Advances in Reliability, Safety and Security, Part 5 – Kołowrocki, Kosmowski (eds.)
 ISBN 978-83-68136-17-3 (printed), ISBN 978-83-68136-04-3 (electronic)

Advances in Reliability, Safety and Security

ESREL 2024
Monograph Book Series

How European Artificial Intelligence Act Impacts On Machinery Products

Luigi Monica[a], Marianna Madonna[b], Silvia Carra[a], Sara Anastasi[a]

[a] *Department of Technological Innovations and Safety of Plants, Products and Anthropic Settlements, INAIL, Rome, Italy*
[b] *Operational Territorial Unit of Naples - Certification, Verification and Research Area, INAIL, Naples, Italy*

Abstract

In light of the rise of generative artificial intelligence, the European Parliament recently reached a provisional agreement with the Council on the Artificial Intelligence Act (AI Act) that will soon become EU law. This regulation was the first major global attempt by a government body to address and mitigate the potentially negative impacts of artificial intelligence technologies, ensuring respect existing law on fundamental rights and Union values. On the other hand, Europe promotes innovation and strengthens itself as a leader in this field. The AI Act was designed as a horizontal EU legislative instrument applicable to all AI systems placed on the market or used in the Union, providing a uniform definition of AI that could be applied to all future AI systems in order to ensure consistency with existing policy provisions in the sector. The proposed regulation was designed to be complementary to cross-sectoral EU legislation, harmonized in the New Legislative Framework (NLF), in particular the updated Machinery Regulation. This paper aims to highlight the risks that could arise by incorporating an AI application into machines, those arising from generative AI which would modify the functionality (intended use) of the machines as designed by the manufacturer. The risks generated by the AI, without the limits imposed by the AI act and in a complementary manner by the Machinery Regulation, would not be considered in the risk assessment carried out by the manufacturer during the design phase. Therefore, we will highlight the risks introduced by AI incorporated into machines and how these are treated and addressed in the new Machinery Regulation.

Keywords: artificial intellegence, machine learning, machinery products, machinery regulation

1. Introduction

On 14 June 2023, the European Parliament approved its first proposal on the Artificial Intelligence, knows as AI Act (European Commission, 2021). The general purpose of the AI Act is to ensure that AI systems placed on the EU market are safe and respect existing law on fundamental rights. Given the complexity of the AI law, in this work only the aspects concerning the coordination of the AI act with the Machinery Regulation will be taken into consideration. The aim of this work, in fact, is to show how the requirements concerning the safety aspects of machines with embedded AI applications, not covered in the current Machinery Directive 2006/42/EC (European Commission, 2006) are instead regulated in a complementary manner by the AI Act and the New Machinery Regulation (European Commission, 2023).

The AI act has been designed as a horizontal EU legislative instrument applicable to all AI systems placed on the market or used in the Union. The proposed regulation was designed to be complementary to cross-sectoral EU legislation, harmonized in the New Legislative Framework (NLF), complemented the New Approach in EU product safety adopted in 1985 (Veale, 2021), in order to ensure consistency and avoid any possible duplication with other ongoing revision of sectoral product legislation (e.g., the Machinery Directive, the General Product Safety Directive). At this end, it will be integrated into the updated Machinery Regulation as regards high-risk AI systems which are safety components of products. The approach used in the regulation on artificial intelligence is risk-based as it divides risks related to artificial intelligence into the following categories: unacceptable, high, limited and low or minimal risk (Chamberlain, 2023). Depending on the level of risk of the AI system, there are different requirements and obligations for the development, placing on the market and use

of AI systems in the EU. The AI Act explicitly prohibits harmful AI practices that are considered a threat to the safety and rights of people, as they generate an "unacceptable risk". It is, therefore, prohibited to place on the market, put into service or use in the EU AI systems that: use manipulative subliminal techniques; exploit physical or mental disability; used by public authorities for social scoring purposes; real-time remote biometric identification systems in spaces accessible to the public for law enforcement purposes.

The AI Act defines "high-risk" AI systems as those that have a negative impact on health and safety of people or their fundamental rights. High-risk AI systems are systems used as safety components of a product or which fall under EU harmonization legislation on health and safety and systems used in specific areas identified in Annex III (i.d biometric identification, education and professional training; law enforcement agencies...) (Ebers, 2021).

High-risk AI systems are subject to an obligation for an ex-ante conformity assessment: providers of high-risk AI systems would ensure that their systems undergo the conformity assessment before placing them on the market or put them into service (Madiega, 2021). In order to minimize the burden on operator and avoid any possible duplication, all AI systems regulated by existing product safety legislation will fall under existing third-party compliance frameworks already applicable (Floridi et al., 2022).

AI systems that present "limited risk", such as systems that interact with humans (e.g., chatbots), emotion recognition systems and AI systems that generate or manipulate images, audio or video content (e.g., example deepfakes), do not have to comply with the requirements and obligations established for high-risk AI systems, but are only subject to a series of transparency obligations.

All other AI systems that pose only "low or minimal risk" could be developed and used in the EU without complying with any legal obligations. However, the AI Act encourage the drawing up of codes of conduct to foster providers of non-high-risk AI systems to voluntarily apply the mandatory requirements for high-risk AI systems.

In this work, particular attention is paid to high-risk AI systems as they undermine health, safety and fundamental rights, especially those that are safety components of machines in order to highlight the complementarity of the AI and Machinery Regulation to cover all health and safety aspects.

1.1. AI Act consistency with Machinery Regulation

The evolution of the machinery sector has led to an increasing use of digital means and the software, that gives the machine a self-evolving behavior, plays an increasingly important role in the design of machinery. As a result, safety components shall be considered not only physical devices but also digital devices.

From these considerations arises the need to review the legislation that regulates the health and safety of products (such as the Machinery Directive 2006/42/EU) which presented gaps in the risk assessment deriving from AI applications integrated into products. Thus, the artificial intelligence act and other regulations under the NLF are being developed in parallel. The AI act, being a horizontal law, regulates the safety requirements of AI systems, both those used as a stand-alone system and those incorporated into products. As a result, where a high-risk AI system is a safety component of a product covered by a relevant NLF sectorial legislation, the manufacturer of the final product has the same obligations of the provider of AI system embedded in the final product, by ensuring that it complies with the requirements of AI act.

As regards, the risks related to the AI systems, that are components ensuring safety functions in machinery or are safety components in machinery with fully or partially self-evolving behavior using machine learning approaches, are addressed by the requirements of AI act, while Machinery Regulation will ensure the safe integration of the AI system into the overall machinery.

This approach is fully reflected in the interplay between AI Act and the Machinery Regulation that apply the same definition of AI system: software that is developed with one or more of the techniques and approaches listed in Annex I (machine learning, logic- and knowledge-based approaches and statistical ones) and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

This broad definition derives from the Organization for Economic Co-operation and Development's attempt to explain the logic of an AI system and which it has represented in a diagram shown in figure 1.

According to the OECD, AI for Europe is a machine-based system that can, for a given set of human-defined explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as makes predictions, content, recommendations, or decisions that can influencing physical real or virtual environments (OECD, 2019). Another point of coordination of the Machinery Regulation with the AI Act is the inclusion of machines that has embedded systems with fully or partially self-evolving behavior using machine learning approaches ensuring safety functions in the list of "high-risk machines" in Annex I.

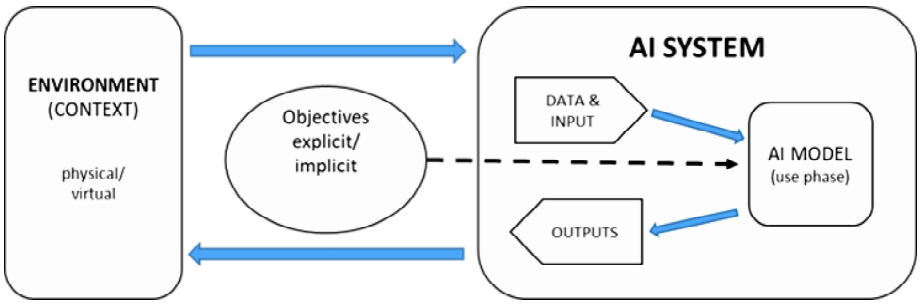

Fig. 1. OECD AI system model.

Furthermore, Article 9 of the Machinery Regulation concerning specific Union harmonization legislation establishes that "where, for a certain product within the scope of this Regulation, the risks addressed by the essential health and safety requirements set out in Annex III are wholly or partly covered by Union harmonisation legislation that is more specific than this Regulation, this Regulation shall not apply to that product to the extent that that specific Union legislation covers such risks". In order to reduce the risks from high-risk AI systems, providers of such systems must comply with the requirements set out in Chapter 2, but the AI Act assumes that often it will not be enough, and some risks will remain. In this sense, the role of Article 9 is to make sure that providers identify those risks and take additional measures to reduce them to an acceptable level (Schuett, 2023).

In the New Regulation the definition of "machine with evolutionary capacity" considers the evolving nature of AI applications incorporated into machines. Therefore, the relevant essential health and safety requirements are also adapted to evaluate the characteristics and capabilities of a machine with software and algorithms with evolving behavior over time, also paying attention to the risks that arise after they are placed on the market. So, the risks identified during the risk assessment should include those risks that might arise during the product's lifecycle due to an intended evolution of its behaviour to operate with varying levels of autonomy.

1.2. The main issues of machinery with embedded AI

More advanced machines that are less dependent on the human operator, recently introduced on the market, are different from previous machines work on defined tasks and in structured environments, so further improvements could lead machines to learn to perform new actions, different from those intended by the manufacturer, and become more autonomous even in unstructured environments. Current product safety legislation, including Directive 2006/42/EC, has a number of gaps regarding the safety and liability implications of AI. Therefore, the new Artificial Intelligence Act and the new Machinery Regulation should cover safety risks arising from new digital technologies and address new challenges in terms of product safety.

An ever-increasing amount of data and, consequently, a greater connection of machines to networks, both external and internal, can lead to risks linked to significant changes in the behavior of the machine itself (Anastasi et al., 2021). The Machinery Directive 2006/42/EC first, and now the new Regulation 2023/1230 of the European Parliament on machinery, introduce the obligation for the manufacturer to carry out risk assessment on machinery during the design and construction phase according to a structured and an iterative process. This structured risk assessment process on CE marked machines is described in the ISO 12100 standard, which also provides the aspects to be taken into consideration to identify all associated hazards and risks.

The design and construction of a machine can no longer be considered a "determined" process, since we know that new machines are equipped with increasingly sophisticated systems (including safety ones) and can have a behavior that evolves as new updates or new technologies take over. Artificial intelligence and the autonomous behaviour of machines can lead to new risks because machines do not always behave as they were initially designed but evolve their behaviour based on environmental conditions and autonomous learning. These behaviours, which may be different from those that the manufacturer considered in the risk assessment, involve the introduction of new hazards that were not taken into account in the hazard identification and therefore the related risks will not have been assessed. However, this scenario in which AI systems decide on their own is averted by the definition of artificial intelligence system according to which the outputs generated by the AI system are based on a set of human-defined objectives (Druetta, 2021).

2. The risk assessment of machines with embedded artificial intelligence

The machinery industry embraces new ways of designing and building machinery or related products that may present higher risk factors, regardless of their intended use or any reasonably foreseeable misuse, such as systems with self-evolving behavior that provide safety functions. Indeed, systems with completely or partially self-evolving behavior using machine learning approaches differ from traditional software, which is incapable of learning or evolving and only programmed to perform certain automated machine functions.

The new EU regulation on machine products 2023/1230, developed to fill the gaps in Directive 42/2006 regarding the introduction of new technologies and autonomous and evolutionary behavior of machines, introduces the following innovative general design principle: "the risk assessment and risk reduction shall include hazards that might arise during the lifecycle of the machinery or related product that are foreseeable at the time of placing the machinery or related product on the market as an intended evolution of its fully or partially self-evolving behaviour or logic as a result of the machinery or related product designed to operate with varying levels of autonomy".

Some embedded AI applications can have implications on the machine function and thus on machinery safety, so these AI systems introduce new hazards or increased risks that are not addressed by the current risk reduction measures. The risk assessment process, conducted in accordance with the ISO 12100 standard (ISO, 2010), in light of the emerging risks related to the integration of AI applications into machines, must take into account these new aspects. The ISO 12100 standard guides the machine manufacturer in the safe design phase. Step by step, for each identified hazard, the manufacturer estimates the associated risk and based on the acceptability criteria, carries out the risk reduction following the three-steps method by favoring reduction measures integrated into the design. As mentioned above, AI incorporated into machines introduces new risks, compared to traditional software installed on machines, linked to self-learning and the evolutionary behavior of machines.

New risks would arise from the new actions performed by the machine that would have not been evaluated a priori (Monica et al., 2020). New research challenge could lead to machines equipped with artificial intelligence capable of gradually assessing themselves emerging risks and acting appropriately measures to reduce them; in this case we could think about "dynamic self-assessment risk". This hypothesized development of research, bringing with it an innovative principle, would lead to opportunities but also to potential negative effects.

New risks would arise from the new actions performed by the machine that would have not been evaluated a priori. New research challenge could lead to machines equipped with artificial intelligence capable of gradually assessing themselves emerging risks and acting appropriately measures to reduce them; in this case we could think about "dynamic self-assessment risk". This hypothesized development of research, bringing with it an innovative principle, would lead to opportunities but also to potential negative effects.

In this work, we want to show the importance of the first step of the risk assessment process on machines during the design phase according to the ISO 12100 standard. This standard currently has gaps related to the evaluation of risk factors linked to the new enabling technologies of industry 4.0, as well as those relating to human-robot interaction and AI applications in machines (Murino et al., 2023).

As shown in figure 2, the aspects, to be considered in the phase of defining limits of machinery with embedded AI applications, must be expanded by also integrating the requirements and definitions reported in the AI Act. The definition of the limits of machinery, in fact, is necessary to identify the hazards associated with that operation mode and the space in which the machine operates. If the machinery's behavior allows itself to perform other actions, different from those defined in the design by the manufacturer, the risk assessment would no longer be reliable. Starting from the diagram of the interactive process of risk assessment (Figure 2), however, it can be deduced that the autonomous and evolutionary behavior of the machine, due to the embedded AI, must respect its defined task (intended use) and its movement space defined in the first step of the above process. According to AI Act, the "intended purpose" means the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation. While the "reasonably foreseeable misuse" means the use of an AI system in a way that is not in accordance with its intended purpose, but which may result from reasonably foreseeable human behavior or interaction with other systems.

Control systems of machines or related products with fully or partially self-evolving behaviour or logic designed to operate with different levels of autonomy shall be designed and constructed in such a way as not to cause the machine or related product to perform actions beyond its defined task and movement space. In essence, the AI system incorporated into the machine must respect the limits for which it was designed, i.e., it must not make decisions that compromise the safety of the machine. Furthermore, since the learning of the AI system continues even after putting into service, it is necessary to ensure that further learning does not compromise the

above. In the updated Machinery Regulations, it is established that "the limits of the safety functions are to be established as part of the risk assessment performed by the manufacturer and no modifications are allowed to the settings or rules generated by the machinery or related product or by operators, including during the machinery or related product learning phase, where such modifications could lead to hazardous situations (EHSR 1.2.2)".

The definition of an artificial intelligence system according to the Artificial Intelligence Act highlights that it is software based on a model which, for "human-defined objective", achieves the intended purpose. This implies that the need for human oversight to ensure that an AI system performs in accordance with its intended purpose under certain conditions. The AI Act requires that data used by AI systems be managed to ensure the highest quality, reducing the risk of bias and ensuring that decisions made are accurate and reliable. The data used to train an AI system must be such as to confirm the expected performance of that system before its placing on the market or putting into service.

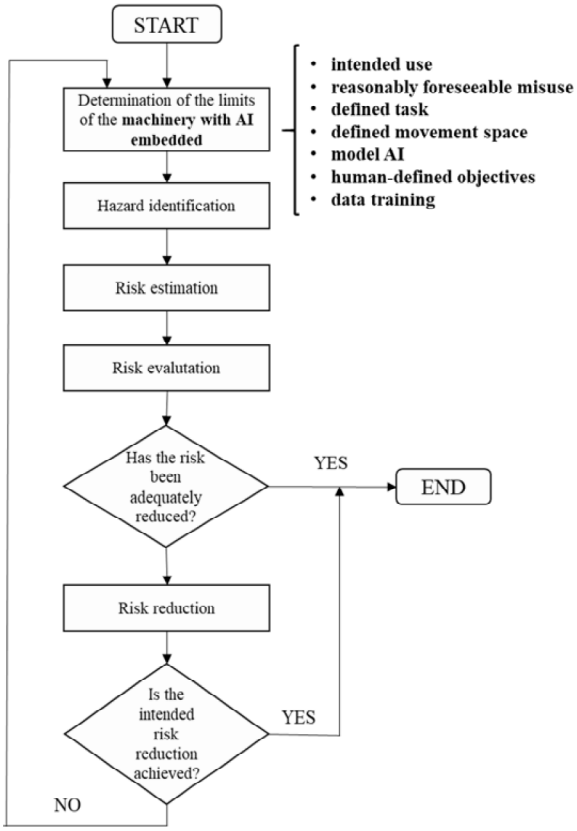

Fig. 2. The risk assessment of machines with embedded AI - adapted from ISO 12100.

In this work, we want to highlight the new risk factors linked to the embedded AI on machines and how to take them into account in the machinery design. In this paragraph, we will list and arrange them in a table to make them easier to understand (Table 1). the measures to be adopted in the design of machines with incorporated AI are deduced both from the essential safety requirements introduced in the New Machinery Regulation and from what is prescribed in the AI Act. The main risk factor related to AI system embedded into machines it is related to its fully or partially self-evolving behaviour. In order to prevent the performance of machinery equipped with artificial intelligence from going beyond its intended use, it is necessary to design an intended evolution of its behavior with actions limited to the defined task and movement space. To avoid health and safety risks, human oversight shall be able to monitor AI system operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible and able to interrupt the system through a stop button or a similar procedure.

The human-machine interface with AI embedded should be appropriately designed according to a human centered approach to guarantee the ergonomics of the interaction (ISO, 2019).

In order to assure ergonomic principles, the human-machine interface must be adapted taking into account that the machine may have a behavior or logic intended to evolve completely or partially and designed to function with different levels of autonomy (ISO/TR, 2020). Furthermore, such machinery with completely or partially self-evolving behavior or logic must be designed to respond to people in an adequate and appropriate manner (e.g., verbally through words and non-verbally through gestures, facial expressions or body movements) and to communicate to operators the planned actions in an understandable way (such as what he will do and why).

Table 1. Link between AI risk factors and measures to be adopted in the design of machines.

AI risk factors	**Measures to be adopted in the design of machines with embedded AI**
• machine with fully or partially self-evolving behaviour	- an intended evolution of its behavior with actions limited to its defined task and movement space - human oversight
• the human-machine interface with AI embedded	- human centered design - designed to respond to people adequately and appropriately - designed to communicate its planned actions (such as what it is going to do and why) to operators in a comprehensible manner - sensors to avoid collisions between humans and autonomous mobile machines
• data on which the decision-making process is based	- data that are critical for the compliance of the machinery or related product with the relevant essential health and safety requirements shall be identified as such and shall be adequately protected against accidental or intentional corruption - the parameters of the machinery or related product shall not change in an uncontrolled way - modifications to the settings or rules, generated by the machinery or related product or by operators, including during the machinery or related product learning phase, shall be prevented, where such modifications could lead to hazardous situations - measures to prevent from cybersecurity attacks by trying to manipulate the training dataset, inputs designed to cause the model to make a mistake

Furthermore, when there is also physical interaction between man and autonomous mobile machine, one of the major related challenges is managing the possible risk of collision. A sensor system must continuously monitor the risk of unwanted interactions and collisions and must have priority over the execution of tasks.

An important aspect to consider when machines are designed to incorporate software is related to the protection of data critical to the machine's compliance with the relevant essential health and safety requirements, against accidental or intentional corruption. So, modifications to the settings (machine parameters) or rules, generated by the machinery or related product or by operators, including during the machinery or related product learning phase, shall be prevented, where such modifications could lead to hazardous situations.

Machines with evolving behavior, due to their characteristics such as data dependency and connectivity, should be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of accuracy, robustness and cybersecurity, and perform consistently in those respects throughout their lifecycle. In fact, the technological development that has led to equipping machines with increasingly sophisticated sensors and technologies has led to new evaluations of safety aspects also linked to the risks introduced by malicious third parties which affect the safety of machinery products. Hence, the need to protect machinery from cybersecurity threats arises, therefore machinery manufacturers should be required to adopt proportionate cybersecurity measures to protect the overall security of the machinery product.

Cybersecurity plays a crucial role in ensuring that AI systems are resilient against attempts to alter their use, behavior, performance or compromise their security properties by malicious third parties who exploit system vulnerabilities. Cyberattacks against AI systems target the AI training dataset (e.g., data poisoning) or AI training models (e.g., adversarial examples), exploiting vulnerabilities in AI digital assets. artificial intelligence. To ensure a level of security appropriate to cybersecurity risks, developers of AI systems should therefore take appropriate measures, also taking into account the underlying ICT infrastructure. Technical solutions to address specific vulnerabilities in AI systems include measures to prevent and control external attacks, both those that attempt to manipulate the training data set (data poisoning) and those that trick the model into committing a mistake (adversarial examples). Furthermore, when there is also physical interaction between man and autonomous mobile machine, one of the major related challenges is managing the possible risk of collision. A sensor system must continuously monitor the risk of unwanted interactions and collisions and must have priority over the execution of tasks.

3. Conclusion

The main challenges in AI arise when machines gain the ability to learn and become autonomous in decision making. One of the main risks arises when, given their algorithmic decision making, AI could make ethical decisions about their actions and interactions with humans (Iphofen, 2021).

AI applications can bring great opportunities but also great risks, with the AI regulation the European Union has taken responsibility to ensure the safe development of AI and avoid distorted uses. The EU, through the AI Act, intended to define governance mechanisms that pose ethical barriers to AI, so that we can achieve human-centered and human-controlled AI design.

In this article, we have highlighted the interplay between the AI Act and the New Machinery Regulation, as AI is a horizontal regulation that cascades into product safety regulations. The analysis conducted has highlighted the potential impact of the requirements that AI systems must meet to comply with the AI Act on Machinery Regulation. The highlighted interactions, consequently, must also be taken into account in standardization relating to machinery to offer manufacturers useful references for the design and construction phase of the product.

However, only when both European regulations come into force it will be possible to highlight their strengths and weaknesses, developments that will be followed by the authors.

References

Anastasi, S., Madonna, M. & Monica, L. 2021. Implications of embedded artificial intelligence-machine learning on safety of machinery. Procedia Computer Science, 180, 338-343.

Chamberlain, J. (2023). The risk-based approach of the European Union's proposed artificial intelligence regulation: Some comments from a tort law perspective. European Journal of Risk Regulation, 14(1), 1-13.

Druetta, C. 2021. AI in the Making: Is EU Artificial Intelligence Regulatory Framework a Good Move for European Manufacturing?. Int'l. In-House Counsel J., 14, 7367.

Ebers, M. 2021. Standardizing AI-The Case of the European Commission's Proposal for an Artificial Intelligence Act. The Cambridge handbook of artificial intelligence: global perspectives on law and ethics.

European Commission. 2006. Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC.

European Commission. 2021. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts.

European Commission. 2023. Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC.
Floridi, L., Holweg, M., Taddeo, M., Amaya Silva, J., Mökander, J. & Wen, Y. 2022. CapAI-A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. Available at SSRN 4064091.
Iphofen, Ron, and Mihalis Kritikos. 2021. Regulating artificial intelligence and robotics: ethics by design in a digital society.Contemporary Social Science 16.2: 170-184.
ISO 9241-210: 2019. Ergonomics of human-system interaction Human-centred design for interactive system.
ISO/TR 9241-810: 2020. Ergonomics of human-system interaction Robotic, intelligent and autonomous systems.
ISO 12100:2010. Safety of machinery – General principles for design – Risk assessment and risk reduction.
Madiega, T. 2021. Artificial intelligence act. European Parliament: European Parliamentary Research Service.
Monica, L., Madonna, M. & Anastasi, S. 2020. Risk assessment on machines with CE marking and with embedded industry 4.0 enabling technologies. WIT Transactions on Engineering Sciences, 129, 55-65.
Murino, T., Nardo, M.D., Pollastro, D., Berx, N., Francia, A.D., Decré, W. & Pintelon, L. 2023. Exploring a cobot risk assessment approach combining FMEA and PRAT. Quality and Reliability Engineering International, 39(3), 706-731.
OECD. 2019. Recommendation of the Council on Artificial Intelligence,.
Schuett, J. 2023. Risk management in the artificial intelligence act. European Journal of Risk Regulation, pp. 1-19.
Veale, M. & Zuiderveen B.F. 2021. Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, vol. 22, no. 4, pp. 97-112.