dynamics that arise from the confluence of two major trends, both already discussed extensively in academic literature and policy-making circles. On the one hand, there is the rapid rise of 'bad actors' making deliberate use of AI/ML-enabled tools and applications to inflict harm upon individuals, groups, or even society at large. The increase in potentially damaging, or even outright criminal activity involving AI/ML has instilled new urgency to the search for

frameworks to define and promote 'trustworthy' and 'ethically responsible' AI/ML.

On the other hand, prospects of an effective, collective global policy response have receded as AI/ML is perceived by governments as a potential source of competitive advantage in an increasingly contested geoeconomic and geopolitical environment. Global efforts to address the risks emanating from the adoption of AI/ML, and especially its use by 'bad actors', continue, of course. Initiatives by international organisations, such as the UN (e.g. the Global Digital Compact,[1] the UN AI Advisory Body's report on 'Governing AI for Humanity'[2], and the UNESCO Recommendation on Ethics in AI[3]), the OECD (AI Principles[4]), and others, as well as bilateral and multilateral initiatives are progressing and most jurisdictions remain committed to maintaining a dialogue on a potential global governance framework for AI/ML. An international declaration on the inclusive and sustainable use of AI was signed in Paris in February 2025 by sixty countries, including all EU member states and China. The absence of the US and the UK from the list of signatories illustrates the lack of a regulatory consensus at present, even among traditionally aligned Western democracies.

This paper will present a brief overview of the potential risks from the use of AI/ML by 'bad actors' (sec. 2) and of the current status of the political debate around regulating AI/ML at the global level (sec. 3). The apparent lack of a political consensus on regulation has shifted much of the attention towards standardisation, which is already closely integrated into existing and emerging regulatory frameworks and may take on an even more critical role (sec. 4). Section 5 looks at current initiatives at the global level, and in two of the major jurisdictions, the European Union (EU) and the United States of America (US). Section 6 provides a brief overview of the limiting factors of a standardisation-led approach, followed by preliminary conclusions (sec. 7).

2. RISK FROM BAD ACTORS

Risks arising from the adoption and/or deployment of AI/ML technology are becoming more widespread and, at the same time, more diverse. While online misinformation has been a concern already for some time, the advent of generative AI, which provides widely-accessible methods for synthesising realistic audio, images, video and human-like text, has further amplified these concerns (Dufour et al., 2024). Generative AI has made AI/ML-enabled tools and applications available to a much broader universe of users. Whereas previously developers and deployers of AI/ML technology were seen as the most likely originators of harmful activity, generative AI has put the

[1]United Nations Office for Digital and Emerging Technologies, Global Digital Compact, 22 September 2024 (https://www.un.org/digital-emerging-technologies/global-digital-compact).

[2]UN High-Level Advisory Body on Artificial Intelligence, Governing AI for Humanity: Final Report, September 2024 (https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf).

[3]UNESCO, Recommendation on the Ethics of Artificial Intelligence, 23 November 2021 (https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence).

[4]OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 22 May 2019 (amended 30 May 2024) (https://oecd.ai/en/ai-principles).

technology within reach of end users, and in the process significantly widened the spectrum of potential 'bad actors'.

This expanding, increasingly diverse population of potential 'bad actors' also finds a growing array of attack vectors and methods at their disposal. Recital 76 of the EU AI Act[5] notes that *"cyberattacks against AI systems can leverage AI specific assets, such as training data sets or trained models, or exploit vulnerabilities in the AI system's digital assets or the underlying ICT infrastructure*". The National Institute of Standards and Technology (NIST) proposes a taxonomy of 'Adversarial Machine Learning' based on generic types of attack, which may target the availability, integrity, or model/data privacy of AI/ML systems (NIST, 2024). In the case of generative AI systems, abuse is added as a fourth category, although the system itself is, in this instance, not necessarily the object of the attack but becomes a vector for carrying out an attack. This taxonomy may be difficult to apply in a stringent way in practice but provides a useful frame of reference to (i) categorise methods of attacks on, or by means of AI/ML, and (ii) inform a structured discussion of potential policy responses.

There is widespread political agreement that any regulation of AI/ML should be proportionate and risk-based[6]. Accordingly, existing regulatory frameworks, such as the EU AI Act, focus on the governance of AI/ML systems that are deemed to be 'high risk'. According to Art. 6(3) of the AI Act, systems are considered to be 'high risk' if they *"pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making"*. The implicit challenge of measuring and classifying what constitutes 'significant risk' is well recognised and the subject of intense debate. Especially in the context of 'bad actors', this challenge is further amplified by the fact that (ex-ante) risk assessments, and hence classifications, tend to focus on the system's intended use rather than their actual use. To address the potential deployment or subsequent modification of AI/ML systems by 'bad actors' both ex-ante testing, including for potential vulnerabilities to non-intended, malicious use, as well as continuous ex-post monitoring will be required (NIST, 2023a).

It appears appropriate in this context to briefly revisit the concept of 'harm'. The protections of the AI Act are limited expressly to harm to the health, safety or fundamental rights of natural persons. This definition is quite narrow in two respects: (i) it does not account for economic loss, which can be as much of a threat to the life prospects of an individual as physical and psychological harm; and (ii) it does not go beyond the individual level to include collective harm, e.g. offences that affect the public order and/or the functioning of political institutions. Other definitions of 'harm', e.g. in the NIST Risk Management Framework (NIST, 2023a), are broader and recognise both 'collective harm', e.g. harm to organisations and ecosystems, and economic harm. The scope and reach of the activities of 'bad actors' are

[5] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act), OJ L 2024/1689, 12 July 2024.

[6] Calls for a risk-based approach to regulating AI/ML are made, e.g., in the Global Digital Compact (Fn. 1) and the Bletchley Declaration (Fn. 9).

also key determinants of their potential impact. At one level, 'bad actors' among developers or deployers may deliberately compromise the integrity of AI/ML tools or applications potentially affecting a wide section of the population, including professional and institutional users, and wide sections of the general public. The same may be effected by end users manipulating generative AI at scale (H.M. Government, 2024). Attacks at this order of magnitude may pose a direct threat to public order, and the functioning and stability of public institutions and processes. Other activities, by end users and on a smaller scale, such as mis- and disinformation amplified by the use of AI/ML and disseminated through social media, may, in the first instance at least, have a more limited impact, primarily on other individuals or specific sub-groups of the population, e.g. by exposing them to hateful, defamatory, and/or discriminatory online agitation, which infringes on their fundamental rights. Malicious activity is not specific to particular types or applications of AI/ML systems but may affect all systems throughout all phases of their life cycle. Left unchecked such attacks may have serious and corrosive implications for democracy, peace, and stability (UNO, 2024). In the EU, these aspects are covered by other horizontal and sector specific legislation, such as the Cyber Resilience Act (CRA) adopted in 2024[7], which applies to all products with digital elements, not only AI systems, and provides a detailed implementation framework, including cybersecurity requirements, monitoring and enforcement, analogous to the framework for high-risk AI systems set out in the AI Act.

3. POLICY RESPONSES AND CHALLENGES

Protection against the activities of 'bad actors' is needed, therefore, at both levels, individually and collectively. From a traditional Western-oriented policy perspective, the protection of the individual revolves, implicitly, around the protection of their fundamental rights. This is the approach that has been embraced by the EU with the AI Act, and which tends to inform most of the relevant international declarations and statements, e.g. under the auspices of the UN and the OECD, the G7 Hiroshima Process[8], the Bletchley Declaration[9], and the Statement on Inclusive and Sustainable Artificial Intelligence[10] signed recently at the AI Action Summit in Paris. The practicalities of implementing legislation to protect fundamental rights are altogether more complex and involve difficult trade-offs in the process of prioritising and balancing individual and collective interests, which are

[7]Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements (Cyber Resilience Act), OJ L 2024/2847, 20 November 2024.

[8]Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems (https://www.soumu.go.jp/hiroshimaaiprocess/en/index.html).

[9]AI Safety Summit, The Bletchley Declaration by Countries attending the AI Safety Summit, 01 November 2023 (https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023).

[10]AI Action Summit, Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet, 11 February 2025 (https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet).

invariably shaped by constitutional arrangements, legal traditions, and cultural values. These trade-offs appear even more challenging when it comes to assessing potential threats at the collective level and determining what is perceived as a threat to the political and social fabric of a society.

At present, there appears to be little consensus, even among Western democracies, on what form of policy response would be appropriate, let alone on how the balance of interests should be struck. Hence the current sharp divergence in policies across major jurisdictions. In Europe, the EU pursues a distinctive, 'rights-based' approach of building trust in technology through compliance with laws, which in turn build on political agreements ('soft-law') and codifications of common European rights and values. Individual legislative measures, such as the Digital Services Act (DSA)[11], and the Artificial Intelligence Act (AI Act)[12], are embedded in a programmatic framework, which invests EU institutions with a wide-ranging mandate to regulate, and sets out general principles based on fundamental and civic rights (Leitner and Stiefmueller, 2025). The US, by contrast, traditionally prefers to let industry formulate their own regulatory and standards regime from the 'bottom-up', with the government intervening only when there is a need for additional or extraordinary measures (Kwan et al., 2024). Regarding AI/ML, the US government appears to be particularly reluctant to take regulatory action that could place US companies at a competitive disadvantage vis à vis overseas competitors. An Executive Order on 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence'[13], issued by the previous presidential administration in November 2023, was repealed by the incoming administration[14] and replaced, on 23 January 2025, with an Executive Order on 'Removing Barriers to American Leadership in AI'[15] (see 5.3). In practice, however, both approaches have in common that they rely, to a large extent, on technical standards.

4. LEGISLATION AND THE ROLE OF STANDARDS

4.1 European Union

The AI Act follows the EU's New Legislative Framework (NLF)[16], which was first introduced in 2008 as a horizontal framework to regulate product safety, in that it formulates essential, high-level legal requirements, which are then operationalised by way of technical standards (known as 'harmonised

[11]Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act), OJ L 277, 27 October 2022, pgs. 1–102.

[12]Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act), OJ L 2024/1689, 12 July 2024.

[13]Executive Order No. 14110 of 30 October 2023 on 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence', 88 Fed. Reg. 751191 (01 November 2023)

[14]Executive Order No. 14148 of 20 January 2025, 90 Fed. Reg. 8237 (28 January 2025).

[15]Executive Order No. 14179 of 23 January 2025 on 'Removing Barriers to American Leadership in Artificial Intelligence', 90 Fed. Reg. 8741 (31 January 2025).

[16]Regulation (EC) No 765/2008 of the European Parliament and of the Council of 09 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products, OJ L 218 of 13 August 2008, pgs. 30–47.

European standards', hENs). These standards are drawn up by the European Standardisation Organisations (ESOs), CEN-CENELEC[17] and ETSI[18], at the request of the European Commission. Under the AI Act, AI/ML systems that are considered high-risk require a conformity assessment, which is based on these standards and carried out by independent private-sector experts. Any system that has passed the assessment may carry the European Conformity (CE) mark, and is presumed to be compliant with the underlying legislation. This implies that, although the general legal and political principles are enshrined in legislation in what may be termed a 'top-down' approach, technical standards ultimately play a decisive role in how they are implemented in practice.[19]

Under EU law[20], a standard is '*a technical specification adopted by a recognised standardisation body for repeated or continuous application, with which compliance is not compulsory*'. A harmonised European standard (hEN) is a standard developed by a European standardisation organisation (CEN, CENELEC, or ETSI) upon a standardisation request from the European Commission for the application of EU harmonisation legislation. According to the case law of the Court of Justice of the European Union (CJEU)[21] hENs, unlike 'ordinary' standards adopted by the European standardisation organisations, should be seen as an implementing acts that form part of EU law. They have de facto mandatory effects in that (i) they establish a legal presumption of conformity, and (ii) it is virtually impossible to confirm compliance with the law without having knowledge of the relevant standard (Gornet and Maxwell, 2024). From a formal perspective, they are adopted by the European Commission, and published in section L of the Official Journal of the European Union, which contains legislation.

4.2 United States of America

In the US, and other jurisdictions which follow a 'bottom-up' approach, legislators tend to leave it to the interaction of industry participants and market forces to produce voluntary or even informal, '*de facto*' industry standards. In the US, this process involves several hundred private-sector standardisation bodies, coordinated and overseen by the American National Standards Institute (ANSI), in close cooperation with the National Institute of Standards and Technology (NIST). US federal agencies are encouraged to use available voluntary industry standards to meet standardisation needs in procurement and regulation so that industry standards are incorporated over time into new or existing regulation, and hence into the Code of Federal Regulations. At this stage, these 'standards incorporated by reference' attain

[17]Comité Européen de Normalisation (CEN) and Comité Européen de Normalisation Électrotechnique (CENELEC).

[18]European Telecommunications Standards Institute (ETSI).

[19]As mentioned previously, this applies to the EU's approach towards cybersecurity more generally, as exemplified by the CRA (see Fn. 7).

[20]Regulation (EU) 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, OJ L 316, 14 November 2012, pgs. 12–33.

[21]CJEU, Judgment of 05 March 2024, *public-resource.org*, C-588/21 P, ECLI:EU:C:2024:201; also CJEU, Judgment of 27 October 2016, *James Elliott Construction*, C-613/14, ECLI:EU:C:2016:821.

legal status and, similar to the EU, an assessment of conformity according to the standard implies a presumption of compliance with the law.

In the US, 'voluntary consensus standards' are technical standards developed by private standard-setting organisations. Similar to their EU counterparts, they are not *a priori* compulsory but may become legally binding by virtue of a process known as 'incorporation by reference' (IBR), which has its origins in the National Technology Transfer and Advancement Act (NTTA) of 1995[22]. Once approved, the referenced material is treated as if it were published in the Federal Register and the US Code of Federal Regulations (CFR) and thus has the force and effect of law[23].

5. RELEVANT INITIATIVES AND WORKSTREAMS

Given the central role played by standards, and in the absence of a political consensus among governments on global regulations for AI/ML, at least in the near term, there is an expectation in some parts that international standard-setters could, to some extent, step into the breach. In particular, efforts to draw up a widely accepted set of standards for 'trustworthy AI systems' and their responsible use, could mitigate risks emanating from the misuse of the technology by 'bad actors'.

5.1 International

At the international level, standardisation efforts are spearheaded by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), which facilitate and coordinate the cooperation of national standard-setting bodies. Together with the International Telecommunications Union (ITU), a specialised agency of the UN, they form the World Standards Cooperation (WSC), an alliance to promote the development and acceptance of consensus-based international technical standards. In its 2023 resolution on digital technologies and human rights[24], the UN Human Rights Council recognised that standards for AI/ML will shape the actual design and usage of the technology, and thus have implications for privacy, freedom of expression, and access to information. It therefore called for integrating these considerations into technical standard-setting in order to effectively align AI/ML standards with international human rights law. The World Standards Cooperation, which comprises the ITU, ISO and IEC, has taken up this mandate to embed that human rights lens into technical standards.

5.1.1 ISO/IEC Joint Technical Committee 1

ISO/IEC JTC 1 is a joint technical committee (JTC) of the ISO and the IEC tasked with developing, maintaining and promoting standards in the fields of information and communications technology (ICT). In 2017, JTC 1 created

[22] National Technology Transfer and Advancement Act, Pub. L. 104–113, 07 March 1996, 110 Stat. 775.

[23] https://ibr.ansi.org/

[24] UNGA, Res. 53/29 (14 July 2023) 'New and emerging digital technologies and human rights', UN Doc. A/HRC/RES/53/29.

a dedicated sub-committee, SC 42, to focus on AI/ML. The secretariat of ISO/IEC JTC 1/SC 42 is provided by the US member organisation of ISO, ANSI. The sub-committee currently has five working groups and are actively collaborating with other committees and organization[25]. Within SC 42, a dedicated working group, WG 3 'Trustworthiness', concerns itself with the analysis of factors that could impact the trustworthiness of AI systems. Examples of these factors are shown in Figure 1.

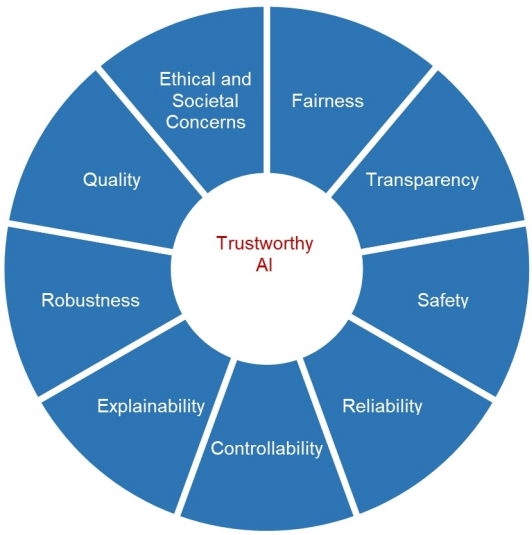

Figure 1: Factors of trustworthy AI.

These factors are the non-functional characteristics/requirements of an AI system, i.e. degree to which a user or other stakeholders has confidence that the AI system will behave as intended. The context of these factors of trustworthy AI from the perspectives of the stakeholders of the AI system is shown in Figure 2 and explained below.

1. Market Demand Analysis – AI producer seeks to understand AI customers and users' desire for trustworthiness (e.g., including factors shown in Figure 1) and risk appetite, etc. in AI systems.
2. The Market Demand Analysis forms the basis for the AI producer to *Make* the AI system with the appropriate AI technology together with considerations to (possibly) Responsible AI Principles, Corporate Ethical Guidelines, Risk Assessment and Mitigation, etc. The AI producer's considerations of Responsible AI Principles include incorporating the trustworthiness factors desired by the customers and users into the AI system.

[25]ISO/IEC JTC 1/SC 42 has liaisons with 22 Category A and B organisations, and 3 Category C organisations. It also has 61 liaisons to other committees and has 51 liaisons from other committees (https://www.iso.org/committee/6794475.html).

3. The produced AI system could then be marketed to potential AI customers and users with an appropriate value proposition and (possibly) license (e.g., EULA, RAIL[26]).
4. The potential AI customers and users could then use the AI producer's conveyance together with their own criteria to perform a cost-benefit analysis as to the desirability and suitability of adoption the AI system.
5. The cost-benefits analysis will feed into the decision to adopt.
6. If the potential AI customers and users decide to adopt, they could then *Use* the AI system for their intended purposes.
7. The *Use* of the AI system will result in value realization for the AI customers and users as well as the AI producer.
8. The *Use* of the AI system will also result in *Impact* on other stakeholders (such as the adopting organization, the community and beyond).
9. The assessment of the *Impact* could then be feedback to the AI producer and the adopting AI customers and users.

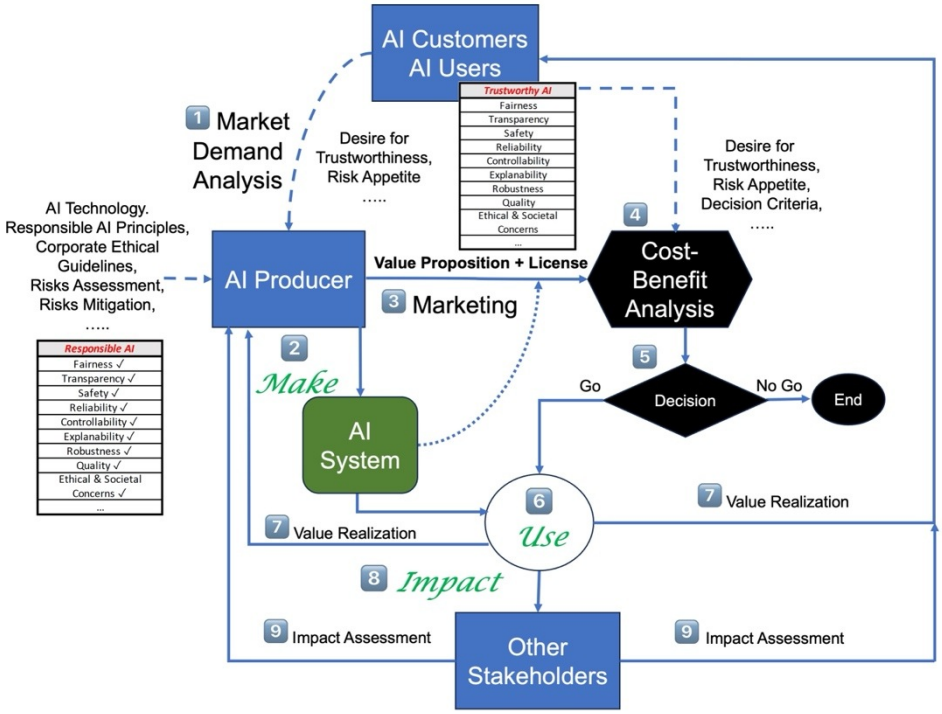

Figure 2: Trustworthy AI in the context of AI system stakeholders.

The work programme of SC 42/WG 3 is shown in Annex 1, including those documents that have been published (P) and in progress (O). Annex 1 also shows the WG 3 documents that are related to the factors of trustworthy AI shown in Figure 1 and 2. In May 2020, SC 42/WG 3 published a technical report of an overview of trustworthiness in artificial intelligence (ISO/IEC, 2020), which surveys existing approaches that could support or improve

[26]EULA – End User License Agreement; RAIL – Responsible AI License (see 5.5).

trustworthiness in technical systems and discusses their potential application to AI/ML. The report also investigates possible approaches to mitigating AI system vulnerabilities and ways to improving their trustworthiness. It identifies specific standardisation gaps in AI and provides a foundation of guidance on how to address these through future standards.

In additional to the work programme of SC 42/WG 3, other working groups of SC 42 are also tackling related aspects of AI systems. For example, SC 42/WG 4 is currently working on ISO/IEC WD TR 21221, Information technology – Artificial intelligence – Beneficial AI systems (ISO/IEC, n.d.).

5.1.2 ISO Committee on Consumer Policy (COPOLCO)

The work of COPOLCO is centred on the consumers' perspective of technology and their impacts. AI is designated as a high priority item in their work programme. Working Group 22 (Consumer standards action group) and Working Group 24 (Consumer safety group) are actively working on seeking more consumer input in the development of AI standards in cooperation with ISO/IEC JTC 1 SC 42.

5.2 European Union

In the EU, work on standards related to AI/ML has been formally assigned in the AI Act (Art. 40) to CEN-CENELEC and ETSI, subject to the guidance and oversight of the European Commission. At CEN and CENELEC a Joint Technical Committee on 'Artificial Intelligence' (CEN-CLC/JTC 21) was established in 2020 to produce standardisation deliverables in the field of AI, and to consider for adoption at European level of those relevant international standards developed in ISO and IEC, in particular in ISO/IEC JTC 1/SC 42. A formal request by the European Commission in accordance with Art. 2(1.c) of Regulation 1025/2023[27] was issued to CEN-CENELEC and ETSI in May 2023[28], well in advance of the eventual finalisation and adoption of the AI Act by the EU co-legislators. At present, CEN-CENELEC JTC 21 is working on more than 30 standardisation activities in fulfilment of the standardisation request. CEN-CLC JTC 21 comprises a dedicated working group, WG 4 'Foundational and Societal Aspects of AI', which concentrates on ethical and societal concerns related to the development of trustworthy AI systems, fundamental values and human rights.

As part of its work programme (see Annex 2), CEN-CLC JTC 21/WG 4 is in the process of drafting its own 'AI trustworthiness framework' (Work Item JT021008), one of a number of standards that were not in the Commission's original standardisation request (Gornet and Maxwell, 2024)[29]. The 'AI trustworthiness framework' will contain terminology, concepts, and high-level horizontal requirements for trustworthy AI systems, along with

[27]Regulation (EU) No. 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, OJ L 316, 14 November 2012, pgs. 12–33.

[28]European Commission Implementing Decision C (2023)3215 of 22 May 2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on Artificial Intelligence, COM/2021/206, 22 May 2023.

[29]Both the AI Act and this work draw on the 'Ethics Guidelines for Trustworthy Artificial Intelligence' issued by the EU's High-Level Expert Group on AI.

guidance and a method to contextualise those to specific stakeholders, domains or applications. The high-level horizontal requirements set out in this standard are meant to address foundational aspects and characteristics of trustworthiness of AI systems. This document is primarily intended for developers and deployers of AI systems but could, if widely implemented, provide a degree of reassurance for end users that the design features of a system fulfil certain formal criteria of 'AI trustworthiness'. While this standard is currently under development and expected to be submitted for approval by year-end 2025, its development runs in parallel, or even ahead of the corresponding ISO/IEC effort. Considering that it was not part of the original standardisation mandate issued by the Commission it remains to be seen whether it will be adopted eventually as a harmonised standard (eHEN; see 4.1).

The AI Act (rec. 121) enjoins the standard-setting bodies to encourage "*a balanced representation of interests involving all relevant stakeholders in the development of standards, in particular SMEs, consumer organisations and environmental and social stakeholders*". To this effect, EU legislation (Art. 5 of Regulation 1025/2012) provides for the participation of a broad range of stakeholders in the work of the CEN-CENELEC Technical Committees. Industry associations, civil-society groups, (non-EU) members of ISO/IEC, and other representative stakeholder groups are eligible to join Technical Committees and attend their meetings in the capacity of non-voting observers. In the case of CEN-CLC JTC 21, civil-society interests are represented by consumer organisations, trade unions and environmental organisations[30]. Moreover, CEN-CLC JTC 21 has created a Task Group dedicated to Inclusiveness, in charge of raising awareness about its activities and to bring stakeholders to the discussion. As with the Advisory Forum (Art. 68 AI Act), stakeholder representatives in CEN-CLC JTC 21 are limited to a consultative and monitoring role since they do not have voting rights. Their involvement nevertheless provides for a degree of transparency and interaction.

Traditionally, standardisation by the European standard-setters has built largely on existing international work, especially global ISO/IEC standards, to ensure a smooth alignment between the European and international standardisation frameworks. CEN and CENELEC have signed dedicated technical cooperation agreements[31] with their global counterparts, ISO and IEC, to ensure market harmonisation. At present, approx. 33% of CEN publications come from ISO and approx. 73% of CENELEC publications from IEC (Gornet and Maxwell, 2024). With the adoption of the AI Act, and the European Commission's request for standards, which are critical to its implementation, it is widely expected that the AI Act could spawn a generation of new, 'home-grown' European standards given that international standards (i) are still evolving and not subject to the time

[30]Notably the European Association for the Coordination of Consumer Representation in Standardisation (ANEC), the European Trade Union Federation (ETUC), and the Environmental Coalition on Standards (ECOS).

[31]The Vienna Agreement was signed between CEN and ISO in 1991, the Frankfurt Agreement between CENELEC and IEC in 2016 (replacing the Dresden Agreement of 1996).

constraints imposed by the implementation timeline of the AI Act; and (ii) may not, in terms of scope and intent, be in fully alignment with the AI Act and its objectives, particularly regarding fundamental rights protection and societal impacts (Kilian et al., 2025).

5.3 United States of America

As mentioned previously (sec. 3), the current US government's position with respect to regulating AI/ML marks a significant departure from the previous administration's. In a previous article (Leitner and Stiefmueller, 2025), we have discussed President Biden's "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"[32]. This Executive Order has since been rescinded. President Trump signed Executive Order 14179, titled "Removing Barriers to American Leadership in Artificial Intelligence", on January 23, 2025[33]. As of the time of writing of this article, there is still uncertainty about the implementation of this Executive Order and overall picture of AI regulations in the US.

Current work on the implementation of the new Executive Order is partially based on NIST's AI Risk Management Framework (AI RMF) (NIST, 2023a) which was published prior to the issuance of the two Executive Orders. As a follow-up to the AI RMF, NIST also published and AI RMF Playbook (NIST, 2023b) which details how organizations can implement the AI RMF functions of Map, Manage, Measure and Govern. For example, the Playbook suggests where adversarial and stress testing of the AI System could help identify weaknesses which could form axes of attack by 'bad actors' (NIST, 2023b).

As part of its outreach program to recruit more consumers to participate in standards development activities, the Consumer Information Forum (CIF) of ANSI established in 2022 a "Consumer Participation Fund". The fund is used to reimburse consumer participants in SDO's for their travel, participation and Technical Advisory Group fees. ANSI has experienced a modicum of success with this Fund and is committed to continue fund-raise and increase awareness in the consumer and SDO community.

The Institute of Electrical and Electronics Engineers (IEEE) has many work projects formulating AI-related standards. Of particular interest is IEEE P2840 Responsible AI Licensing (RAIL) (IEEE, 2024). This standard pertains to how AI system producers could define the responsible use of their product to prevent possible misuses and abuses. This type of license could be part of the conveyance from the AI producer to potential customers for consideration in their adoption/use decisions ("Value proposition + License" in Figure 2).

6. EFFECTIVENESS AND LIMITATIONS

6.1 Effectiveness

The use of technical standards in the context of regulating the adoption of AI/ML is likely to test their capacity to complement, or even substitute

[32] see Fn. 13.

[33] see Fn. 15.

for statutory law to its limits. The traditional role of technical standards, which is to ensure product safety, quality, performance, and compatibility/interoperability equally applies to AI/ML, especially where AI/ML becomes an integral part of the functionality of another product. These 'embedded' applications of AI/ML have shaped the EU AI Act to a large extent, and led the EU legislators to base their regulatory framework largely on the NLF for product safety legislation. AI/ML is an enabling technology, however, which lends itself to being incorporated in a wide range of products and services, both in 'embedded' and 'stand-alone' applications.

Technical standards that enhance the security, resilience and robustness of an AI/ML system mainly deal with its safety, quality, performance, and compatibility/interoperability, criteria that are empirically observable and quantifiable, and thus fall well within the traditional 'comfort zone' of technical standard-setting. As such, they are likely to protect end users effectively, especially in the early stages of its lifecycle. This is because technical standards primarily govern the design and functionality of an AI/ML system and compliance is usually tested prior to its market introduction and/or deployment. They are therefore likely to be most effective in preventing 'bad actor' developers and deployers from bringing systems to market, or in flagging up unintended vulnerabilities in systems that may be exploited by 'bad actors' after deployment. Over time, the effectiveness of technical standards in protecting end users is dependent as much on regular, and effective monitoring of compliance as on timely and comprehensive updating of technical standards to keep up with technological developments and the appearance of new threats.

The benefits of a positive standardisation of 'Beneficial AI' are more difficult to assess at this stage in that they rely critically on broad acceptance, not only by the industry but also by end users. Unlike standards that regulate safety, quality, performance, and compatibility/interoperability, which can be based most of the time on empirical metrics, standards that are designed to establish a level of 'trust' operate on legal (fundamental rights) and ethical (values) criteria, which do not lend themselves to quantification but require complex, multivariate balancing and trade-offs. In most jurisdictions these complex assessments are the prerogative of the legislators and the courts, based on a popular consensus enshrined in the law that embodies constitutional arrangements, legal traditions, and cultural values (see 2). Operationalising such complex concepts in the form of technical standards is likely to be challenging, even with a supporting legal framework in place.

6.2 Limitations

6.2.1 Fundamental Rights and Values

True to its origins in the world of physical products and product safety the concept of 'harm' that technical standards are designed to protect against is fairly narrow and typically focuses on bodily harm, i.e. physical injury. Being anchored in the physical world also implies that technical standards usually rely on empirical observations and measurements, which can be operationalised, e.g. by way of target or threshold values, and performance

levels. Quantitative indicators are less well suited when it comes to assessing other harms, especially the infringement of personal (fundamental) rights, and, at the collective level, offences that affect the public order and/or the functioning of political institutions.

Standards for 'trustworthy AI' can define a valuable general frame of reference, e.g. by defining the characteristics of, and criteria for a value proposition which may be qualified as beneficial for the end user (see ISO/IEC WD TR 21221 Beneficial AI systems). Ideally this frame of reference would be widely adopted across jurisdictions. Operationalising these criteria in a consistent way across jurisdictions is likely to remain a challenge.

6.2.2 Stakeholder Participation

It is worth remembering that SDOs are private member organisations comprising private companies, research institutes, public establishments, and sometimes individual members. Their primary mandate is to provide a forum for developing – and negotiating – technical standards for use in a particular sector. The primary motivation of suppliers of goods and services, i.e. those parties expected ultimately to implement and adhere to these standards, to participate in voluntary standard-setting is predominantly commercial. i.e. to establish, consolidate or enhance their market positions in a particular sector. Accordingly, standard-setting processes are *a priori* structured to exchange technical and commercial arguments and – at least as long as they remain voluntary – do not have to satisfy the more varied and demanding requirements, e.g. for legality, legitimacy, due process, transparency, and inclusion, that apply to the development and implementation of binding laws and regulations. They are therefore not usually set up to allow for the inclusion of a wide range of stakeholders. Giving end users, especially consumers and small businesses, the opportunity to actively participate in the co-creation process of standard-setting, and influence technical standards that shape the value proposition of a product or service (ex-ante), usually remains a secondary consideration (for a contrast see ANSI's Consumer participation Fund in (5.3)). Instead, the role of these stakeholders is usually limited to an (ex-post) purchasing decision, which may over time produce *de facto*-standards that could in some cases supersede formal technical standards.

It is true that standard-setting processes that are intended to complement and/or implement legislation, and thus expected to become legally binding to some degree, are sometimes subject to particular procedural requirements that provide for an enhanced degree of public scrutiny and debate (see 4.1). In comparison with the level of public deliberation and stakeholder involvement typically associated with laws and regulations, the process of standardisation appears distinctly less suitable to decide on trade-offs involving public goods, or to determine acceptable levels of risk concerning fundamental rights (Gornet and Maxwell, 2024).

6.2.3 Enforceability

Perhaps the most frequently raised concern about the effectivenees of technical standards is about their enforcability. Technical standards are, *a*

priori voluntary, and failure to comply therefore not sanctionable. This continues to be the case even if technicals standards are incorporated into law (see 4). While the developer or deployer of a system that has been certified to be in conformity with relevant technical standards is granted a legal presumption of compliance with the law they are, in principle, still at liberty to fulfil the relevant legal requirements by other means. Therefore a developer or deployer of an AI/ML system which has not been certified cannot be deemed *a priori* to be in breach of the law. This may complicate the prosecution of, and enforcement against 'bad actors' intentionally deploying an uncertified system.

Adding to the complexity, a side-effect of establishing a presumption of compliance with the law through standards is that it could potentially raise the barrier for an end user to hold a developer or deployer legally responsible for any harm inflicted by their system. In addition to the information asymmetry between the developer and the end user, which applies to most products and services, the novelty and complexity of AI/ML technology pose particular challenges for end users, e.g. in demonstrating harm, establishing causality, and attributing legal responsibility. This technical complexity is compounded by the often equally complex supply chain, which may involve a number of different parties at different stages, e.g. the design and development of a foundational model, the implemenation of a specific application, training, testing, and deployment. Entry points for 'bad actors' exist at each stage. Moreover, as mentioned previously (see 2), other end users may be able to manipulate AI/ML systems to inflict harm, especially in the case of generative AI systems. In the EU, the potential difficulties for end users to obtain legal redress in such cases have instigated a debate about the need for additional legislation. The European Commission's proposal for an 'AI Liability Directive'[34], published in September 2022, was opposed by several member states and withdrawn in February 2025.

7. CONCLUSION

In the absence of a political consensus among major jurisdictions at the global level on regulating AI/ML, standard-setting plays important role. While the prospect of a broad regulatory consensus appears remote given the current geopolitical environment international cooperation at the technical level is essential to maintain a constructive dialogue, identify common ground, and preserve a degree of convergence. In particular, the international community should be aligned in seeking to protect their citizens against the misuse of AI/ML by 'bad actors'. The work of standardisation bodies on common criteria for 'trustworthy AI' could be an important contribution to this effort. At the same time, however, the inherent limitations of standard-setting imply that a global dialogue on regulation is still needed. While standards could play an important complementary role, they cannot realistically substitute for an international agreement on the beneficial use of AI/ML.

[34]European Commission, Proposal for a Directive on adapting non contractual civil liability rules to artificial intelligence, COM (2022) 496 (final), 28 September 2022.

ANNEX 1: ISO/IEC JTC 1/SC 42 WG 3 – WORK PROGRAMME

ISO/IEC SC 42 Artificial intelligence - WG 3 Trustworthiness - Work programme (P=Published, O=Open) - Retrievedlanuary, 2025

Trustworthiness - non-functional characteristics/requirements of an Al system - degree to which a user or other stakeholders has confidence that the system will behave as intended - characteristics/requirements are from ISO/IEC 24028:2020

Overview	P	ISO/IEC TR 24028:2020	Information technology - Artificial intelligence - Overview of trustworthinessin artificial intelligence
	0	ISO/IEC PWI 42117	Information technology - Artificial intelligence Trustworthiness fact labels for Al systems
Eaimess - absence of inappropriate bias	P	ISO/IECTR 24027:2021	Information technology - Artificial intelligence (Al) - Bias in Al systems and Al aided decision making
	P	ISO/IECTS 12791:2024	Information technology - Artificial intelligence - Treatment of unwanted bias in classification and regression machine learning
Transparency - provides visibility to the features, components and procedures of an Al system	0	ISO/IEC FDIS 12792	Information technology - Artificial intelligence - Transparency taxonomy of Al systems
Safety - freedeom from unacceptable risk	P	ISO/NEC TR 5469:2024	Artificial intelligence - Functional safety and Alsystems
Reliability - property of consistent intended behaviour and results	P	150/1EC 23894:2023	Information technology - Artificial intelligence - Guidance on risk management
	0	ISO/IEC PWI 42118	Information technology - Artificial intelligence - Reliability of Al systems
	0	ISO/IEC NP TS 25570	Information Technology - Artificial Intelligence - Reliability assessment of Al systems
	0	ISO/IEC NP TS 25568	Information technology - Artificial Intelligence - Guidance on addressing risks in generative Al systems
Controllability - can be achieved by providing reliable mechanisms by which an operator can take over control from the Al system	P	ISO/IEC TS 8200:2024	Information technology - Artificial intelligence Controllability of automated artificial intelligence systems
	0	ISO/IEC CD 42105	Information technology - Artificial intelligence - Guidance for human oversight of Al systems
	0	ISO/IEC PWI 18966	Artificial Intelligence - Oversight of Al systems
Explanability - Explanations of processes relevant to the development, implementation and use of an Al system to achieve effective transparency	0	ISO/IEC DTS 6254	Information technology - Artificial intelligence - Objectives and approaches for explai nability and interpretability of ML models and Al systems
Robustness - the ultimate ability of a system to maintain its level of	P	ISO/IEC TR 24029-1:2021	Artificial intelligence (A) - Assessment of the robustness of neural networks - Part 1: Overwiew
performance under any circumstances as it was intended by its developers	P	150/1EC 24029-2:2023	neural networks - Part 2: Methodology for the use of formal methods

	0	150/1EC AWI 24029-3	Artificial intelligence [A] - Assesment of the robusthess of neural networks - Part 3: Methodology for the use of statistical methods
	0	150/IEC PWI 24029-5	Artificial intelligence (A) - Assessment of the robustness of neural networks - Part 5: Applicability of the methodology to other artificial intelligence algorithms
Quality (software) - freedom from risks	P	ISO/IEC TS 25058:2024	Requirements and Evaluation (SQuaRE) - Guidance for quality evaluation of artificial intelligence (A) systems
	P	150/1EC 25059:2023	Software engineering - Systems and software Quality Requirements and Evaluation (SQuarE) - Quality model for Al systems
	0	ISO/IEC CD TR 42106	Information technology - Artificial intel ligence - Overview of differentiated benchmarking of A system quality characteristics
Ethical & Societal Concerns ability, integrity, benevolence	P	ISO/IEC TR 24368:2022	Information technology - Artificial intel ligence - Overview of ethical and societal concerns
	0	ISO/IEC NP TS 25571	Artificial Intelligence - Example template for documenting ethical issues of an Al system
	0	ISO/IEC AWI TS 22443	Information technology - Artificial intelligence - Guidance on addressing societal concerns and ethical considerations
Other characteristics	0	ISO/IEC PWI 42108	Artificial intelligence - Operational design domain (ODD) for Al systems
	0	ISO/IEC NP TS 25566	Terminology and concepts for domain engineering of Al systems
	0	ISO/IEC AWI 25029	Artificial intelligence - Al-enhanced nudging.

ANNEX 2: CEN/CLC/JTC 21 – WORK PROGRAMME

Project reference	Project title	Status
prEN ISO/IEC 12792 (WI=JT021022)	Information technology - Artificial intelligence - Transparency taxonomy of AI systems (ISO/IEC DIS 12792:2024)	Pending approval
prEN ISO/IEC 5259-1 (WI=JT021040)	Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 1: Overview, terminology, and examples (ISO/IEC 5259-1:2024)	Pending approval
prEN ISO/IEC 5259-2 (WI=JT021041)	Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 2: Data quality measures (ISO/IEC 5259-2:2024)	Pending approval
prEN ISO/IEC 5259-3 (WI=JT021042)	Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 3: Data quality management requirements and guidelines (ISO/IEC 5259-3:2024)	Pending approval
prEN ISO/IEC 5259-4 (WI=JT021043)	Artificial intelligence - Data quality for analytics and machine learning (ML) - Part 4: Data quality process framework (ISO/IEC 5259-4:2024)	Pending approval
prEN ISO/IEC TR 23281 (WI=JT021002)	Artificial Intelligence - Overview of Al tasks and functionalities related to natural language processing	Drafting
prEN XXX (WI=JT021008)	AI trustworthiness framework	Drafting
prEN ISO/IEC 23282 (WI=JT021012)	Artificial Intelligence - Evaluation methods for accurate natural language processing systems	Drafting

Continued

Project reference	Project title	Status
prEN XXX (WI=JT021019)	Competence requirements for AI ethicists professionals	Drafting
prEN ISO/IEC 24970 (WI=JT021021)	Artificial intelligence — AI system logging	Drafting
prEN XXX (WI=JT021024)	AI Risk Management	Drafting
prEN XXX (WI=JT021025)	Artificial Intelligence – Evaluation methods for accurate computer vision systems	Drafting
prEN ISO/IEC 25059 rev (WI=JT021027)	Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems (ISO/IEC 25059:2023)	Drafting
prEN XXX (WI=JT021029)	Artificial intelligence - Cybersecurity specifications for AI Systems	Drafting
EN ISO/IEC 22989:2023/prA1 (WI=JT021031)	Information technology — Artificial intelligence — Artificial intelligence concepts and terminology — Amendment 1	Drafting
EN ISO/IEC 23053:2023/prA1 (WI=JT021032)	Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) — Amendment 1	Drafting
prEN XXX (WI=JT021036)	Artificial Intelligence - Concepts, measures and requirements for managing bias in AI systems	Drafting
prEN XXX (WI=JT021037)	Artificial Intelligence -- Quality and governance of datasets in AI	Drafting
prEN XXX (WI=JT021038)	AI Conformity assessment framework	Drafting
prEN XXX (WI=JT021039)	Artificial intelligence - Quality management system for EU AI Act regulatory purposes	Drafting
prEN XXX (WI=JT021044)	Artificial Intelligence - Taxonomy of AI tasks in computer vision	Drafting
prEN ISO/IEC 42102 (WI=JT021045)	Information technology - Artificial intelligence – Taxonomy of AI system methods and capabilities	Drafting
prEN ISO/IEC 25029 (WI=JT021046)	Artificial intelligence - AI-enhanced nudging	Drafting
prCEN/CLC/TR XXX (WI=JT021009)	AI Risks - Check List for AI Risks Management	Prelim.
prEN ISO/IEC 42001 (WI=JT021011)	Information technology - Artificial intelligence - Management system	Prelim.
prEN ISO/IEC 24029-2 (WI=JT021015)	Artificial intelligence (AI) - Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods	Prelim.
prCEN/CLC/TR XXX (WI=JT021026)	Impact assessment in the context of the EU Fundamental Rights	Prelim.
prCEN/TS (WI=JT021033)	Guidance for upskilling organisations on AI ethics and social concerns	Prelim.
prCEN/TS (WI=JT021034)	Guidelines on tools for handling ethical issues in AI system life cycle	Prelim.
prCEN/TS (WI=JT021035)	Sustainable Artificial Intelligence – Guidelines and metrics for the environmental impact of artificial intelligence systems and services	Prelim.
(WI=JT021028)	Reference architecture of knowledge engineering based on ISO/IEC 5392	Prelim.
(WI=JT021030)	Contributions towards ISO/IEC 27090	Prelim.

REFERENCES

Cuccuru, P. (2019). Interest Representation in European Standardisation: The Case of CEN and CENELEC, Amsterdam Centre for European Law and Governance Research Paper No. 2019-06, 17 December 2019 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3505290).

Dufour, N., Arkanath, P., Pouya, S., Hariri, N., Shashi, D., Dudfield, A., Guess, C., Hernández Escayola, P., Tran, B., Mevan, B., Bregler, C. (2024). AMMEBA: A Large-Scale Survey and Dataset of Media-Based Misinformation In-The-Wild, 21 May 2024 (pre-print; unreviewed) (https://arxiv.org/pdf/2405.11697).

Gornet, M., Maxwell, W. (2024). The European approach to regulating AI through technical standards. Internet Policy Review, vol. 13/3 (https://doi.org/10.14763/2024.3.1784).

H. M. Government, Department for Science, Innovation & Technology (2024). Research and Analysis: Cyber security risks to artificial intelligence, 15 May 2024 (https://www.gov.uk/government/publications/research-on-the-cyber-security-of-ai/cyber-security-risks-to-artificial-intelligence#list-of-case-studies).

High-Level Expert Group on AI (AI HLEG) (2019). Ethics Guidelines for Trustworthy AI, European Commission, Brussels. (https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419)

Institute of Electrical and Electronics Engineers (IEEE) (2024). 2840-2024, Approved Draft Standard for Responsible AI Licensing (https://standards.ieee.org/ieee/2840/7673/).

International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC). ISO/IEC WD TR 21221, Beneficial AI Systems, (in progress).

International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC) (2020). ISO/IEC TR 24028:2020, Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence.

Kilian, R., Ebel, D., Jäck, L. (2025). European AI Standards – Technical Standardization and Implementation Challenges under the EU AI Act, 24 March 2025. (https://ssrn.com/abstract=5155591)

Kwan, S. K., Stiefmueller, C. M., Leitner, C. (2024). "Exploring Regulatory Frameworks for AI/ML through Different Lenses: A Comparative Approach" in: Leitner, C., Nägele, R., Bassano, C., Satterfield, D. (eds), The Human Side of Service Engineering: AHFE 2024 International Conference (Proceedings), AHFE Open Access, vol. 143 (http://doi.org/10.54941/ahfe1005080).

Leitner, C., Stiefmueller, C. M. (2025). "Chapter 3: Legal and Regulatory Frameworks for Digital Technologies" in: Baimenov, A., Liverakos, P. (eds.), Public Administration in the New Reality, Springer Nature, Singapore, pgs. 63–97. (http://doi.org/10.1007/978-981-96-3845-1_3)

National Institute of Standards and Technology (NIST) (2023a). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100–1. January 2023. (https://www.nist.gov/itl/ai-risk-management-framework)

National Institute of Standards and Technology (NIST) (2023b). NIST AI RMF Playbook. January 2023. (https://airc.nist.gov/airmf-resources/ playbook/)

National Institute of Standards and Technology (NIST) (2024). NIST AI 100–2 E2023. Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations, January 2024 (https://csrc.nist.gov/pubs/ai/100/2/e2023/final).

Soler Garrido, J., De Nigris, S., Bassani, E., Sanchez, I., Evas, T., André, A.-A., Boulangé, T. (2024). Harmonised Standards for the European AI Act, Joint Research Centre Science for Policy Brief JRC 139430, 24 October 2024 (https://publications.jrc.ec.europa.eu/repository/handle/JRC139430).

Stiefmueller, C. (2022). "The Soul of a New Machine: Promises and Pitfalls of Artificial Intelligence in Finance" in: Leitner, C., Ganz, W., Bassano, C., Satterfield, D. (eds), The Human Side of Service Engineering: AHFE 2022 International Conference (Proceedings), AHFE Open Access, vol. 62 (http://doi.org/10.54941/ahfe1002577).

Tovo, C. (2018). Judicial review of Harmonized Standards: Changing the paradigms of legality and legitimacy of private rulemaking under EU law, Common Market Law Review, vol. 55(4), pgs. 1187–1216 (http://doi.org/10.54648/cola2018096).

United Nations Organisation (UNO) (2024). Artificial Intelligence 'Must Serve Humanity Equitably, Safely', Secretary-General Stresses, in Message for International Day of Democracy, SG/SM/22347, United Nations, New York, NY. (https://press.un.org/en/2024/sgsm22347.doc.htm)

Note: All online sources as of 30 April 2025.