AI RISK MANAGEMENT IMPLEMENTATION CHALLENGES

Ewelina Szczekocka

Orange Polska S.A.

ABSTRACT

The article presents a state-of-the-artreview on AI Risk Management, a first result in aresearch company project. It highlights crucial questions on practical implementations anddepicts major challenges for organizations, finally proposingfurther directions in solving these challenges (ongoing work).

KEYWORDS

AI Risk Management, organizational challenge, standardisation.

1. INTRODUCTION

AI applications have developed rapidly and massively in recent years. Gartner [1] predicts that during the next six years, the AI software market will grow at a 19.1% annual growth rate (Generative AI from 8% in 2023 to 35% in 2027) and reach $297 billion in 2027, while in 2022, it was $124 billion. AI supports a competitive advantage, andcompanies, countries, and regions are racing to develop AI systems (US, China).

It is indispensable to manage risks arising from their undesired operation. Risk management has long been a well-known approach for various fields of human activity, in particular IT systems. Risk management is one of the fundamental concepts of project management [2]. It is developed and evaluated in many scientific, standardization, and industrial organizations by defining specific frameworks, which should be defined at a highlevel and adapted to different vertical organizations' particular needs and objectives(like industries and services).Referring to the Gartner Report[1] 2023 and [3], the AI Trust, Risk and Security Framework Management Framework (AI TRiSM) is a fundamental framework for organizations to deliver Responsible AI that will reach mainstream adoption within two to five years.

This article aims to clarify the specificity of AI risks, present state-of-the-art AI risk management, and consider a practical approach to AI risk management.There are different perspectives on risks, for instance, business (e.g., company reputation, penalty linked to regulation), technical (functional and non-functional, e.g., related to security), and societal (related to fundamental values and their protection). AI technologyincreases existing risks and brings new ones concerned with new technologies, algorithms, societal and regulatory approaches [4], [5]. It can completely change the risk perspective, impact regulation, standardization, economies, and even governments (an example of Generative AI blow in 2022 with ChatGPT). Europe will soon have a specific law regulating the development and use of AI products and services. A considerable challenge for organizations established in Europe arisesrelated to implementing and operationalizing

[1]https://www.gartner.com/en/newsroom/press-releases/2023-10-11-gartner-says-more-than-80-percent-of-enterprises-will-have-used-generative-ai-apis-or-deployed-generative-ai-enabled-applications-by-2026

David C. Wyld et al. (Eds): CCSIT, NLPCL, AISC, ITE, NCWMC, DaKM, BIGML, SIPP, SOEN, PDCTA – 2024
pp. 95-110, 2024. - CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.141710

this set of regulations.It is primarily dueto the gap between a high-level description of regulations and the need for detailed requirements to implement risk managementeffectively. AI regulations will be followed byharmonized standards that will help transformthem into practice. In particular, standards can play a crucial part in clarifying a risk subject.It is challenging, especially when it is expected to cover the protection of fundamental rights, and it raises questions on democratic processes in standardization organizations[2].

The new landscape of European regulations raises awareness that managing AI risks is more than merely a technical challenge. It requires addressing complex regulatory, ethical, and legal concerns. The AI Act and other regulations, like the GDPR and sector-specific laws, require organizations to ensure a necessary balance of innovation (based primarily on new algorithms, models, applications and complex value chains, including a wide range of stakeholders, technologies,and processes) with accountability and transparency. Ensuring compliance with regulations while protecting fundamental rights, such as privacy and non-discrimination, becomes critical as AI systems can be broadly applied across sectors. Moreover, the dynamic and evolving nature of AI technologies, including Generative AI, presents organizations with the difficulty of staying ahead of regulatory developments. Harmonized standards will be vital in bridging the gap between legal frameworks and practical implementation. Still, organizations must proactively interpret and apply these standards to their specific contexts. Lastly, embedding ethical considerations into AI risk management frameworks ensures that societal impacts and human values are consistently addressed, preventing harmful outcomes and fostering public trust in AI systems. On the other side, they must be managed appropriately. These are broad challenges in AI risk management, where the interplay between regulations, standards, ethical rules (including fundamental rights), and practical implementations are particularly demanding for organizations,arousing the need for a proactive and adaptable approach.

2. RISKS AND AI RISKS

Referencing the NASA handbook [6], the risk is characterized by three essential components: the scenario that leads to risk materialization, the likelihood of its occurrence, and the consequences. Risks also have a mutual correlation with quality [7]. Quality of products and services leads to decreasing risks, whileeffectiverisk management through mitigating risks leads to increasing quality. Quality is a final objective for organizations concerning product or service delivery, while risk is the effect of uncertainty on that objective [8]. It is also important to include safety in this complex picture. Safety for people using products and services is a fundamental demand [9]. It is also a subject of several regulations.

Risk management developed in the middle of the 20th century with the fast development of IT systems (including business platforms and CRMs). Risk perception variesdepending on different aspects (cultural, geographical, political, educational). It can be differently perceived,e.g., depending on age and gender. There are also differences in the country's viewing risk, e.g., according to the survey, the Japanese have concerns about climate change, Chinese - geopolitical risks, and French - pollution. Europeans cope differently with geopolitical tensions (the UK has trust in international organizations at 55% while the French at 44%).Several standardization organizations (SDOs) worked on defining and describing risks and guiding risk management, including an emerging issue of AI risks. Among them, there are especially international ISO and European CEN-CENELEC organizations offering, e.g., ISO/IEC 31000:2018 [10] for general risk approach, ISO Guide 73:2009 [11] with guidelines and vocabulary for risk management, and also NIST with its Risk Management Framework [12]. Other European organizations, ENISA and ETSI, provide guidance for security and cybersecurity approaches supporting minimizing cyber-

[2]https://policyreview.info/articles/analysis/regulating-ai-through-technical-standards

security risks, such as ENISA [13] and ETSI [14]. SDOs describe risk in different ways and perspectives, which shows that there is no single understanding of it. For instance, in cybersecurity, „risk" is always perceived as having a negative impact being related to threats and vulnerabilities to losses due to a cyber-attack or data breach (see, for instance, CRA[3]). On the contrary, ISO perceives risks as negative and positive (working on both business and technical aspects of risks). It reflects that in business, risk impact, for example, concerned with the projects and their outcomes, may be positive or negative. It may lead to deeper implications, like discrepancies in understanding among various legislations.According to the European Cyber Resilience Act (CRA) [15], risk refers to a product before an attack, while another concept, that is impact, refers to consequences after the attack. Risk is related to the product definition and properties, describing potential ways of attack and known vulnerabilities. It can change due to the appearance of new methods of attack, ecosystem evolution, or discovery of unknown vulnerabilities, but it can also evolve with technological development.There is a clear link among risks, threats, and harms described by the SDOs like ISO, ENISA, or NIST [16], [17]. For instance, several known digital risks were described and managed by ISO 27005 [18] and NISTstandards dedicated to information and cybersecurity risks, e.g., [19].Concerning AI, risk can be connected, for instance, with the use of an AI system out of the intended domain, e.g., driving an autonomous car out of its operational design domain (qualifying as a threat) with the corresponding risk of car damage and passengers or pedestrians' harms.AI strengthens some risks and their impacts (e.g., much higher risk in data and its quality related to unintentional or intentional errors in data) and brings some new risks, in particular, for data (like bias or data drift) and models (like some risks related to autonomous systems and autonomous cars, AI-supported decision-making automation and human control aspects, lack of system specification, uncertainty in output).There is a specific relation of AI risk to AI lifecycle and Responsible AI principles. Failure to consider that may cause negative consequences of risk increase. Linking AI risks to responsible AI principles and the AI lifecycle is then critical. It presents a good way to facilitate risk management considering AI Act requirements and helps properlyidentify risks.Concerning the AI lifecycle, it isimportant to consider data processing steps across the lifecycle stages. From the cybersecurity risks perspective (AI Act, Article 15, Recital 66), it is also crucial to link threats to it. ENISA, in its report on the AI cybersecurity threats landscape [20], associated data processing steps and threats to the AI lifecycle stages. Risks can be associated with different responsible AI principles[4](examples: 1.COMPAS tool supporting legal decisions in the US - lack of model transparency, and appearing risks toward fairness concerned with bias in data leading to discriminacy, 2. ChatGPT, providing extensive information from different sources possibly linked –a lack of mechanisms to check for privacy protection causing significant risks towards privacy, 3. insufficient quality of data and process of data handling - impacting robustness of the systems).Missing application of Responsible AI principles in AI systems is a source of significant risks, as examplesshow (see Table 1). Managing AI risks and following the Responsible AI principles helps decreasethe likelihood of risk occurrence.Table 1 (drawn during our research analysis) provides some examples of correlations between risks and responsible AI principles.

Table 1. Responsible AI Principles and Risks.

Responsible AI Principle	**Risk Example**

[3] https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act

[4]https://artificialintelligenceact.eu/recital/27/ (on ethical principles for AI) and https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Ethics guidelines for Trustworthy AI)

Transparency	Lack of model transparency trap: ex. Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in US – lack of judicial transparency due to lack of algorithmic transparency (and biased algorithm). Excessive openness: ex. ChatGPT- prompting computer code, policy briefs compromising company knowledge. Important: relevant transparency level (type and purpose of the system).
Explainability	Black box effect of credit scoring application decision of rejecting loan
Accountability	Ex. for healthcare (ScienceDirect, 2022): Accountability as a process in which healthcare practitioners have the potential responsibility to justify their "clinical actions" to patients and are held liable for the consequences. With an AI-based decision support system, clinicians are held accountable if they decide to follow AI, even resulting in patient harm. Clinicians are also held responsible if they deviate from the standard protocols. Facing lack of accountability (and trustworthiness) clinicians will only follow AI if it matches their judgment and aligns with the standard protocol - potentially making the AI underused.
Equitability/Fairness/ Bias	1) COMPAS biased algorithm: predicting twice as many false positives for recidivism in black offenders than white ones; 2) Facebook's Advertisement Algorithm (2019): allowed advertisers to target people based on their race, gender, and religion; 3) Twitter Image Cropping (Sept 2020) - the image cropping algorithm favoured white faces over black.
Robustness	1) The data quality and handling process impacts the robustness, e.g., data augmentation may cause risks; 2) adversarial modification of pixels in the image, to fool the identification and classification of the image by the AI (ex. of adversarial attack).
Security	Intentional breaches or accidental leaks of information caused by not proper data handling in AI systems (this also falls to privacy risk); AI systems to be hacked or manipulated (they become still more complex and autonomous, causing the increase of risk of cyber-attacks). These attacks may allow malicious actors to take control of the AI system, causing it to make harmful decisions for individuals, groups or whole society.
Privacy	ChatGPT – not enough protection of sensitive information; social credit system already existing in China; federated learning models attacks may cause the models to fail and also infer private information (empirical research showed it).

It is also crucial to understand an AI risk cartography. It is a matter of the comprehensive understanding and mapping of risks for AI systems. In the scope of our research, a review of theAI risk specifics and risk cartography approach was provided based on differentSDO works (a brief summary presented in Table 2).

Table 2. AI risks and risk cartography

Risk Characteristics	**ETSI**	**ENISA**	**CEN-CENELEC**	**ISO**	**NIST**
Link to Responsible AI	Not directly linked	Yes (AI principles included to AI cybersecurity properties)	Yes (adopted from ISO and new standards to come)	Yes	Yes
Risk definition	Yes	No specific definitions added	Adopted from ISO fundamentals: concepts/ terminology (EN 22989), framework for AI (EN	Concepts/ terminology (ISO/IEC 22989), framework for AI (ISO/IEC 223053)	Like ISO definition

			223053)		
Framework of risks	Not available	Interoperable RM Framework	Works ongoing with direct link to AI Act	Risk Management Framework ISO 31000, AI aspects in ISO 23894	AI Risk Management Framework (AI RMF): AI RMF core function, AI RMF profiles
Risk management (RM)	Risk level assessment	Interoperable RM framework	Works initiated, considered development of original approach	Risk management process. (risk: assessment, treatment monitoring and review, recording and reporting)	Management of AI risks in context of Responsible AI ; Risk measurement, Prioritization, Tolerance
Risk cartography	Several related approaches to classification	Risk classification, AI assets and threat taxonomy	Check list and catalogue of AI risks (family of risks);	Classification of risks not proposed; Bias categorization	Characteristics of trustworthy systems
Risk level assessment	Defined as the likelihood of an attack and the impact of the attack on the system	Yes	Ongoing works	Not available	Yes
RM guidance document	White papers	Toolbox and reports	Not available	Guidance on AI Risk Management	AI RMF Playbook
AI Lifecycle	Not considered	Defined and Risk linked with it	Adopted from ISO	Described but not clearly linked with AI lifecycle	AI RM processes, procedures, and functions can be applied at specific stages of the AI lifecycle
AI risks related problems	Security of AI support	Support for AI cybersecurity; Identification of malicious use of AI	Trustworthiness considered	Trustworthiness, Transparency, ethics (bias, fairness), robustness	Transparency, transparency ontology (draft)
Impact definition	Not provided	Not provided	Impact is taken from ISO (by adopting ISO standard for risk management)	Impact is defined as „consequence”	Impact is ISO-based (understanding) but naming „impact” not „consequence”
Impact assessment	Calculation methods for occurrence likelihood	Calculation of impact	Not considered	AI system impact assessment process (initial step)	Relation of impact and risk (in AI RMF)
Impact cartography	Not available	Not available	Not available	Not available	Yes (e.g., actors across AI dimen-

					sions)
Conform-ance	Test suites	Requirements on EU certifica-tion	TR AI Con-formity As-sessment under consideration	Requirements for bodies providing audit and certifica-tion of artificial intelligence (ongoing)	Not considered (there are docu-ments on con-formance but not for AI)

This research, which involved a comprehensive analysis of different standardization documents and reports in an AI risk management context, provides a solid foundation for companies to design and build effective AI risk management systems. A brief summary of this in-depth study, focusing on a few major selected organizations, is presented in Figure 1.Several frameworks proposed by vertical organizations for AI risk management were also reviewed, as they can be a practical starting point for organizational AI risk management systems.This review also helps to see a feasibility of implementations of AI risk management systems.

CEN-CENELEC (JTC21)	ISO (SC42)	ESTI	ENISA	NIST
• **EN ISO/IEC 22989:2023** • **EN ISO/IEC 23894:2024** (harmonized EU) • **EN ISO/IEC 42001** (Preliminary) (harmonized EU) • **prCEN/CLC AI Risk Management** (Drafting) • **pr EN AI trustworthiness framework** (Drafting) • **prCEN/CLC TR AI Risks - Check List for AI Risks Management** (Preliminary) • **prCEN/CLC TR Impact assessment in the context of the EU Fundamental Rights** (Preliminary) • **FprCEN/CLC/ TR 17894 AI Conformity assessment** • EN 23053:2023 Framework for AI Systems Using ML (harmonized EU)	• **ISO/IEC 22989:2022 AI Concepts and terminology** • **ISO/IEC 23894:2023 AI, Guidance on risk management** • **ISO/IEC 42001:2023 AI Management system** • **ISO/IEC 42005 AI system impact assessment** (Draft International Standard) • ISO/IEC 42006 Requirements for bodies providing AI audit and certification (Draft International Standard) • ISO/IEC 23053 Framework for AI Systems Using ML • **ISO/IEC 31000:2018 Risk Management – Guidelines** • **ISO Guide 51:2014 Safety aspects**	• **GR CIM 007 Security and Privacy** • **GS SAI 003 Securing AI (SAI), Security Testing of AI** • **GR SAI 001 SAI, AI Threat Ontology** • ETSI ISG SAI GR-004 SAI: Problem Statement • ETSI TS 102 165-1 Method and pro forma for Threat, Vulnerability, Risk Analysis (TVRA) • Draft ETSI GR SAI 009 ETSI GR SAI 009 V1.1.1 (2023-02) (SAI), AI Computing Platform Security Framework	• **Interoperable EU Risk Management Framework** • **Interoperable EU Risk Management Toolbox** • **AI Threat Landscape** • Cybersecurity of AI and Standardization	• **NIST AI RMF, AI Risk Management Framework,** • Draft Taxonomy of AI Risk • NIST AI Risk Management Framework Playbook

Figure 1. Standards and reports supporting AI risk management (extract)

The result of this review provided for vertical AI risk management frameworks is summed up in Figure 2.

Risk management framework	Useful terminology	RMF objectives	Measurements	Comments (details in separate report)
Green Climate Fund RMF (GCF)	Risk appetite	maintain the residual risk level within risk appetite and tolerance	Probability of occurrence Impact Key Risk Indicators	dedicated for the organizational aspects of granting projects proposals by the funding organization
Asian Infrastructure Investment Bank (AIIB)	Risk culture Risk appetite	Business continuity	Key Risk Indicators	dedicated for protecting business continuity and investment capital of the AIIB; concept of risk management is based on risk culture, which forms the base for risk definition
Department of Education of Queensland Government (DoE) Enterprise RMF	Risk appetite	reduce exposure to risk	Risk estimation	supports managing risk in schools, regions and in divisions and supports a consistent approach to identifying, analyzing, evaluating and treating risk
Risk Management Software	Decision tree	Customized tool	Not defined	provides means for the risk analysis in complex way with consideration of all relevant aspects
Trial Master File (TMF) Management	none	Effort prioritization	Not defined	leads to a more efficient and consistent flow of documents and data throughout the clinical process
Risk Management File (applicable for manufacturing of medical devices)	none	Documentation management Support for audits	Not defined	management of risk management related documents requires implementation of detailed process to enable tracking and updating of any creation or modification of information related to the manufacturing and delivery process of medical device
EUDG Transport of Dangerous Goods (TGD) RMF	none	Accessibility by many users	Not defined	aims to be a comprehensive instrument that can help to build robust decisions for managing residual TDG risks and to share harmonized results between interested parties

Figure 2. AI in verticals: risk management frameworks.

3. AI REGULATORY CHALLENGES IN EUROPE

3.1. AI European Regulatory Context

AI regulation in Europe is coming in the form of the AI Act[5] based on risk levels, where some systems will have high-risk level assigned by default depending on their type of application.Several obligations will be introduced for high-risk systems. Non-compliance with the AI Act will cause heavy consequences (financial and legal). Providers of AI systems established in the EU will have an obligation to comply with these new European regulations, including those outside Europe providing their businesses in Europe. The AI Act mentions a wide range of risks, like accident, misuse, or structural risks [21], making the risk management approach even more difficult, as the AI Act does not provide additional guidance on how to deal with those different types of risks. It is the organization's responsibility to identify and manage risks in an appropriate way.The AI Act is Europe's successive, most revolutionary regulation of digital technologies. It develops in the complex regulatory environment proposed for the Digital Market in Europe, with several data, cybersecurity, and other regulations, examples of which are provided in Table 3.

Table 3. Examples of EU regulations for Digital Market.

Act	Description	Status
	1 Data Protection and Privacy	
EU 2016/679	GDPR, General Data Protection Rules	In force (2018)
EU 2022/868	The Data Governance Act	In force (2022)
EU 2023/2854	The Data Act	In force (2024)
	2Cybersecurity	
EU 2022/454	CRA, Cyber Resilience Act	Ongoing (2024), mandatory (2027)
EU 2014/53	RED on Cybersecurity and Privacy; delegated regulation 2022/30	In force (2022), man-datory (2025)

[5] https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf

EU 2016/1148	NIS2 Directive	In force (2023)
EU 2019/881	EU Cybersecurity Act	In force (2019), amended (2023)
	3Other	
EU 2022/1925	DMA, Digital Market Act	In force (2022), applicable (2023)
EU 2022/2065	DSA, Digital Services Act	In force (2024)
EU 2023/988	GPSR, the General Product Safety Regulation	In force (2023), applicable (2024)
EU 2023/1230	EU Machinery Regulation that replaces EU Machinery Directive (2006/42/EC)	In force (2023)

AI Act will be followed by harmonized standards elaborated for Europe[6].There is an ongoing work of European SDOs, coordinated by CEN-CENELEC, playing a pivotal role in this process and cooperating with other international SDOs like ISO. The conformity assessment procedure proposed within these standards will lead to the European Conformity (CE) marking of AI products, showing adherence to EU regulations. As seen above, a degree of complexity of the new regulation and legislation relatingto AI may profoundly impact AI stakeholders.This complex landscape requires a comprehensive understanding of mutual relations among regulation acts. Dependencies among regulations and their understanding are still under discussion. [22] provided an overview of the policy actionscorresponding to the needs of the upcoming Artificial Intelligence (AI) Act and Cyber Resilience Act (CRA). The AI Act emphasizes risks related to the area of use, considering its impact on the vendor ‘s market position.In contrast, the CRA, according to product category and intended use, considers the severity of the incident. CRA alone considers the financial impact of adopting the regulation by market players.European Cyber Resilience Act will introduce cybersecurity obligations and essential requirements for market players, for which assessment of riskswill be the first step in verifying compliance and part of a tool for manufacturers to comply with those essential requirements. It will apply to products with digital elements (with several exclusions) [23].Its scope is broad and not precise enough for the current stage, and it has to define product categoriesbetter and clarify them.In general, there may be an issue of compatibility between various regulations, potential overlaps, or contradictions. Dependencies and precedence for particular sectoral and specific regulationcategories should be clarified.

3.2. AI Act Implementation Level Issues

The AI Act defines essential requirements and delegates precise implementation decisions to European Standardisation Organisations.The ambiguity of these requirements raises many open questions for stakeholders involved in the AI Lifecycle on how to operationalize managing AI risks and achieve fundamental rights protections [24]. There are several policy areas difficult to quantify and operationalize. The risk categories defined in the AI Act apply to broad fields of AI application, which may cause the risk magnitude to be wrongly estimated. Consequently, the AI Act may not be enforced effectively, primarily when regulating general-purpose AI (GPAI) with its versatile and unpredictable applications[25]. The proposed solution applies the risk categories to specific AI scenarios rather than solely to fields of application, enabling the estimation of the magnitude of AI risk by considering the interaction between risk determinants, individual drivers of determinants, and multiple risk types. Two major artefacts of the proposed AI Risk assessment model are enhancement of the AI Act proposal with more effective risk management measures, and scenario-based risk assessment introducing granularity related to scenarios. It should be treated as an iterative approach. There is still a gap in ensuring a coverage of all representative

[6] https://artificialintelligenceact.eu/standard-setting/

practical cases. The AI Act also posed highly political questions to standards development, which was not typical for standardization.In particular, standardization bodies do not have the legitimacy and expertise to make decisions on interpreting human rights law and other policy goals[7]. Specific mechanisms should be established to enable democratic control of human rights protection. The AI Act is intended to protect humans against safety risks and malevolent results on human fundamental rights. Harmonized standards and CE marking could apply to the protection of fundamental rights. This extension of the product safety approach to fundamental rights protection is new. It raises numerous challenges that should be addressed.

Human (fundamental) rights impact assessment is crucial but also challenging for AI systems, as AI introduces several risks to this area.GDPR is the foundation for a privacy impact assessment provided under localauthorities' guidelines and control. AI and fundamental rights introduce new risks for privacy. The relevant impact assessment should evolve to consider new requirements concerned with the protection against these risks.Charter of Fundamental Rights (CHFR)[8], developed before the European Convention of Human Rights (ECHR), expresses the concept of 'human rights' within a specific European Union (EU) context, called further fundamental rights. The European Commission places health, safety, and fundamental rights at the centre of the AI Act.

CHFR (extending ECHR) addresses some modern issues that are not occurring in the ECHR (such as human cloning and data protection)[9].It contains articles related to various fundamental human aspects, including human dignity and freedom, cultural, economic, social, ethical (incl. equality and fairness), and legal. It also concerns the right to privacy, such as personal data protection related to GDPR,which is interrelated with the AI Act, both grounded in TFEU (Treaty on the Functioning of EU). GDPR will continue its role for AI systems (Article 22, and beyond, stating that personal data processing may create significant risks to fundamental rights and freedoms (e.g., may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data).Thanks to GDPR rules, privacy impact assessment is already in force, supporting the achievement of transparency and accountability of AI systems. This trend of identifying whether systems fulfil GDPR will be continued.

Moreover, AI technology may cause more elements to be checked [26]. Privacy impact assessment would be mandatory for high-risk AI systems.There are important questions about how privacy impact assessment should evolve to consider AI risks and how to include fundamental rights in the AI risk management approach.Another important aspect is that AI systems will act crossing EU borders, and then a global perspective should be applied to human rights in the light of AI applications, not limited to EU countries.

It will be particularly important in cases where AI systems assist or replace human decision-making or perform other tasks relevant to such contexts. AI systems shall only be used in a way that that does not, directly or indirectly, endanger or undermine democratic processes.

4. EVOLVING AI RISKS IMPACTING AI RISK MANAGEMENT

Risks evolveand require a specific, relevant, and constant risk management. Moreover, the impacts of risks should be identified and regularly measured. Risk management and impact assessment in common can help decrease the risk of the products or services [27]. A risk management system is intended to monitor and control risks. In turn, the impact is related to possible negative

[7]https://www.adalovelaceinstitute.org/blog/role-of-standards-in-ai-governance/

[8]http://fra.europa.eu/en/content/what-are-fundamental-rights (2009)

[9]https://www.citizensinformation.ie/en/government_in_ireland/european_government/eu_law/charter_of_fundamental_rights.html (1953)

effects of risk materialization and can also change during the product lifecycle due to internal or external changes [28].There is worldwide consensus on the importance of AI risk management topic, even if there are differences in its understanding (e.g., China - strong governmental impact, US business-driven, lobbying, Europe - fundamental rights protection and related regulations).

Given that AI is developed and appliedglobally (crossing borders, including beyond Europe), it is fundamental to establish a common global approach to basic principles governingthe development and use of AI systems. Risk management systems shall be set up, implemented, documented, and maintained to ensurethat high-risk AI systems comply with the AI Act. These systems will monitor AI systems (products or services) to identify, estimate, and evaluate risks (using relevant measures), reduce risk, mitigate consequences, and build awareness. An impact assessment analysis shall also be provided. Standards play a crucial role in supporting the implementation of proper risk management. ISO is a major player in this field, with its long tradition in system, quality, and risk management,. For instance, ISO 31000 [29] is a fundamental standard for risks and risk management of IT systems, and it serves as a solidfoundation for new standards concerning AI risk management. It contains several important definitions, primarily for risk (seen as an effect of uncertainty on objectives, deviation from expected results). It also describes consequence as an outcome of an event affecting objectives.This standard focuses on essential processes of risk assessment and management (identifies their different components).

Moreover, there are also works ofdifferent organizations on different AI risk management aspects, like AI Risk Management Framework (NIST) [30], OECD working on AI risk management interoperability[10], and some specific aspects like piloting G7code of conduct application[11], UNESCO focusing on ethical considerations[12]. An important work conducted by ENISAproposed some methods for AI risk level assessment and AI risk measurement[13].Other ongoing or published standardization works concerning AI system risk management also exist. For instance, ISO 23984 on AI risk management [31] and the other ISO 42001 on AI system management [32], already published, are referenced by other works. NIST 800-39 [33] and ISO 27005 [34] are the most common information and cybersecurity risk managementstandards. Several known digital risks are described in ISO 27001 [35]. ISO Fundamentals concern concepts and terminology (ISO 22989 [36]), a framework for AI (ISO 223053 [37]) that lists risk sources, events and outcomes, and stakeholders.CEN-CENELEC recognizes levels of risk corresponding to the AI Act. Various risk management frameworks are defined for different domains of AI implementation (e.g., ETSI[14], ENISA[15]). They should consider the different aspects of responsible AI principles. There is also a need for an AI lifecycle generic reference model. It should take into account data specifics in AI systems.

Data is one of the most valuable assets in AI. It is subject to continuous transformation along the AI Lifecycle stages, like Data Ingestion, Data Exploration, Data Pre-processing, Feature Importance (Selection), Training, Testing and Evaluation.

AI Lifecycle engages various actors and various types of assets such as computational resources, software, and even non-tangible assets like processes. Cultural aspects and how actors experience the knowledge can bring potential non-intentional threats (like non-intentional bias).It is essential to understand the AI Threat Landscape, have a common and unifying foundation for understand-

[10]https://doi.org/10.1787/ba602d18-en, Common guideposts to promote interoperability in AI risk management (OECD)

[11] https://oecd.ai/en/wonk/pilot-g7-monitoring

[12] https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

[13] https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation?v2=1

[14]https://doi.org/10.1080/13511610.2024.2349627, The role of ETSI in the EU's regulation and governance of artificial intelligence

[15] https://www.enisa.europa.eu/publications/multilayer-framework-for-good-cybersecurity-practices-for-ai

ing the potential of threats, and accordingly conduct targeted risk assessments.ENISA provides both, proposing an AI Lifecycle tosecurely manage the development and maintenance of AI systems [38].Risk mitigation is an element of risk management and its implementation will differ depending on the organization. It is a strategy to prepare for risks (by a proper risk identification) and minimize the risk impacts. It is also a process that has to be implemented in a company (see Figure 3). As such, comparable to risk reduction, risk mitigation takes steps to reduce the negative effects of threats and disasters on business continuity (like cyberattacks, weather events, and other reasons for physical or virtual damage).

Proper risk mitigation will be critical for the appropriate risk management of AI systems in light of the AI Act, especially for high-risk systems [39]. A case of residual

risks is not fully clear in the AI Act, and several requirements will relate to acceptability of them [40].

Figure 3. An exemplary process of risk mitigation. Source: TechTarget

5. RESEARCH PROBLEM

For companies in Europe, the European Union, through the AI Act, will dictate in a regulatory way how to treat risks in AI systems. A high-risk system (i.e., the one with a potentially high level of risk) will be an AI system from a specific application domain regardless of how high the probability of threat materializing would be for a specific system. The European approach can force companies to legitimize whether a particular AI system will be safe and respectful of fundamental rights and personal data protection. However, it creates confusion for many companies, from the cybersecurity perspective that is commonly applied by organizations (considering real risks of negative consequences identified for a particular system) due to the different understanding of the concept of risk by the AI Act and the cybersecurity approach (explained in the previous section). It also poses several open questions for the risk management of AI systems. In order to effectively manage risks and protect AI systems and users against the consequences of risks, an integrated approach must take place.

It should be done at all levels, beginning at the regulatory level(with different regulatory directives), across standardization, and towards practical implementation. In particular, from the business perspective, there is a significant gap between regulatory and implementation. Organizations have several questions and will soon be on the verge of facing specific practical issues concerning how AI risk management is implemented in compliance with AI Act regulations. Standards will not answer all the questions. They are at the intermediary level between regulation and im-

plementation, providing guidance, recommendations and explanations of mandatory elements that should be included in organizations' risk management systems. Still, a significant gap between the regulatory and practical implementation remains, which should be further investigated and supported. All this unambiguity in regulation and lack of clear interpretation causes an implementation challenge for companies. It is unclear what they can do or how to minimize the gap between the formal understanding of the concept of AI risks in the regulatory approach and the practical approach of the business world, bridged by standardization guidance.Standardisation can be helpful at a generic level, but proposing an appropriate implementation is still on the companies' side.

In the case of AI systems, we may have to deal not only with purely technical risks but also with risks related to so-called human or fundamental rights. Such risks should also be identified and managed for AI systems.Concerning ethical aspects of AI systems, companies could follow their Customer Social Responsibility (CSR) strategies and try to extend them to ethical aspects of the use of AI. For instance, [41] addresses the ethical challenges posed by the AI Act for construction engineering. It explores the concept of Corporate Digital Responsibility (CDR) as an extension of CSR. It is seen as a holistic approach integrating ethical issues and proposing appropriate digital transformation processes, emphasizing the necessity to implement ethicsbydesign. Case studies and expert interviews can be helpful in the proper implementation of efficient processes.

The roots of CSR date back to the end of the nineteenth century, and its fast development was observed inthe 1950s and later. An ISO 26000 standard proposes a pragmatic approach to CSR. It provides CSR Management System logic, including risk mitigation and value-creating processes (see Figure 4).In general, there is a question in which manner businesses can try to fill this gap. There is no one simple answer.Companies will not be able to fully manage AI risks without automating the way they are managed and creating strategic processes for risk management to effectively identify risks, or determine their severity (low, high probability of occurrence) and prioritize them. It should be a strategic approach at a top company level.The development and adoption of strategic business processes, primarily for AI risk and incident identification within a company, could be a solution for better management of existing AI risks. It should feed AI risk management systems in a systemic way. They should be cataloged and unified. Here, support from standardization could be offered in the form of a risk catalog proposed by CEN-CENELEC[16].

One of its outputs will be AI cartography, which can be used for high-risk systems with a challenge of keeping a relevant list of risks up-to-date. Providing a regularly updated catalog of risks can lead to the risk decrease thanks to risk structuring and systematization. It exemplifies the types of risks (potential of harm), harms (e.g., loss, damage or destruction of assets caused by a threat), and associated threats (processes magnifying the likelihood of a negative event) for AI applications. Understanding the above-mentioned elements of AI risk cartography and associating risks with responsible AI principles and the AI lifecycle phases (framing high level objectives for AI risk management) is crucial for proper AI risk management. There are several responsible AI principles, including fairness, security, privacy, transparency or explainability, as well as others, such as safety, environmental impact, availability. Besides that, data quality is critical for AI systems, and data processes should be protected by risk management. Providing concrete examples of threats and harms is beneficial for better understanding and identifying them. In order to receive a whole picture, risk and impact should be measured continuously. Such a risk catalog should be provided as a multi-layered approach using a list of risks reported by standardization bodies (as a generalized layer) and the organization's risks in the more specific layer. The specific

[16]https://www.etuc.org/sites/default/files/page/file/2024-05/AI%20standardisation%20Inclusiveness_Newsletter3.pdf

risks should be mapped to the more generic ones. It would ensure cooperation across the whole AI value chain of partners.

It is important to classify risks in an organization by introducing risk categories. It will allow further interoperation. Taxonomies should also be applied for that purpose [42]. For instance, NIST began to build an AI risk taxonomy [43] as a first attempt to collect and structure terminology related to AI risks, aimed at achieving consensus on it and using the value of community engagement.The proposed taxonomy structure consists of two levels of terms, i.e., categories of risk (technical, perceptual or human context, regulatory context) and characteristics of trustworthy systems.There are three types of attributes (technical, e.g., accuracy and robustness, socio-technical, e.g., explainability and privacy, and, finally, guiding principles to AI Trustworthiness like fairness, accountability and transparency).

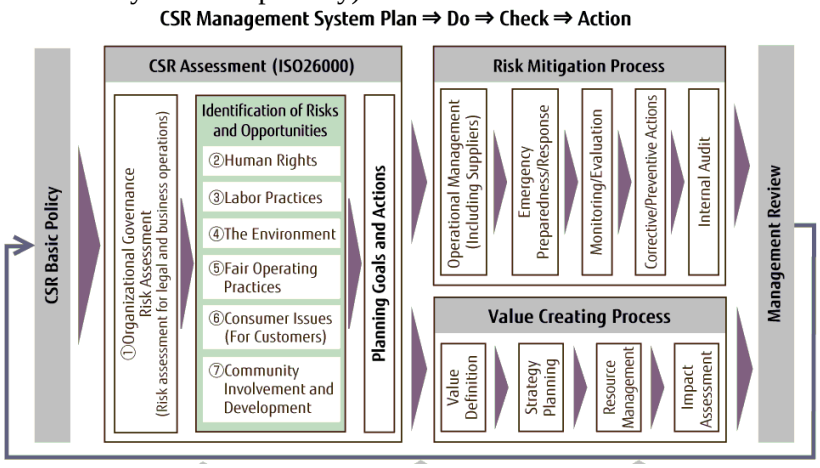

Figure 4. CSR Management System, source ISO 26000[17].

OECD builds a catalog of AI tools and metrics for trustworthy AI that is thought to support organizations. It is intended for a cooperative development by practitioners willing to participate in these activities. Another initiative that underpins the importance of a systematic approach for an effectively building trustworthy AI applications is the AI Assessment Catalog described in the guidelines for Trustworthy Artificial Intelligence [44]. It also shows that the requirements for trustworthy AI that are described abstractly must be made clear and tangible to enable their practical use.The following steps, built upon this systematization, i.e., introducing ontology and semantic layer in the AI risk management system, could help in automation, further systematization, and generalization of AI risk management. Good examples of existing works on ontologies are AIRO (AI Risk Ontology) presented in [45] or HART ontology of AI risks in the health domain.

As a result of the research conducted so far, we have outlined the first draft version of the resolving procedure and its steps, aiming at solving the problem of effective AI risk management. The high-level diagrampresents the draft procedure steps (see Figure 5) that will be further discussed with business and technical experts within the organization.

[17] https://www.iso.org/iso-26000-social-responsibility.html

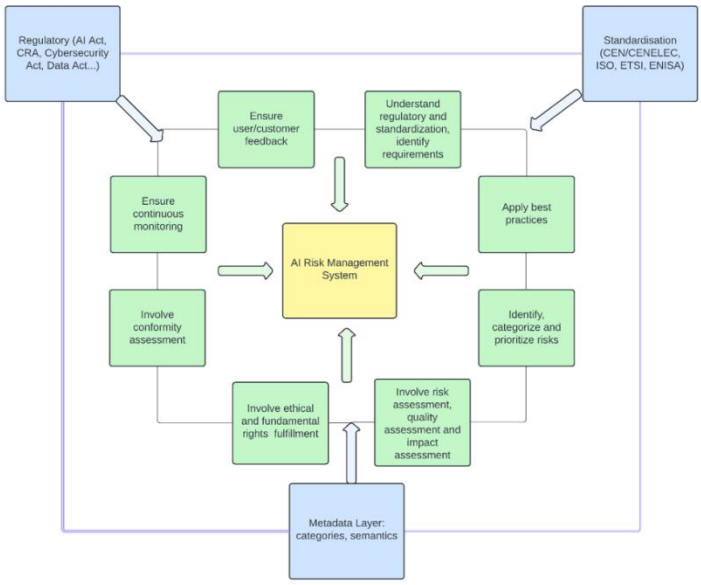

Figure 5. Draft of the AI risk management procedure (high-level view).

6. Discussion

AI systems introduce a high degree of complexity in managing, measuring, and assessing risks. In particular, they bring a significant challenge concerning societal and ethical aspects and the protection of human (fundamental) rights in this context. In Europe, the AI Act will regulate the development and use of AI systems. It is essential for businesses to build their strategies on what AI risks mean for them and how they should be managed according to this new regulation. It is also vital to prepare catalogs of risks at the level of organizations that are identified for the developed AI systems.

AI risk cartography developed by standardization can significantly help identify, prioritize, and mitigate risks and manage them. However, there is a problem concerning a non-exhaustive list of risks that may result from various reasons, such as the high variability of risks associated with Artificial Intelligence systems, insufficient knowledge about all the effects of the operation of this type of systems (not always predictable results), as well as an important aspect related to the specificity of individual domains of functioning of AI systems. From this research perspective, it is important to highlight the latter reason, as it depends on actions from different organizations. Therefore, the importance of cataloging risks within an organization, ensuring interoperability between organizations and risk generalization should beemphasized. It may impact the future use of predictions to anticipate risks. Several companies already have experience in building IT systems usinga risk-driven approach. However, they apply an approach where risks are associated with an IT system rather than focusing on a domain of their use. The AI Act introduces a categorization of IT systems depending on their applications. Companies should begin with revisiting their expertise and best practices in IT risk management to apply them to AI systems, and in the next step, enhance the approach and look for specificities. AI risk management should be provided on top of the „classical" risk management, considering that AI duplicates and strengthens existing risks and also introduces new ones (AI-specific).With new technologies (like Generative AI), new risks appear, but the methodology and processes for risk management are similar, i.e., identify, evaluate, and take actions.Concerning AI risks for societies, awareness of risks but also benefits are growing and people begin to participate in the collective risk evaluation. People are ready to accept the risk when they see the value in usage.Otherwise they do not accept it.

7. CONCLUSION

The AI risk and risk management environment is very complex and diverse compared to other kinds of digital product services and systems. There still remains a significant gap between the level of regulation and the implementation propositions. There are yet questions for organizations on how to implement risk management to be efficient and minimize risk materialization. This article points out elements that can be helpful in providing efficient implementations and better control of AI risk management. It highlights a strong need of systematization, generalization and interoperation, and way towards automation for more efficient AI risk management. It recalls that implementation is ultimatelythe responsibility of organizations andrequires concrete tools and facilities.

In order to answer appearing questions and validate the research results so far, further steps are required. They will include several workshops with business and technical experts across the organization (in the international environment) to validate the research findings and first draft procedures proposed to address the described challenges. Further work will require building a test environment and performing experiments involving a metadata and semantic layer to support proper risk identification and categorization. A tool supporting cataloging risks can be a part of further research. In the following stages of the research, the application of AI could be considered to operate the AI risk management system. It would allow for a more efficient identification of existing risks and the prediction of future ones. However, this process should be subject to specific research and oversight considerations. It would require setting up a complex Big Data environment and developing specific algorithms. A state-of-the-art review of tools capable of supporting this task is foreseen.

Acknowledgments

The completion of the study and research paper would not have been possible without great guidance from Emilie Sirvent-Hien (Orange research program Resp AI lead).

Disclosure of Interests

The author has no competing interests to declare that are relevant to the content of this article.

REFERENCES

[1] Jaffri A., Khandabattu H.: Hype Cycle for Artificial Intelligence. Gartner Report (2024)
[2] Pritchard CL.: Risk Management: Concepts and Guidance (5th ed.). Auerbach Publications. (2015)
[3] Habbal A., Ali M. K., Abuzaraida M.: Artificial Intelligence Trust, Risk and Security Management (AI TRiSM). Frameworks, applications, challenges and future research directions, Expert Systems with Applications. Vol. 240,2024, 122442, ISSN 0957-4174(2023)
[4] Raso, F., HilligossH., Krishnamurthy V.: Artificial Intelligence & Human Rights: Opportunities and Risks. Berkman Klein Center Research Publ. No. 2018-6, (2018)
[5] Zhang R.: The Prospects and Risks of Artificial Intelligence Industry. BCP Bus.& Mgmt. 34. (2022)
[6] NASA Risk Management Handbook. NASA SP, U.S., Vol. 3422 (2017)
[7] Moon J.: Foundations of Quality Risk Management, ISBN 9781951058333, ASQ Quality Press (2022)
[8] Williams R., Bertsch B., Dale B.: Quality and risk management: What are the key issues?. The TQM Magazine. 18. (2006)
[9] AmodeiD., OlahC., Steinhardt J.: Concrete problems in AI safety. arXiv (pp. 1–29) (2017)
[10] ISO/IEC 31000:2018, Risk Management - Guidelines. ISO. (2018)
[11] ISO Guide 73:2009, Risk Management - Vocabulary.ISO. (2009)

[12] Guide for Applying the Risk Management Framework to Federal Information Systems: A Security Life Cycle Approach. NIST Special Publication 800-37 Rev. 2. (2018)
[13] Threat Landscape 2021 - Cybersecurity Threats and Trends. ENISA. (2021)
[14] Tackling the Challenges of Cybersecurity (WP No. 18). ETSI.(2018)
[15] Cyber Resillience Act, https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act
[16] Threat Landscape 2023. ENISA. (2023).
[17] Guide for Conducting Risk Assessments. NIST Special Publication 800-30 Rev. 1). (2012)
[18] ISO/IEC 27005:2018, Information security risk management.ISO.(2018)
[19] Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1. NIST. (2018)
[20] ENISA Threat Landscape 2022 - Cybersecurity Threats and Trends. ENISA. (2022)
[21] Schuett J.: Risk Management in the Artificial Intelligence Act. Europ. J. of Risk Reg. (2023)
[22] Mueck M. D., On A. E. B. and Du Boispean S.: Upcoming European Regulations on Artificial Intelligence and Cybersecurity, in IEEE Comm. Mag., vol. 61, no. 7, pp. 98-102. (2023)
[23] Cyber Resillience Act, https://digital-strategy.ec.europa.eu/en/library/cyber-resilience-act
[24] Looking before we leap. Expanding ethical review processes for AI and data science research. Ada Lovelace Institute. (2022)
[25] Novelli C., Casolari, F. and Floridi L.: Taking AI risks seriously: a new assessment model for the AI Act. AI & SOCIETY. 1-5. 10.1007/s00146-023-01723-z. (2023)
[26] Liu, H-Y.: What makes AI regulation so difficult? European Liberal Forum (2024)
[27] Hopkin P.: ”Fundamentals of Risk Management: Understanding, Evaluating and Implementing Effective Risk Management”. Institute of Risk Management. (2018)
[28] Meagher, H., Dhirani, L.L:. Cyber-Resilience, Principles, and Practices.Springer.(2024)
[29] ISO/IEC 31000:2018, Risk management - Guidelines. ISO. (2018)
[30] AI Risk Management Framework. NIST AI-101. (2022)
[31] ISO 23894:2023, Artificial Intelligence - Guidance on Risk Management. ISO. (2023)
[32] ISO 42001:2023, Artificial Intelligence -Management System. ISO. (2023)
[33] Managing Information Security Risk, NIST 800-39. (2011)
[34] ISO 27005:2022, Guidance on managing information security risks.ISO. (2022)
[35] ISO 27001:2022, Information Security Management Systems – Requirements, ISO. (2022)
[36] ISO 22989:2022, Artificial Intelligence concepts and terminology. ISO. (2022)
[37] ISO 23053:2022Framework for Artificial Intelligence Systems Using Machine Learning. ISO. (2022)
[38] AI Threat Landscape 2021. ENISA. (2021)
[39] Jastroch, N.: Applied Artificial Intelligence: Risk Mitigation Matters. Springer. (2022).
[40] Fraser H. L, Bello y Villarino J-M.: Where Residual Risks Reside: A Comparative Approach to Art 9(4) of the European Union's Proposed AI Regulation.SSRN E. J. (2021)
[41] van der Merwe J, Al Achkar Z. Data responsibility, corporate social responsibility, and corporate digital responsibility. Data & Policy. (2022)
[42] Zeng Y., Klyman K., Zhou A.,AI Risk Categorization Decoded (AIR 2024): From Government Regulations to Corporate Policies. arXiv:2406.17864. (2024)
[43] Taxonomy of AI Risks – Draft. NIST. (2021)
[44] The Guideline for Trustworthy AI - AI Assessment Catalogue. arXiv:2307.03681. (2023)
[45] Golpayegani D., Pandit H. J.: AIRO: An Ontology for Representing AI Risks Based on the Proposed EU AI Act and ISO Risk Management Standards, Studies on the Semantic Web, Volume 55. (2024)