Foreseeing the Impact of the Proposed AI Act on the Sustainability and Safety of Critical Infrastructures

Foreseeing the Impact of the Proposed AI Act on the Sustainability and Safety of Critical Infrastructures

Francesco Sovrano*
Department of Computer Science and Engineering,
University of Bologna
Bologna, Italy
francesco.sovrano2@unibo.it

Giulio Masetti*
Istituto di Scienza e Tecnologia dell'Informazione -
Consiglio Nazionale delle Ricerche
Pisa, Italy
giulio.masetti@isti.cnr.it

ABSTRACT

The AI Act has been recently proposed by the European Commission to regulate the use of AI in the EU, especially on high-risk applications, i.e. systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity. On the other hand, IEC 61508, one of the most adopted international standards for safety-critical electronic components, seem to mostly forbid the use of AI in such systems. Given this conflict between IEC 61508 and the proposed AI Act, also stressed by the fact that IEC 61508 is not an harmonised European standard, with the present paper we study and analyse what is going to happen to industry after the entry into force of the AI Act. In particular, we focus on how the proposed AI Act might positively impact on the sustainability of critical infrastructures by allowing the use of AI on an industry where it was previously forbidden. To do so, we provide several examples of AI-based solutions falling under the umbrella of IEC 61508 that might have a positive impact on sustainability in alignment with the current long-term goals of the EU and the Sustainable Development Goals of the United Nations, i.e., affordable and clean energy, sustainable cities and communities.

KEYWORDS

AI Act, IEC 61508, safety standards, sustainability

1 INTRODUCTION

Harnessing the full potential of AI could lead our society to more efficient and thus sustainable energy production, storage and transportation, in accordance with many of the Sustainable Development Goals of the United Nations [1], including: affordable and clean energy (goal 7), industry, innovation and infrastructure (goal 9), sustainable cities and communities (goal 11), responsible consumption and production (goal 12), climate action (goal 13). Indeed, AI can be used to optimally recognise, predict, detect, identify, determine, control, generate, and classify [37] in a wide range of tasks, sometimes also achieving or exceeding human performance in problems such as strategy games [22, 44], image and object recognition [24, 41], etc. Nonetheless, the adoption of AI-based technological solutions for more sustainable energy (e.g., for decreasing the carbon emissions of coal-fired thermal power plants [45]) has been held back in the last decades by conservative international standards (i.e., IEC 61508 [46]: a standard that regulates safety-critical electronic components and that practically forbid AI in many critical infrastructures).

Despite this, in 2021, the European Commission published a proposal of AI Act[1] [12] that is expected to become a legally binding regulation to all the Member States of the EU by 2024. Importantly, the objective of the AI Act is to set a common regulatory and legal framework for AI that applies to all sectors (except for military), and to all types of artificial intelligence, including (high-risk) AI for the *management and operation of critical infrastructure.*

Considering that one of the goals of the proposed AI Act is to regulate the use of AI also on those systems covered by IEC 61508, i.e. 'systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity' (see Annex III, point 2.a), the research questions we are trying to answer with the present paper are the following:

- Will the AI Act be disruptive with respect to IEC 61508?
- What will happen to those industries currently regulated by IEC 61508?

In fact, we believe that answering these has the potential to help both industry and academia to quickly seize the opportunities offered by the new European policies enshrined in the AI Act.

In order to answer these questions, we analyse the main differences between IEC 61508 and the proposed AI Act. Then, we identify significant and concrete examples of technical solutions that could increase sustainability and energy efficiency but which, at the moment, are not feasible due to IEC 61508. Furthermore, we also study how such new technological solutions might impact industry, trying to understand how disruptive the AI Act could be by loosening the tight laces of IEC 61508 on AI. Hence, we try to align our findings to the medium- and long-term objectives of the EU on sustainability and support for innovation.

This paper is structured as follows. Section 2 discusses the adopted methodology. In Section 3 we give enough background to understand IEC 61508 and its implications for industry. While, in Section 4 we analyse the position of the AI Act on IEC 61508 and other standards, providing in Section 5 our understanding of how AI could improve sustainability and energy efficiency whilst maintaining safety. Finally, in Section 6 we try to give a conclusive answer to each research question, discussing the consequences of our findings as well as some possible issues.

2 METHODOLOGY

The adopted methodology employed for analysing the proposed AI Act and answering the aforementioned research questions is

*Both authors contributed equally to this research.

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

as follows. First of all, we start from the identification of climate-neutrality (by 2050) as one of the main and most actual objectives of the EU. In particular, we refer to the European Green Deal[2] and the EU's commitment to global climate action under the Paris Agreement, considering them as interpretative keys for the proposed AI Act. Hence, we study how articles such as article 54.1.a, interpreted under the lenses of the European Green Deal, may impact on the adoption of new AI-based technologies in industrial contexts currently regulated by non-harmonised technical standards (i.e., IEC 61508) that practically forbid the use of AI. We do it by looking at how the AI Act may help to improve the sustainability of our society, thanks to state-of-the-art advancements in AI that cannot currently be deployed.

3 BACKGROUND

With this section we provide a minimal amount of information about safety, AI, the IEC 61508 standard and the proposed AI Act.

3.1 The safety standard IEC 61508

IEC 61508 [28] is an international standard describing how to design, deploy and maintain an Electrical or (Programmable) Electronic safety-related system. Examples of safety-related systems to which IEC 61508 can be applied are: emergency shut-down systems, remote monitoring, operation or programming of a network-enabled process plant, information-based decision support tool where erroneous results may affect safety. In particular, programmable electronic safety-related systems typically incorporate programmable controllers, programmable logic controllers, microprocessors, application specific integrated circuits, or other programmable devices (e.g., 'smart' devices such as sensors/transmitters/actuators). The focus is in particular on safety functions and on the relative level of risk reduction that they provide. Those levels are grouped in four Safety Integrity Levels (SILs), the higher the Safety Integrity Level the greater the risk of failure.

Notice that it is expected[3] that IEC 61508 can be published as EN 61508, an European standard, but it does not have the status of a *harmonized* European standard in relation to any EC product directive and it is not therefore listed in the EC Official Journal. However, this does not prevent compliance with relevant parts of EN 61508 being used to support a declaration of conformity with an EC product directive, if that is appropriate. In any cases, IEC 61508 is followed worldwide[4].

3.2 Safety vs AI

Quoting [47]: ' There is no such thing as zero risk. This is because no physical item has zero failure rate, no human being makes zero errors, and no piece of software design can foresee every operational possibility '. Thus, perfect *safety*, i.e., the absence of catastrophic consequences on the user(s) and the environment [5], is out of reach. During the last decades, several standards on how to develop hardware and software artefacts in safety critical contexts have been defined. These standards crystallize lessons learned, common practices and scientific research into concrete guidelines. Each industry sector has its own standard, but the idea behind all of them is the same: a risk-based approach that characterize the entire product life-cycle.

With respect to Artificial Intelligence (AI), at row 5 in Tables A.2 and C.2, Part 3, of IEC 61508 [28], it is clearly stated that AI is not recommended for Safety Integrity Level 2 or above because it may complicate the achievement of one or more of the following properties: correctness with respect to software safety requirements specification, freedom from intrinsic design faults, simplicity and understandability, related with the observability-in-depth principle, aimed at avoiding as much as possible a false sense of safety due to lack of information, predictability of behaviour, verifiable and testable design.

IEC 61508 has influenced other standards [47], here called 'second tier standards', that are as rigid as IEC 61508 Part 3 with respect to AI. Among those, examples are software for train EN 50128 [15], process industry [29] and machinery IEC 62061 [30]. Parallel to the family of standards originated from IEC 61508, other really important examples where AI is banned, for high Safety Integrity Level, from computer-based systems employed in nuclear power plants, IEC 60880 [27], and avionic, DO-178 C [3].

To the best of the authors' knowledge, the only safety standard that allows the employment of AI (because it does not mention it, and then it is not 'not recommended') is ISO 26262 for the automotive industry sector [19, 25, 43].

3.3 The Proposed AI Act

The AI Act [12] is a proposed European law on AI. Differently from other domains, this act is specific to AI systems and requires an *ad hoc* discussion rather than the framing of these systems in the discussion of other legal domains. This is because AI technologies are not placed within an existing legal framework (e.g., banking), but the whole legal framework (i.e., the proposed AI Act) is built around AI technologies.

The proposed AI Act assigns applications of AI to three risk categories. First, applications and systems that create an unacceptable risk, such as government-run social scoring of the type used in China, are banned. Second, high-risk applications, such as a CV-scanning tool that ranks job applicants, are subject to specific legal requirements. Lastly, applications not explicitly banned or listed as high-risk are largely left unregulated.

Examples of high-risk AI are given by the proposed AI Act in Annex III, as they broadly include applications for: biometric identification and categorisation of natural persons, management and operation of critical infrastructures, etc. In particular, for all those applications defined as 'high-risk', the AI Act provides several limitations and safety assurance procedures including: a risk management system (art. 9), appropriate data governance and management practices (art. 10), detailed technical documentation (art. 11).

Finally, the AI Act defines in Annex I what are the AI techniques and approaches referred by the proposal. Among them we have: machine learning approaches (e.g., neural networks), logic- and knowledge-based approaches (e.g., inductive logic programming), and statistical approaches (e.g., Bayesian estimation).

[2] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2019:640:FIN
[3] https://www.iec.ch/functional-safety/faq
[4] https://www.iec.ch/national-committees

4 ANALYSIS OF THE POSITION OF THE AI ACT ON IEC 61508 AND OTHER STANDARDS

It is crystal clear from the European Green Deal and the EU's commitment to global climate action under the Paris Agreement that one of the big goals of the EU is to be climate-neutral by 2050. Citing the words of the European Commission: 'The EU can lead the way [to climate-neutrality] by investing into realistic technological solutions, empowering citizens and aligning action in key areas such as industrial policy, finance and research, while ensuring social fairness for a just transition.' The reason why we are citing these statements is that we are going to use them as interpretative key for the proposed AI Act, especially with respect to the importance of article 54.1.a, stating that 'innovative AI systems shall be developed for safeguarding substantial public interest in [...] a high level of protection and improvement of the quality of the environment'.

In fact, the AI Act is (as mentioned in Section 3) regulating a vast range of AI applications, with due focus on those listed as high-risk in Annex III. In particular, it covers, among others, the AI applications for the 'management and operation of critical infrastructure', i.e. the 'AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.' But, considering that many critical infrastructures are currently following a non-harmonized IEC 61508 standard that is de facto excluding the involvement of any AI, we can see a non-alignment of it to the proposed AI Act.

Indeed, according to article 40 only the 'harmonised standards or parts thereof the references of which have been published in the Official Journal of the European Union' are considered to be in conformity with the requirements set out in the proposed AI Act for high-risk AI systems (see Chapter 2 of Title III). In other words, article 40, together with the *Explanatory Memorandum*, article 54.1.a and the fact that IEC 61508 is a non-harmonised standard, make us understand that the intent of the proposed AI Act is to promote innovative AI systems also for the 'management and operation of road traffic and the supply of water, gas, heating and electricity'. As consequence, we envisage that the proposed AI Act, without further modifications, can have a disruptive effect in the industry of critical infrastructures. This effect can be disruptive in a positive way, by opening to new technological solutions that have the potential to improve even further our quality of life, reducing costs and increasing efficiency. Nonetheless, it can be disruptive also in a negative way, by ceding the control of critical infrastructures to automatic decision makers that are possibly opaque, greedy, unfair and non-transparent in a way that would not allow to understand where the responsibility lies.

Although, despite the fact that IEC 61508 is a non-harmonized standard, thus not covered by article 40, we can see that the proposed AI Act shares several and important similarities with it, suggesting that it is not the intent of the EU Commission to fully upset existing standards.

Overall, we see that the intent of the proposed AI Act is to modernize existing critical infrastructures, to make them more sustainable. To do so, the AI Act does not ignore or try to eliminate the currently adopted standards, although it wants them harmonised with the EU's policies. This is why the CEN-CENELEC has established a joint technical committee on AI[5] and defined a road map for AI standardization [2] that includes the harmonization of IEC 61508 and other standards. In fact, according to article 2(1)(c) of Regulation (EU) No 1025/2012, the CEN-CENELEC is the European Union (EU) authority for standards. Nonetheless, we can see also that European countries start producing guidelines and roadmaps [48] on this subject.

So, given the very clear position of the proposed AI Act with respect to the possibility of using AI systems in particular critical infrastructures, we believe that the CEN-CENELEC, together with IEC will adapt IEC 61508, eventually opening to a safe use of AI systems also in critical infrastructures. Importantly, the CEN-CENELEC [2] has already identified article 41 as a possible source of uncertainty in industry, given that it would explicitly cut out any non-harmonized IEC standard (i.e., IEC 61508). In fact, article clearly 41.1 says that: 'Where harmonised standards referred to in Article 40 do not exist or where the Commission considers that the relevant harmonised standards are insufficient or that there is a need to address specific safety or fundamental right concerns, the Commission may, by means of implementing acts, adopt common specifications in respect of the requirements set out in Chapter 2 of [Title III]. Those implementing acts shall be adopted in accordance with the examination procedure referred to in Article 74(2).'

Consequently, given all the aforementioned facts, we see an harmonization of IEC 61508 or its replacement by 2024, and this will open to at least one of two scenarios. In the first scenario, we will have an opening to the use of AI systems in the context of critical infrastructures, whereas they can improve sustainability whilst guaranteeing safety. While in a second scenario a very strict policy against AI systems in critical infrastructures will be maintained.

Again, as consequence of the analysis presented in this Section, we believe that this very first scenario is the most likely. If that is correct, we envisage that a new stream of research on AI for critical infrastructures will be opening by the end of 2024, paving the way for AI systems to improve the sustainability of our society. Nonetheless, it is important to stress that the use of AI does not come free of problems related to safety, fairness, transparency and sometimes even sustainability. For this reason, in the following section we will discuss and classify existing AI techniques, to analyse their impact on sustainability and safety and to understand which AI-based solutions are likely to be allowed by a future harmonization of IEC 61508.

5 CLASSIFICATION AND DISCUSSION OF THE IMPACT OF AI ON SUSTAINABILITY AND SAFETY

As mentioned in Section 3.3, the techniques and approaches covered by the proposed AI Act include both symbolic (e.g., logic-based) and non-symbolic (e.g., neural networks, statistics) techniques. Nonetheless, each different type of AI may have its own characteristics, impacting on safety differently from others. Indeed, as suggested by Mohseni et al. in their taxonomy of machine learning safety [36], the decisions of state-of-the-art machine learning techniques can be,

[5] https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence

Table 1: AI Act vs IEC 61508: AI Act is centred on transparency while IEC 61508 on safety. This table shows how the proposed AI Act and IEC 61508 address the same process for risk-assessment, analysis, development and document production in different ways.

	Differences	
	IEC 61508	**Proposed AI Act**
Risk-based approach, in particular to establish the belongings to a predefined category	Quantitative (hazard analysis, risk assessment and identify the Safety Integrity Level)	Qualitative (one of the alternatives: no risk, application listed in Annex III, AI not applicable)
Normalised life cycle, with focus on accountability	V-shape development (focus on modularity and decomposability)	Clear definition of datasets (focus on data management, in particular for training the AI and how to use the product)
Ex-ante and ex-post analysis	Statistical methods (hardware), study of qualitative techniques (hardware and software) and structured testing campaigns	Declarative (identify high level characteristics, provide general description of components behaviour)
Document production	Assessment performed by an external institution	Fill a form in EU database, part of the information is of public domain (focus on transparency)

in some cases, completely unexplainable, non-transparent, biased and non-robust. On the other hand, the automatic decisions of fully symbolic approaches can be explainable by design but not as good as those of a state-of-the-art neural network [26]. Therefore, given such a trade-off between explainability and performance, being able to foresee and analyse the impact of AI on safety is not trivial, forcing us to analyse it differently for different types of applications.

This is why in the present paper we will study the impact of AI, on the safety and sustainability of critical infrastructures, by using as reference point the 4 Safety Integrity Level defined by IEC 61508. In fact, for each safety level we will show concrete examples of technological solutions based on AI that have the potential for significantly improving sustainability, analysing what is the trade-off between sustainability and safety and how important that is.

5.1 Examples of "Forbidden" AI-based Solutions that Could Improve Sustainability in Safety-Critical Systems

Safety-critical subsystems of cyber-physical systems[6] compliant with IEC 61508 are required to have the properties listed in section 3.1 and these normally do not include AI. Nonetheless, few examples can illustrates how impactful can be AI on safety-critical systems, considering that in scientific and technical literature are available several studies that directly address the issue or propose promising approaches that well fit the kind of data relevant for safety-critical functions. In table 2 we show these examples aligned to the Safety Integrity Level of IEC 61508.

Safety Integrity Level 4 systems compliant to IEC 61508 are quite rare. Nevertheless, the nuclear power plants industry offers examples of such systems [33]. Here, AI is envisioned to have a great impact in the relatively close future, in particular for safeguard and surveillance (filter and identify signatures of nuclear materials), monitoring and diagnosis of severe accidents or nuclear power plant transients [13]. All these actions are crucial to avoid environmental consequences of accidents and lives lost [23, 38].

For Safety Integrity Level 3 consider the railway industry, and in particular those systems for which energy efficiency is crucial, with focus on the heating system for rail-road switches [11]. This is a critical subsystem, responsible for keeping the switches free from snow and ice, necessary to guarantee the correct operation of the switches and so the correct train routing (always turned on increases safety). Depending on the climate conditions of the place where the railway system operates, the energy consumed by this heating system can be very relevant, i.e. the heating always turned on implies greater ambient impact and cost. To provide concrete examples, in [40] it is reported that the cost for heating the 6800 switches and crosses in Sweden can amount to $10-15$ Million Euros per year. In Germany, Deustche Bahn (DB) alone has 64000 switches heated with electrical resistance and gas heaters, a combined power of 900 MW which consume up to 230 GWh/year [39]. AI is expected to empower this subsystem with snow or ice prediction/detection and by making the turning on/off algorithm more responsive.

Regarding Safety Integrity Level 2 and 1, [7] shown concrete examples in the process industry, where the second tier standard IEC 61511 [29] apply, that can be generalized to the chemical industry. The chemical industry is one of the most energy-intensive manufacturing industries and a major source of greenhouse gas emissions. Besides that, chemical production often involves hazardous materials and high-pressure/high-temperature conditions, which may lead to fire, explosion, and other types of chemical accidents. Those chemical accidents could cause casualties, financial and social losses [34]. According to a survey conducted by Accenture [4], most of the companies in the chemical and advanced materials industry expect an industry-wide digitisation, and AI plays an essential role in enabling the digital revolution [17]. In particular, fault detection and diagnosis is crucial to both safety and sustainability. As an example, consider fault detection for a Tennessee Eastman process (chapter 8 of [16]) with few modes, where unit operations include

[6] A cyber-physical system comprises physical mechanisms that are monitored and/or controlled by Information and Communication Technologies.

Table 2: AI on Safety-Critical Environments: This table shows examples of possible applications of AI on some safety-critical contexts. For each context we identify its Safety Integrity Level (SIL) and possible tasks where AI can be deployed to improve sustainability.

SIL	Context	Use of AI
4	Nuclear power plant [23]	• Anomaly detection [8, 9] • In-core full management [38]
3	Railway, station management [39, 40]	• Turning on/off switch heaters [11] • Fault detection of sensitive components [6]
2 & 1	Chemical industry [4, 17]	• Predicting chattering alarms [49] • Plant health diagnosis [51]

a reactor, a condenser, a recycle compressor, a vapour-liquid separator and a stripper [51]. Notice that the adoption of AI is not 'not recommended' for Safety Integrity Level 1 in IEC 61508.

Indeed, AI has the potential to cope with high dimensional data, being able to generalise, handling novel inputs and incomplete knowledge [32]. These features are expected [50] to greatly impact the way goals and targets in the 2030 Agenda for Sustainable Development are addressed.

Overall, we can say that AI may be critical to anomaly detection, for taking timely countermeasures, being able to find patterns in data that do not conform to expected behaviour [10].

5.2 Discussion

Even though several metrics for AI performance and robustness appeared in literature and have been tested in several contexts [52], only preliminary ones have been defined specifically to address safety or sustainability (e.g., [18, 20]), and are yet to be tested extensively before some AI can become amenable for safety critical applications (where quantification has a central role). Thus, it is expected that those AI for which will be available reliable metrics will be the first to be employed in safety functions or safety critical systems.

It is desirable that interpretable or explainable-by-design AI [35] are the first to be employed, in particular for handling tabular data [42]. This is indeed expected to cover, at least in part, simplicity, understandability and observability-in-depth (section 3.2).

The heart of the problem is that AI is difficult to be framed in safety standards because of the way it fails. Deterministic software fails systematically, whereas hardware fails randomly [47]. Safety standards recommend to address hardware failures through statistical methods and mitigate/tolerate deterministic software failures employing qualitative techniques. In some standards, statistical methods for quantifying software failures are allowed (e.g., suggestions are provided in Part 7 of IEC 61508 [28]) in others (e.g., DO-178 C [3]) are not recommended. After about forty years of discussions, in industry and academia, no consensus has been reached, and strong opinions continue to emerge [14].

Among those listed as AI in Annex I of the proposed AI Act, some (e.g., statistical models or neural networks) are intrinsically non-deterministic [31], and then does not fit current safety standards framework. Seen from a different perspective, though, this removes many of the assumptions that prevent the use of statistical methods, opening up new ways to address AI failures. Indeed, a positive by-product of the discussions on statistical methods for deterministic software is the huge body of knowledge that is available but not enough explored for addressing non-deterministic software.

6 CONCLUSIONS

First of all, with this paper we performed an analysis of how the proposed AI Act might impact on the sustainability and safety of critical systems (e.g., power plants). We did it by looking at the differences, incompatibilities and similarities of the AI Act with IEC 61508, one of the most important non-harmonised standards for safety-critical infrastructures. Importantly, among the main differences, we show the incompatibility of IEC 61508 with the use of any AI in systems requiring a Safety Integrity Level greater than 1, pointing to the disruptive effect that the proposed AI Act might have on that part of industry aligned with IEC 61508. Then, we identified examples of AI-based solutions falling under the umbrella of IEC 61508 with a Safety Integrity Level greater than 1 that might have a positive impact on sustainability in alignment with the current long-term goals of the EU and the proposed AI Act.

Eventually, we collected enough material to answer our initial research questions and foresee a future where critical infrastructures may harness the full potential of AI to improve both sustainability and safety in accordance with the following Sustainable Development Goals of the United Nations [1]: affordable and clean energy (goal 7), industry, innovation and infrastructure (goal 9), sustainable cities and communities (goal 11), responsible consumption and production (goal 12), climate action (13). To be more precise, in accordance with the analysis we carried out in this paper, we believe that the AI Act will eventually soften the position of IEC 61508 with respect to AI, leading to a new generation of critical infrastructures harmonised with the European vision embodied by the proposed AI Act. This would clearly open to new research and technological solutions on this topic by the end of 2024.

Overall, with this paper, our focus was exclusively on those safety-critical contexts where AI is expected to enhance economic/environmental aspects of sustainability but is not employed yet because considered not enough mature or potentially in conflict with safety or technical aspects of sustainability, as per IEC 61508. Nonetheless, despite the promises made by state-of-the-art AI we can sceptically argue that using AI in safety-critical systems does definitely come with a risk. This risk is posed by the fact that ceding control to machines might lead to new unregulated unethical and immoral behaviours as well as a dangerous lack of transparency and accountability. Importantly, with respect to this specific issue, there are several flourishing discussions in literature and among policy

makers, also taking into account that similar issues are addressed in other contexts as well [21]. This gives us hope that the technology of the future will be able to cope with such urgent problems to give us solutions based on AI capable of addressing the sustainability goals that have been set for the future. For this reason, we argue that any forthcoming harmonised version of IEC 61508 is unlikely to completely close to application of AI in safety-critical systems with a Safety Integrity Level greater than 1. This is why we are all waiting for the CEN-CENELEC and its technical commission to give us a final answer to our research questions in the form of new harmonised standards.

REFERENCES

[1] 2016. *The Sustainable Development Goals Report.* Technical Report. United Nations. https://unstats.un.org/sdgs/report/2016/The%20Sustainable%20Development%20Goals%20Report%202016.pdf

[2] 2020. Road Map on Artificial Intelligence (AI). https://www.standict.eu/sites/default/files/2021-03/CEN-CLC_FGR_RoadMapAI.pdf

[3] RTCA (Firm). SC 167. 1992. *Software considerations in airborne systems and equipment certification.* RTCA, Incorporated.

[4] Accenture. 2016. Global Digital Chemistry - Survey quantitative findings. https://www2.deloitte.com/content/dam/Deloitte/de/Documents/consumer-industrial-products/Deloitte%20Global%20Digital%20Chemistry%20Survey2016Extract.pdf

[5] Algirdas Avizienis, Jean-Claude Laprie, Brian Randell, and Carl E. Landwehr. 2004. Basic Concepts and Taxonomy of Dependable and Secure Computing. *IEEE Trans. Dependable Secur. Comput.* 1, 1 (2004), 11–33. https://doi.org/10.1109/TDSC.2004.2

[6] Raul Barbosa, Stylianos Basagiannis, Georgios Giantamidis, H. Becker, Enrico Ferrari, J. Jahic, A. Kanak, Mikel Labayen Esnaola, Vanessa Orani, David Pereira, Luigi Pomante, Rupert Schlick, Ales Smrcka, Ahmet Yazici, Peter Folkesson, and Behrooz Sangchoolie. 2020. The VALU3S ECSEL Project: Verification and Validation of Automated Systems Safety and Security. In *23rd Euromicro Conference on Digital System Design, DSD 2020, Kranj, Slovenia, August 26-28, 2020.* IEEE, 352–359. https://doi.org/10.1109/DSD51259.2020.00064

[7] D. Barone and A. Damiani. 2016. Esperienza pratica nella applicazione delle analisi SIL (IEC 61508/61511) relative ai sistemi di sicurezza ad alta affidabilità, per uno stabilimento a rischio di incidente rilevante. (2016). http://conference.ing.unipi.it/vgr2016/images/papers/133.pdf Valutazione e Gestione del Rischio negli Insediamenti Civili ed Industriali.

[8] Roger Boza. 2019. *Subtle Process Anomalies Detection using Machine Learning Methods.* Technical Report. U.S. Department of Energy, Office of Nuclear Energy. https://lwrs.inl.gov/Advanced%20IIC%20System%20Technologies/Subtle_Process-Anomalies_Detection_Using_Machine-Learning_Methods.pdf

[9] Francesco Calivá, Fabio De Sousa Ribeiro, Antonios Mylonakis, Christophe Demazière, Paolo Vinai, Georgios Leontidis, and Stefanos D. Kollias. 2018. A Deep Learning Approach to Anomaly Detection in Nuclear Reactors. In *2018 International Joint Conference on Neural Networks, IJCNN 2018, Rio de Janeiro, Brazil, July 8-13, 2018.* IEEE, 1–8. https://doi.org/10.1109/IJCNN.2018.8489130

[10] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly Detection: A Survey. *ACM Comput. Surv.* 41, 3, Article 15 (jul 2009), 58 pages. https://doi.org/10.1145/1541880.1541882

[11] Silvano Chiaradonna, Giulio Masetti, Felicita Di Giandomenico, Francesca Righetti, and Carlo Vallati. 2021. Enhancing sustainability of the railway infrastructure: Trading energy saving and unavailability through efficient switch heating policies. *Sustain. Comput. Informatics Syst.* 30 (2021), 100519. https://doi.org/10.1016/j.suscom.2021.100519

[12] European Commission. 2021. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Inteligence and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[13] European Commission, Joint Research Centre, J Tanarro Colodron, K Simola, A Liessens, G Joanny, G Renda, J Colle, J Griveau, S Vigier, H Gerbelova, D Vanleeuw, M Cihlar, and A Cambriani. 2022. *Horizon scanning for nuclear safety and security yearly report 2021 : creating an anticipatory capacity within the JRC.* Publications Office of the European Union. https://doi.org/doi/10.2760/645368

[14] D. Daniels and N. Tudor. 2022. Software Reliability and the Misuse of Statistics. *Safety-Critical Systems eJournal* 1, 1 (2022). https://scsc.uk/journal/index.php/scsj/article/view/8

[15] European Committee for Electrotechnical Standardization 2020. *Railway applications - Communication, signalling and processing systems - Software for railway control and protection systems.* European Committee for Electrotechnical Standardization.

[16] Richard D. Braatz Evan L. Russell, Leo H. Chiang. 2000. *Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes.* https://doi.org/10.1007/978-1-4471-0409-4

[17] World Economic Forum. 2017. Digital transformation initiative chemistry and advanced materials industry. http://reports.weforum.org/digital-transformation/wp-content/blogs.dir/94/mp/files/pages/files/white-paper-dti-2017-chemistry.pdf

[18] Mohamad Gharib and Andrea Bondavalli. 2019. On the Evaluation Measures for Machine Learning Algorithms for Safety-Critical Systems. In *15th European Dependable Computing Conference, EDCC 2019, Naples, Italy, September 17-20, 2019.* IEEE, 141–144. https://doi.org/10.1109/EDCC.2019.00035

[19] Mohamad Gharib, Paolo Lollini, Marco Botta, Elvio Gilberto Amparore, Susanna Donatelli, and Andrea Bondavalli. 2018. On the Safety of Automotive Systems Incorporating Machine Learning Based Components: A Position Paper. In *48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops, DSN Workshops 2018, Luxembourg, June 25-28, 2018.* IEEE Computer Society, 271–274. https://doi.org/10.1109/DSN-W.2018.00074

[20] Mohamad Gharib, Tommaso Zoppi, and Andrea Bondavalli. 2021. Understanding the properness of incorporating machine learning algorithms in safety-critical systems. In *SAC '21: The 36th ACM/SIGAPP Symposium on Applied Computing, Virtual Event, Republic of Korea, March 22-26, 2021*, Chih-Cheng Hung, Jiman Hong, Alessio Bechini, and Eunjee Song (Eds.). ACM, 232–234. https://doi.org/10.1145/3412841.3442074

[21] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam. 2021. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet, digital health* 3 (2021), E745–E750. Issue 11.

[22] Elizabeth Gibney et al. 2016. Google AI algorithm masters ancient game of Go. *Nature* 529, 7587 (2016), 445–446.

[23] Mario Gomez-Fernandez, Kathryn Higley, Akira Tokuhiro, Kent Welter, Weng-Keen Wong, and Haori Yang. 2020. Status of research and development of learning-based approaches in nuclear science and engineering: A review. *Nuclear Engineering and Design* 359 (2020), 110479. https://doi.org/10.1016/j.nucengdes.2019.110479

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision.* 1026–1034.

[25] Jens Henriksson, Markus Borg, and Cristofer Englund. 2018. Automotive Safety and Machine Learning: Initial Results from a Study on How to Adapt the ISO 26262 Safety Standard. In *1st IEEE/ACM International Workshop on Software Engineering for AI in Autonomous Systems, SEFAIAS@ICSE 2018, Gothenburg, Sweden, May 28, 2018*, Reinhard Stolle, Stephan Scholz, and Manfred Broy (Eds.). ACM, 47–49. https://doi.org/10.1145/3194085.3194090

[26] Andreas Holzinger. 2018. From Machine Learning to Explainable AI. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA).* 55–66. https://doi.org/10.1109/DISA.2018.8490530

[27] International Electrotechnical Commission 2006. *Nuclear power plants – Instrumentation and control systems important to safety – Software aspects for computer-based systems performing category A functions.* International Electrotechnical Commission.

[28] International Electrotechnical Commission 2010. *Functional safety of electrical/electronic/programmable electronic safety-related systems.* International Electrotechnical Commission.

[29] International Electrotechnical Commission 2016. *Functional safety - Safety instrumented systems for the process industry sector.* International Electrotechnical Commission.

[30] International Electrotechnical Commission 2021. *Safety of machinery - Functional safety of safety-related control systems.* International Electrotechnical Commission.

[31] Bonnie Johnson. 2022. Metacognition for artificial intelligence system safety – An approach to safe and desired behavior. *Safety Science* 151 (2022), 105743. https://doi.org/10.1016/j.ssci.2022.105743

[32] Zeshan Kurd, Tim Kelly, and Jim Austin. 2007. Developing artificial neural networks for safety critical systems. *Neural Comput. Appl.* 16, 1 (2007), 11–19. https://doi.org/10.1007/s00521-006-0039-9

[33] Jussi Lahtinen, Mika Johansson, Jukka Ranta, Hannu Harju, and Risto Nevalainen. 2010. Comparison between IEC 60880 and IEC 61508 for Certification Purposes in the Nuclear Domain. In *Computer Safety, Reliability, and Security, 29th International Conference, SAFECOMP 2010, Vienna, Austria, September 14-17, 2010. Proceedings (Lecture Notes in Computer Science)*, Erwin Schoitsch (Ed.), Vol. 6351. Springer, 55–67. https://doi.org/10.1007/978-3-642-15651-9_5

[34] M. Liao, K. Lan, and Y. Yao. 2022. Sustainability implications of artificial intelligence in the chemical industry: A conceptual framework. *Journal of industrial ecology* 26, 1 (2022), 164–182.

[35] Ricards Marcinkevics and Julia E. Vogt. 2020. Interpretability and Explainability: A Machine Learning Zoo Mini-tour. *CoRR* abs/2012.01805 (2020). arXiv:2012.01805 https://arxiv.org/abs/2012.01805

[36] S. Mohseni, H. Wang, Z. Yu, C. Xiao, Z. Wang, and J. Yadawa. 2021. Taxonomy of Machine Learning Safety: A Survey and Primer. https://arxiv.org/abs/2106.04823

[37] Mark Munro, Jacob Whiton, and Robert Maxim. 2019. What jobs are affected by AI? (2019).

[38] Ephraim Nissan. 2019. An Overview of AI Methods for in-Core Fuel Management: Tools for the Automatic Design of Nuclear Reactor Core Configurations for Fuel Reload, (Re)arranging New and Partly Spent Fuel. *Designs* 3, 3 (2019). https://doi.org/10.3390/designs3030037

[39] International Union of Railways. [n.d.]. Technologies and Potential Developments for Energy Efficiency and CO2 Reduction in Rail Systems. https://uic.org/IMG/pdf/_27_technologies_and_potential_developments_for_energy_efficiency_and_co2_reductions_in_rail_systems._uic_in_colaboration.pdf Online; accessed 15 January 2019.

[40] A. Parida P. Norbbin, J. Lin. 2016. Energy efficiency optimization for railway switches & crossings: a case study in Sweden. In *WCRR 2016, 11th World Congress on Railway Research*. SPARK knowledge sharing portal. https://www.diva-portal.org/smash/get/diva2:1010747/FULLTEXT01.pdf

[41] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017).

[42] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 5 (2019), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[43] Rick Salay, Rodrigo Queiroz, and Krzysztof Czarnecki. 2017. An Analysis of ISO 26262: Using Machine Learning Safely in Automotive Software. *CoRR* abs/1709.02435 (2017). arXiv:1709.02435 http://arxiv.org/abs/1709.02435

[44] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, et al. 2020. Mastering atari, go, chess and shogi by planning with a learned model. *Nature* 588, 7839 (2020), 604–609.

[45] T Sivageerthi, Bathrinath Sankaranarayanan, Syed Mithun Ali, and Koppiahraj Karuppiah. 2022. Modelling the Relationships among the Key Factors Affecting the Performance of Coal-Fired Thermal Power Plants: Implications for Achieving Clean Energy. *Sustainability* 14, 6 (2022), 3588.

[46] David J Smith and Kenneth GL Simpson. 2020. *The Safety Critical Systems Handbook: A Straightforward Guide to Functional Safety: IEC 61508 (2010 Edition), IEC 61511 (2015 Edition) and Related Guidance*. Butterworth-Heinemann.

[47] David J. Smith and Kenneth G. L. Simpson (Eds.). 2020. *The Safety Critical Systems Handbook* (fifth edition ed.).

[48] DKE standards. 2020. German standardization roadmap on artificial intelligence. https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf

[49] Nicola Tamascelli, Nicola Paltrinieri, and Valerio Cozzani. 2020. Predicting chattering alarms: A machine Learning approach. *Comput. Chem. Eng.* 143 (2020), 107122. https://doi.org/10.1016/j.compchemeng.2020.107122

[50] Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Langhans, Max Tegmark, and Francesco Fuso Nerini. 2019. The role of artificial intelligence in achieving the Sustainable Development Goals. *CoRR* abs/1905.00501 (2019). arXiv:1905.00501 http://arxiv.org/abs/1905.00501

[51] Hao W. and Jinsong Z. 2020. Fault detection and diagnosis based on transfer learning for multimode chemical processes. *Computers & Chemical Engineering* 135 (2020), 106731.

[52] T. Wu, Y. Dong, Z. Dong, A. Singa, X. Chen, and Y. Zhang. 2020. Testing Artificial Intelligence System Towards Safety and Robustness: State of the Art. *International Journal of Computer Science* 47, 3 (2020).