EasyChair Preprint

№ 10639

Risk Management of AI in Industry: a Literature Review

Paul Somer and Stefan Thalmann

July 31, 2023

RISK MANAGEMENT OF AI IN INDUSTRY: A LITERATURE REVIEW

Research full-length paper

Somer, Paul, BEHAM Techn. Handels GmbH, Vienna, Austria, p.somer@beham.com

Thalmann, Stefan, University of Graz, Graz, Austria, stefan.thalmann@uni-graz.at

Abstract

The fourth industrial revolution (Industry 4.0) is significantly changing industrial work by introducing interconnected machines and leveraging Artificial Intelligence (AI) to collect and analyze vast amounts of data. This technological advancement has led to increased reliance on AI systems, with some decisions being made automatically. However, the opaque nature of AI presents significant risks, demanding the development of suitable risk management approaches tailored to industrial settings. Currently, there is a lack of comprehensive research in this area, highlighting the need to conduct a structured literature review (SLR) to bridge this gap. Through the SLR, we identified various risk management approaches, including algorithm regulation, governance, certification, and auditing. Building upon these findings, we propose a research agenda for the management of AI risks in industrial settings.

Keywords: Industry 4.0, Artificial Intelligence (AI), AI Governance, De-risking AI, Regulation of AI.

1 Introduction

Digital transformation as an essential driver of Industry 4.0 is significantly changing business models and operations in the industry. At the heart of this transformation are data-driven approaches and Artificial Intelligence (AI) offering many benefits (Ibarra, Ganzarain and Igartua, 2018; Königstorfer and Thalmann, 2020). In addition, both the complexity of business activities and the amount of data to be processed have continuously grown (Bartodziej, 2017; Iafrate, 2018), dramatically increasing the pressure on companies to be competitive and efficient (Dossou, 2019; Wan *et al.*, 2018). As a result of this, data-driven approaches and AI are more and more adopted and responsibility in decision making is increasingly transferred from humans to machines (Stuurman and Lachaud, 2022).

AI is able to handle complex relationships hidden in big data sets and learn from them, which means that the decision-making process of AI remains opaque (Bathaee, 2018; Eschenbach, 2021). This characteristic of AI is often called the black box character of AI, which entails a corresponding risk potential for the user of AI – also in an industrial context (Savage, 2022). This can lead to erroneous decisions by AI leading to serious consequences for the organization. Thus a risk management (RM) framework is required, taking the specifics of AI applications into account (Erdélyi and Goldsmith, 2018; Iphofen and Kritikos, 2021; Reed, 2018). As a result, severe consequences should be minimized, corresponding user risks mitigated and adoption rates of AI in industrial applications should be increased.

Recently, law makers acknowledge this need and proposed regulations focusing on the protection of consumers and employees. The European Parliament passed the AI Act in June 2023 (artificialintelligenceact.eu, 2023; Pagallo, Ciani Sciolla and Durante, 2022; Veale and Zuiderveen Borgesius, 2021), whereas the National Institute of Standards and Technology (NIST) in the USA is working on a framework as a comparable regulatory approach to mitigate potential risks from corresponding AI applications (Barrett *et al.*, 2022). As a result, current research focuses primarily on the legal regula-

tion of AI from a consumer perspective. However, RM approaches focusing on the operations and especially in industrial settings are scarce.

This research gap forms the basis for the present publication and thus attempts to answer the following research question (RQ):

Which approaches to manage risks of industrial AI applications exist in the relevant literature?

The authors want to answer this research question by conducting a structured literature review (SLR) according to Webster & Watson (Webster and Watson, 2002). The SLR focuses on publications that consider approaches to risk minimization with regard to RM in relation to industrial AI applications.

2 Background

Digital transformation can be seen as integration of digital technologies and solutions in all areas of a company, whereby this change is based not only on a technological but also on a cultural basis (Schellinger, Tokarski and Kissling-Näf, 2020). In this regard, data analytics became the key driver of process- and product-innovations in industrial settings and the key source of competitive advantage in industry (Thalmann *et al.*, 2018). This trend builds on digital infrastructures and data-driven applications, which enable company-specific systems to communicate with each other along the entire Supply Chain (SC) and automatically take over corresponding tasks (Sorger *et al.*, 2021). This transformation process not only affects internal company structures (processes, employees, workflows, corporate structures, etc.), but also the embedded connections to customers and supplier systems leading to new business models (Appelfeller and Feldmann, 2018; Hess, 2019). From a research perspective, the industrial focus is relevant since the realizable potential benefits of the targeted applications of AI in industry are particularly high compared to other business areas (Abioye *et al.*, 2021).

The term AI was first established in 1956 and defined as '*the science and engineering of making intelligent machines, especially intelligent computer programs*' (Kersting, 2018). In accordance with increasing complexity and higher degree of use, the difficult question of a uniform definition of AI also gradually changed. Today, AI is generically defined as a task-processing technology which produces results that are supposed to be similar to those resulting from human action (Buiten, 2019). However, it must be stated at this point that there is still no uniform definition of AI to date (Monett and Lewis, 2018; Regona *et al.*, 2022; Wang, 2019). From a technical perspective, AI has meanwhile established itself as an important future technology that can take on corresponding tasks completely autonomously using algorithm-based data processing and also makes its own decisions regarding the way of processing (Cadavid *et al.*, 2019; Demary and Goecke, 2019). Accordingly, this development has promoted the emergence of completely new products and services as well as innovative business models due to the variety of use-cases that can be processed (Regona *et al.*, 2022).

The digital transformation and the associated structural change to the smart factory in the sense of Industry 4.0 also mean a strong increase of AI in the industrial environment (Balamurugan *et al.*, 2019). As a result, AI is a rapidly developing technology that is transforming the manufacturing industry (Arinez *et al.*, 2020). By enabling machines to learn from data and make decisions on their own, AI can help manufacturers to optimize production processes, reduce costs, and improve product quality (Javaid *et al.*, 2022). Thus, the potential benefits of AI in the industrial sector are very high and can therefore enable a substantial comparative competitive advantage (Cubric, 2020).

One of the most important benefits of AI in the industrial sector is the reduction of human intervention in the processing of business processes (Jarrahi, 2018). In this context, AI is able to automatically take over industrial processes and also make corresponding decisions (Jakhar and Kaur, 2020). In some areas, this enables labor-intensive processes to be shifted from people to machines. The capabilities of AI extend across the entire SC of an industrial company, from the procurement of raw materials through the entire manufacturing process to the distribution of finished goods (Liu, Song and Liu, 2023). In planning, AI can support the **optimization of processes** (Ammar *et al.*, 2021) and thus help

to reduce inventory costs and lead times (Sanchez, Exposito and Aguilar, 2020). In addition, almost the entire procurement of a company can be automated using AI (Bueno *et al.*, 2022) by determining the required raw materials or items on the basis of corresponding input data and ordering them automatically (Shivam and Gupta, 2023). Complementary to this, AI is capable to elaborate predictive analytics by analyzing historical production data to identify patterns and trends, helping manufacturers optimize production processes and anticipate demand (My, 2021).

Another field of application of AI in manufacturing processes is **quality control**, whereby images of products can be analyzed to detect defects or anomalies, allowing manufacturers to identify quality issues in real-time (Salkin *et al.*, 2018). This is realized by gathering, analyzing and interpreting production data in order to optimize production processes (Thalmann *et al.*, 2018). Furthermore, it is also possible to predict the quality of a product during manufacturing, which is particularly valuable for long-lasting production processes or very expensive quality inspections (Schuetz *et al.*, 2023). The data processing and transmission required for this can be fully automated in this context (using sensors and actuators), which in turn produces more accurate results with less human effort (Ammar *et al.*, 2021; Javaid *et al.*, 2022). A similar effect can be achieved by using AI in the field of robotics, in which AI can enable robots to **perform complex tasks**, such as assembly and packaging, with greater precision and efficiency than humans (Vrontis *et al.*, 2022). In addition, such AI applications in the field of robotics are able to detect any root cause of faults in automation at an early stage. In this context, AI-supported autonomous vehicles can also be used, which can transport production goods autonomously between factories and warehouses (Buntak, Kovacic and Mutavdzija, 2021). As a result, manufacturing processes are smarter and more productive while also ensuring a more efficient use of resources (Javaid *et al.*, 2022).

With regard to the stability and performance of industrial machines, the minimization of downtimes is an essential criterion for success (Henríquez-Alvarado *et al.*, 2019). In this context, corresponding AI applications are able to make a substantial contribution through **predictive maintenance** by analyzing time series data (Divya, Marath and Santosh Kumar, 2023; Gashi and Gursch *et al.*, 2022). AI can analyze sensor data from machinery to predict when equipment is likely to fail, allowing manufacturers to schedule maintenance before breakdowns occur (Gashi and Mutlu *et al.*, 2022). This allows necessary maintenance work to be scheduled and machine failures to be minimized.

The extensive technical capabilities of AI and the rapid development of applications mean that decision-making power is gradually being shifted from humans to machines (Zhang *et al.*, 2021). This suggests that appropriate RM is required as a safety mechanism (Zhang *et al.*, 2022). In general, RM is a systematic corporate process designed to support companies in dealing with risks (Tupa, Simota and Steiner, 2017). There are numerous models in the literature, such as model risk management and enterprise risk management, which are widely used in RM (Aristi Baquero *et al.*, 2020; Olson and Wu, 2020). With regard to their approach, they are basically carried out in the following steps: risk analysis (identification of risks), risk assessment, risk management and risk controlling (Tupa, Simota and Steiner, 2017). Such models are rather static in their mode of operation and are therefore hardly suitable for dynamic purposes, as would be necessary in relation to AI (Aristi Baquero *et al.*, 2020).

In this regard, regulators start to approach this problem as there are still no clear rules for regulating AI to date (Pagallo, Ciani Sciolla and Durante, 2022). For example the EU AI Act was passed by the European Parliament in June 2023 and in the USA the NIST is developing a corresponding framework that deals with the corresponding risks in dealing with AI. These regulations are mainly targeted towards protecting the end users and thus ensuring product safety. From an industry perspective this is a very important point, but risk in operations and production processes needs to be targeted as well. Thus, industry needs to establish holistic RM approaches taking the specifics of industry and AI into account. To the best of our knowledge, no overview of requirements, existing approaches or challenges of implementing a RM framework in industry exists so far.

3 Procedure

To analyze RM approaches for AI in industry, a Structured Literature Review (SLR) was conducted. The authors followed the approach proposed by Webster & Watson (Webster and Watson, 2002). The relevant publications were identified by defining primary and secondary expressions that appear to be relevant for the corresponding subject area (e.g. RM, artificial intelligence, regulation), whereby internationally common abbreviations were also taken into consideration (e.g. RM, AI etc.). The relevant journals, proceedings and other conference papers were also clarified via the search results (i.e. frequency of occurrence). The authors focused on key journals of every of the relevant four domains according to VHB[1] to investigate the best publications in a rigorous way (Webster and Watson, 2002). As an overview, the search terms were clustered into those domains (based on the VHB categories) and the publishing journals were also taken into account. This enabled the researched publications to be categorized accordingly and summarized in a matrix (see Appendix). The timespan-period of relevant publications was set from 2015 to 2023, since the topic around RM of AI has only become increasingly important in recent years.

The authors used Scopus, Web of Science and Google Scholar as target databases. The search focused on the presence of at least two secondary keywords or one primary keyword (regarding the title of the publications and their defined keywords). The terms risk management, artificial intelligence, AI, industry 4.0 and supply chain were used as primary keywords and the terms framework, regulation, governance, auditing and certification as secondary keywords (see Appendix). The search led to 226 publications in total (see Figure 1).

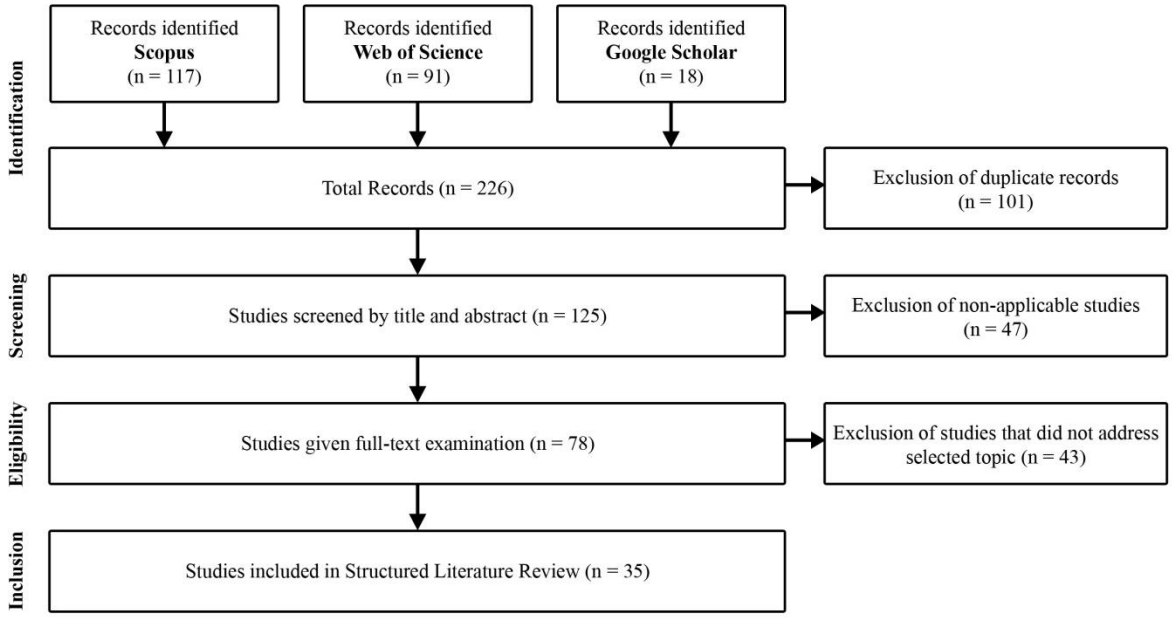

Figure 1. Visual Illustration of Structured Literature Review

Since a large number of these publications were published in several databases, a total of 101 duplicate records had to be removed. This left 125 publications for a corresponding abstract screening. A further 47 publications were excluded because they primarily focused on medicine or pharma. The

[1] https://vhbonline.org/wk-/-fachgruppen

subsequent content analysis of full texts was carried out with 78 relevant publications and a further 43 were excluded from the SLR because they did not address the selected topic. Some publications were included here that focused on how to do RM with AI. However, since these cannot contribute to answering the present RQ, these publications were also excluded. As a result of this procedure, **35 publications remained** for the SLR.

The authors analyzed these remaining 35 publications by applying the qualitative content analysis according to Mayring (Mayring, 2015). The content description of each individual publication was summarized in tabular form in order to provide an overview of the approaches to the subject area (RM of AI in industry). In the next step, the risks of AI in industry were identified and the approaches of RM of industrial AI applications were determined. Numerous frameworks in different areas have emerged as approaches to risk minimization (governmental and organizational). The authors examined these frameworks from the perspective of their implementation (design) phase, their operation and the necessary control and monitoring phase. The knowledge gained from this formed the basis for recommendations for corresponding regulators of AI applications in industrial surroundings. Finally, the results were interpreted and synthesized in a concept-centric way, as proposed by Webster & Watson (Webster and Watson, 2002).

4 Discussion of Results

As a result of the SLR we identified the major risks of AI in industry and found different approaches how organizations deal with AI risks in industrial settings. In this context, **regulation of algorithms and governance, certification and auditing of AI applications** were identified as core approaches. The results are structured regarding to the following phases: **planning-/development phase**, in the **operational phase** and in the **control-/monitoring phase**.

4.1 AI Risks in Industry

AI applications are getting more and more popular as they offer many advances to optimize industrial processes (Jaekel *et al.*, 2022; Ralph and Stockinger, 2020). As a result, AI applications are deployed in numerous settings, thus the associated risks have also increased substantially (Zhang *et al.*, 2022). AI is primarily based on historical data and learns independently from humans and is therefore prone to errors (Nikitaeva and Salem, 2022). If AI models are not sufficiently trained, validated or monitored, there is a risk of erroneous decisions or predictions that could lead to errors in production or other operational processes along the corporate SC (Vyhmeister, Gonzalez-Castane and Östbergy, 2023).

Since AI applications base on models trained with extensive amounts of data, the primary internal IT risks in the context of AI in the industry are model risks and data risks (Zhang *et al.*, 2022). Hence, both incorrect data (data bias) and the underlying models for processing the data (model bias) can lead to erroneous AI outcomes and thus to risks for the stability of an SC (Gopal *et al.*, 2022). For example, incorrect predictions of production figures could arise or AI-based quality controls in the manufacturing process could fail (Kehayov, Holder and Koch, 2022). Additionally, AI systems can be vulnerable to cyberattacks or hacks (external IT risks), which can lead to data leaks, production downtime, or other security-related issues (Zhang *et al.*, 2022). In addition, the enormous complexity of AI models often results in a lack of transparency and explainability (Vyhmeister, Gonzalez-Castane and Östbergy, 2023). AI models are therefore often viewed as ‘black boxes’ because they use complex algorithms that are difficult to understand and their decision-making path almost impossible to follow. For this reason, AI-supported decisions in the manufacturing process can hardly be traced, which means that the industrial user is exposed to high risks (Arinez *et al.*, 2020). This lack of transparency can lead to lower acceptance of AI systems by employees, customers or regulators (Cheatham, Javanmardian and Samandari, 2019).

Another risk factor is the lack of consideration for human-machine interaction (Bécue, Praça and Gama, 2021). The human oversight or '*Human in the Loop*' and thus human responsibility must be ensured (Bannister and Connolly, 2020; Mosqueira-Rey *et al.*, 2023). If this risk aspect is not taken into account, especially in autonomous transport systems, manufacturing, and infrastructure systems, ethical questions arise regarding the safety of workers involved (Cheatham, Javanmardian and Samandari, 2019).

The application of AI in the manufacturing industry is subject to various regulatory requirements, such as data protection, liability and compliance. Even if there are no uniform regulations to date, companies must ensure that they comply with existing regulations in order to minimize legal risks. Consequently, numerous frameworks relating to RM of AI are currently being established to minimize the corresponding risks and their probability of occurrence (Nikitaeva and Salem, 2022). The implementation of a corresponding RM framework for AI applications is usually divided into three phases. These are to be subdivided into the planning phase (design), the operational phase (operation) and the control and monitoring phase (control).

4.2 Design of AI Risk Management Frameworks

Based on the analysis of the literature we identified four relevant key activities during the design of a RM Framework for industrial AI applications:

First, the **identification of AI use cases** as well as their description along the entire SC is essential (Stuurman and Lachaud, 2022). This limitation of the company's internal fields of application subsequently leads to a clarification of the scope and the boundaries of the AI-supported system (Zhang *et al.*, 2022). In this context, typical applications in industry are, for example, (mobile) fraud detection, float optimization with regard to internal processes, accident propensity prediction, and predictive maintenance (Masood and Hashmi, 2019). The use of specified templates is recommended here, with which use cases can also be explained to other actors involved (Brunnbauer, Piller and Rothlauf, 2021). These templates should at least contain a description of their own application and its dependency on other use cases, ideally also the associated risks and probabilities of occurrence (Balamurugan *et al.*, 2019). In addition, in the industrial sector, the understanding of the production processes (business understanding) and the connection with the existing data (data understanding) is important to map the individual entrepreneurial needs (Brunnbauer, Piller and Rothlauf, 2021). Completely formulated use cases can be prioritized in a second round to plan their concrete order and final implementation (Greiner, Berger and Böck, 2022). At this point it makes sense to identify the potential risks of the individual use cases and to formulate the options for risk mitigation (Aristi Baquero *et al.*, 2020; Lauterbach, 2019). As a result, this analysis is also important to evaluate whether the maturity-level of the current IT system is even capable of deploying a corresponding RM framework for AI (Mäntymäki *et al.*, 2022; Quest *et al.*, 2022). If it is determined that the existing IT solutions are not sufficient to implement an adequate RM framework successfully, it is necessary to define the requirements at this point (Wirtz, Weyerer and Kehl, 2022).

After inspection of the suitability of the company's IT is clarified, it is recommendable to screen existing published frameworks in advance and examine them for their eligibility for use in the company (Butcher and Beridze, 2019). The **evaluation and adaptation of existing RM frameworks** usually makes more sense than the implementation of completely new approaches, since this allows the advantages and experiences of existing RM frameworks to be taken into account (Almeida, dos Santos and Farias, 2021). The layered model for AI risk regulation (Wirtz, Weyerer and Kehl, 2022) and the three-stage model (Clarity, Breadth and Nuance) for alleviating AI risks can be mentioned here as examples (Cheatham, Javanmardian and Samandari, 2019). At this point, it is advisable to ensure that the researched frameworks are used in the same industry or at least for the same purpose in terms of risk minimization of similar AI applications (Chesterman, 2019). This can optimize and facilitate the

elaboration and implementation of an adequate RM framework (Cheatham, Javanmardian and Samandari, 2019).

Another challenging step in the design phase of an AI RM Framework is the consideration of **risk measurement and quantification** (Bannister and Connolly, 2020). Based on the previously defined use cases, the focus here is on the identification of corresponding risks and their effects and probability of occurrence (Wirtz, Weyerer and Sturm, 2020). Especially in industry, the question arises which process should be handled by humans or AI and how the resulting risks can be measured and quantified in advance (Schneider *et al.*, 2022). This is necessary to measurably compare corresponding risks with expected benefits. Based on classic model risk management, it is advisable to adapt the elements for risk assessment (Aristi Baquero *et al.*, 2020). Taking into account the corresponding AI application and the respective use case, the following components should be considered in the RM Framework: type of algorithm, transparency, comprehensibility, and impact (Bannister and Connolly, 2020).

Finally, when implementing a RM framework for AI, the **consideration of legal aspects** in the industrial environment is important (Chambers, 2021; Mäntymäki *et al.*, 2022). This requires an overview of the applicable national AI regulations, which must be taken into account in any case (Mäntymäki *et al.*, 2022), although these governmental policies typically lag behind technological developments concerning AI (Stuurman and Lachaud, 2022). Nevertheless, it is advisable to implement any RM frameworks in accordance with international AI governance rules (Chambers, 2021; Ellul *et al.*, 2021). Due to the current research and the rapid spread of the use of AI applications, it can be assumed that there will be internationally similar approaches and governmental rules to regulate AI in the future – at least in the medium term.

Overall, an appropriate RM framework should be designed as a support tool to mitigate industrial risks under consideration of existing legal regulations (Quest *et al.*, 2022). The most important aspect in relation to industrial risks from the use of AI is the maintenance of machine performance with optimized process flows, which means that the RM framework must be able to minimize potential deficits in the areas mentioned (Nikitaeva and Salem, 2022). As a result of this, the design of a corresponding framework must at least ensure that is able to support safety and reliability of AI in the company (Zhang *et al.*, 2022). However, it should not contain any requirements that restrict AI innovation and thereby disrupt agile working methods (Aristi Baquero *et al.*, 2020; Wirtz, Weyerer and Kehl, 2022). In addition, conditions should already be established in the course of the design, which enable dynamic regulation of algorithms (Almeida, dos Santos and Farias, 2021; Chambers, 2021). The need for this arises on the one hand from the rapid technological developments in the field of AI and on the other hand from the currently volatile legal policies, which are not yet standardized internationally (Nikitaeva and Salem, 2022).

4.3 Operation of AI

With regard to the operation of AI in industrial applications, a corresponding RM framework must also meet proper requirements. One of the most important principles is that an AI application is never operated without **human oversight** as part of a corresponding risk mitigation strategy (Stuurman and Lachaud, 2022). In many cases, the use of corresponding AI applications in industry can therefore only be regarded as a supporting assistance, since wrong decisions by algorithms can lead to fatal consequences (machine downtime, supply bottlenecks as a result of wrong procurement decisions, etc.). The collaboration between humans and AI should therefore not be limited to the development phase (Mosqueira-Rey *et al.*, 2023). For this reason, the **use of explainable approaches** (such as XAI) seems advantageous, so that both the reasons for decisions and their connections can be representable in the RM framework. In addition, a corresponding RM framework should be designed in such a way that cross-departmental compliance requirements are defined and taken into account (Cheatham, Javanmardian and Samandari, 2019). However, corresponding internal company user rules for algorithm-based decision making must also be defined.

For security reasons, an **integrated corporate disaster management** must also be included at this point (including respective action plan) to minimize any damage resulting from the operation of AI in industry (Bannister and Connolly, 2020). Even if the majority of resulting errors in dealing with AI are caused by human actions (Senders and Moray, 2020), the organizational embedding of humans in the AI loop is essential. Ultimately, only the permanent consideration of humans in the loop as the highest control body can reduce the probability of errors occurring as a result of AI (Almeida, dos Santos and Farias, 2021; Lauterbach, 2019). This should be integrated into the RM framework and ideally be extended by appropriate risk classifications in terms of scope and scale to optimize early risk detection and risk prevention (Wirtz, Weyerer and Kehl, 2022). The factor of early detection appears to be essential, especially in the manufacturing process, since errors in consecutive process handling continue to cause enormous costs (interrupted SC, machine downtime, increased production rejects, etc.). For this reason, the **transparency of autonomous systems** is also an important risk minimization factor in the industrial environment to minimize algorithmic bias (Arslan, 2020; Chesterman, 2019; Zhang *et al.*, 2022). Since both the possible applications and the potential of AI in an industrial context are promising, from the point of view of RM, corresponding **requirements for algorithm-based decision making** are necessary and must therefore be mapped in a suitable RM framework (Zhang *et al.*, 2022).

4.4 Control and Monitoring of AI

With regard to the control and monitoring phase, a suitable RM framework must be able to cover two fundamental approaches to risk regulation, namely the consideration of **regulation ex ante and regulation ex post** (Butcher and Beridze, 2019). On the one hand, regulations that are already considered in the design phase and, in turn, those that arise from problems during operation will have to be taken into account. In course of this, it must be ensured that the control systems meet the industry-specific requirements and are in line with legal regulations (Butcher and Beridze, 2019). In the design phase, control mechanisms to manage analytical risks are usually applied after development is complete (Aristi Baquero *et al.*, 2020). In the subsequent deployment, continuous regulatory monitoring and reporting usually starts. Here, the implementation of regulatory metrics in the RM framework based on individual needs is highly recommended (Kurshan, Shen and Chen, 2020).

Referring to the embedding of control mechanisms in the course of implementing a RM framework, the **use of corresponding scenarios** is recommended (Aristi Baquero *et al.*, 2020). These should be similar to common RM models and broken down into best case, most likely case and worst case (Chambers, 2021). In general, the control and monitoring should take into account all risks of AI in operation, but those that are essential for maintaining the stability and performance of the SC and the manufacturing process are primarily mapped (Zhang *et al.*, 2022). Since machines are nowadays increasingly networked as part of Industry 4.0, it must be ensured in the monitoring phase that machine data is available in real time and analyzed in order to reveal corresponding risks in the SC process flow (machine breakdowns, robot malfunctions, etc.) as quickly as possible to identify and mitigate their consequences (Deshpande and Jogdand, 2020; Kim *et al.*, 2022).

The practical implementation of appropriate control mechanisms in a RM framework to reduce the occurrence of respective risks is recommended in three related steps (Aristi Baquero *et al.*, 2020): **context monitoring** (consideration of regulatory or legal changes, company policy changes and consideration of usage appropriateness etc.), **model monitoring** (data drift, model metrics, bias metrics etc.) and **model maintenance** (database of metrics and trend tracking, model optimization and updates of respective documentation etc.). The monitoring steps mentioned should ensure that the industrial company is able to continuously monitor and manage bias risk in production processes (Aristi Baquero *et al.*, 2020). Since the primary objective here is to minimize the error rate of decisions using AI to protect the stability of the SC, additional internal and external audits can be specified in the RM framework (Arslan, 2020; Ellul *et al.*, 2021).

5 Research Agenda and Implications

5.1 Research Agenda for AI in industry

Based on the presented findings and their implications, we can identify several research areas that should be pursued. These findings have led us to develop a research agenda specifically focused on managing the RM of industrial AI. It is important to acknowledge that in the industrial environment the requirements for a risk minimization framework significantly differ from those in other societal domains (Nikitaeva and Salem, 2022; Zhang *et al.*, 2022).

First, future research in RM of AI in industry should consider the unique challenges and complexities inherent to industrial settings. These may include factors such as high-stakes decision-making, safety-critical operations, and complex systems integration (Vyhmeister, Gonzalez-Castane and Östbergy, 2023). Our work show that some scattered challenges are mentioned in literature, but neither a comprehensive empirical inquiry nor a systematic conceptual work is currently available.

Second, specific and holistic RM for industrial AI applications does not exist yet. We found several attempts focusing on specific aspects or phases, but no approach covering the entire industrial RM life cycle. Thus, research and practice should put emphasis on the development and evaluation of industry-specific RM frameworks for AI (Nikitaeva and Salem, 2022).

Third, a classification of AI use cases in industry seems promising to guide practitioners as well as researchers. The classification of use cases should be discussed in the light of the industrial RM. In this regard it seems particular relevant to focus on sensitive industrial use cases in which the use of AI is generally recommended and beneficial, but currently not implemented due to the existing high risk potential (Brunnbauer, Piller and Rothlauf, 2021). It is important to work out the impact of the risk together with the corresponding industrial escalation levels and concrete preventive measures.

Fourth, a classification of risk mitigation strategies and measures for AI in industry is needed. Hitherto, we found general approaches and scattered discussion about industry-specific requirements. But a comprehensive work on measures taking the specifics of AI in industry into account is missing so far. Specifically, it seems promising to investigate which approaches are suitable for which industrial use cases and which are not. Existing RM frameworks should be examined for their aptitude in all sub-areas of manufacturing industry and, if necessary, checked whether they could be extended for sensitive industrial AI use cases (Vyhmeister, Gonzalez-Castane and Östbergy, 2023). Furthermore, it seems promising to investigate which measures are in line with existing industry standards or if industry standards need to be revised.

Fifth, exploring the implementation of suitable RM frameworks for AI within a broader context of organizational AI governance seems promising. Previous discussions on AI RM in industry have focused on isolated aspects so far. However, with the introduction of emerging regulations, such as the EU AI Act, a more comprehensive and holistic perspective is now required, particularly for high-risk use cases (Mäntymäki *et al.*, 2022; Papagiannidis *et al.*, 2023). As a part of this research avenue, future research should aim to investigate techniques for real-time risk assessment and mitigation while considering the impact of emerging no- and low-code data analytics as a service platform (Pangsuban, Nilsook and Wannapiroon, 2020). These platforms play an increasingly significant role in industrial AI and should be integrated into the overall RM framework.

Sixth, it is essential for research to concentrate on developing approaches to mitigate the challenges associated with RM in the context of AI. One significant challenge in this regard is the black-box nature of AI systems, which poses a particular obstacle for effective human-AI interaction (Savage, 2022). Therefore, exploring methods that enhance explainability and interpretability of AI models in industrial contexts holds significant promise (Gashi and Vuković *et al.*, 2022; Polzer *et al.*, 2022). Additionally, it is crucial to investigate into the concepts of the implementation of human-in-the-loop and human oversight, which have been frequently mentioned in the literature as suitable mitigation

strategies (Mosqueira-Rey *et al.*, 2023), but require careful design of the interaction between humans and AI (Kloker *et al.*, 2022). However, the implementation and enforcement of these strategies in real-world industrial settings remain unclear and require further investigation. Finally, the exploration of ethical and legal implications related to the deployment of AI in high-risk industrial environments is an area that has not been thoroughly scrutinized yet (Mäntymäki *et al.*, 2022). Understanding and addressing these implications are of importance to ensure the responsible and accountable use of AI technologies in industry (Königstorfer and Thalmann, 2022).

5.2 Implications

From a **managerial point of view**, the rapid adoption of AI in industry, driven by technological advancements, has created significant risks for industrial companies that need to be managed effectively (Kehayov, Holder and Koch, 2022; Kim *et al.*, 2022). Implementing robust RM frameworks specific to industrial settings is crucial to mitigate these risks (Nikitaeva and Salem, 2022; Vyhmeister, Gonzalez-Castane and Östbergy, 2023). Industrial companies must find ways to balance increasing AI adoption rates with the management of associated risks to remain competitive (Na *et al.*, 2022). Specifically, industry-specific approaches to AI RM should be developed to meet the unique needs of each sector (Jaekel *et al.*, 2022). In this regard our research agenda provides guidance for important areas concerning AI development.

From a **scientific perspective**, we showed that existing RM approaches focus on the influence of data processing on people, organizations, and society, but do not adequately consider the unique characteristics of AI systems, where machines learn and make independent decisions (Al-Qudah, 2022; Howard, 2019). Future theoretical models should incorporate the decision-making power of AI systems to provide a comprehensive understanding of AI RM (Amraoui *et al.*, 2019). Furthermore, existing machine directives in engineering sciences, such as industrial mechanical engineering, need to be expanded or adapted to account for the increased influence of AI in risk assessment for industrial applications. Our research agenda clearly points out the most relevant and urgent areas in industrial RM for AI.

From a **policy perspective**, managing the risks of industrial AI applications requires a combination of technical tools and consensus-driven standards (Butcher and Beridze, 2019). Regulatory frameworks must satisfy both individual company needs and legal requirements, considering potential rapid changes in governmental rules (Cath, 2018; Lauterbach, 2019). It is advisable to incorporate established tools such as model interpretability, bias detection, and performance monitoring into the regulatory measures (Aristi Baquero *et al.*, 2020; Kurshan, Shen and Chen, 2020). Additionally, the inclusion of specific roles can provide added value for risk mitigation (Ellul *et al.*, 2021).

5.3 Conclusion

The fourth industrial revolution (Industry 4.0) became reality and has greatly accelerated the process of digital transformation in the industrial landscape (Dohale *et al.*, 2023). Consequently, the technical features and their practical application in practice have continued to rapidly evolve (Ralph and Stockinger, 2020). The use of AI has also has also been progressively increasing in recent years, accompanied by an escalation in the associated risks for users (Wirtz, Weyerer and Kehl, 2022). For this reason, the topic of RM of AI is currently a hot topic for regulators worldwide. Both the NIST on the US side and the European Commission (implementation of the EU AI Act) are actively working to develop solutions to regulate the risks posed by these technologies. However, the approaches developed so far are primarily focussing on the protection of persons and personal data (in terms of ethics and values), which is of minor importance in the industrial context. Consequently, the requirements for risk-minimizing approaches are fundamentally different. In the industrial context, RM aims to minimize risks associated with decisions made by AI and is primarily concerned with the performance of processes and the stability of the corporate SC (Baryannis *et al.*, 2019).

For this reason, the present paper is intended to bring up a contribution by carrying out the question, **which approaches to manage risks of industrial AI applications exist in the relevant literature**. Based on this existing research gap, the authors conducted a SLR (Webster and Watson, 2002). The most important finding from the SLR is the fact that there is hardly any industry-specific framework in relation to RM from AI. The approaches found often have insufficient industrial focus and are therefore not able to meet their requirements. For this reason, the adoption rate of AI in industrial applications is lower than in those from other fields of application due to their high risk potential (Battistoni *et al.*, 2023). In conclusion, a research agenda with six promising areas for research on RM of AI in industry was proposed.

References

Abioye, S.O., Oyedele, L.O., Akanbi, L., Ajayi, A., Davila Delgado, J.M., Bilal, M., Akinade, O.O. and Ahmed, A. (2021) 'Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges', *Journal of Building Engineering*, 44, p. 103299. doi: 10.1016/j.jobe.2021.103299.

Almeida, P.G.R. de, dos Santos, C.D. and Farias, J.S. (2021) 'Artificial Intelligence Regulation: a framework for governance', *Ethics and Information Technology*, 23(3), pp. 505–525. doi: 10.1007/s10676-021-09593-z.

Al-Qudah, A.A. (2022) 'Artificial Intelligence in Practice: Implications for Information Systems Research, Case Study UAE Companies', in Musleh Al-Sartawi, A.M.A. (ed.) *Artificial Intelligence for Sustainable Finance and Sustainable Technology*. (Lecture Notes in Networks and Systems). Cham: Springer International Publishing, pp. 225–234.

Ammar, M., Haleem, A., Javaid, M., Walia, R. and Bahl, S. (2021) 'Improving material quality management and manufacturing organizations system through Industry 4.0 technologies', *Materials Today: Proceedings*, 45, pp. 5089–5096. doi: 10.1016/j.matpr.2021.01.585.

Amraoui, S., Elmaallam, M., Bensaid, H. and Kriouile, A. (2019) 'Information Systems Risk Management: Litterature Review', *Computer and Information Science*, 12(3), p. 1. doi: 10.5539/cis.v12n3p1.

Appelfeller, W. and Feldmann, C. (2018) *Die digitale Transformation des Unternehmens*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Arinez, J.F., Chang, Q., Gao, R.X., Xu, C. and Zhang, J. (2020) 'Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook', *Journal of Manufacturing Science and Engineering*, 142(11). doi: 10.1115/1.4047855.

Aristi Baquero, J., Burkhardt, R., Govindarajan, A. and Wallace, T. (2020) *Derisking AI by design: How to build risk management into AI development*. Available at: https://www.mckinsey.com/business-functions/quantumblack/our-insights/derisking-ai-by-design-how-to-build-risk-management-into-ai-development (Accessed: 7 August 2022).

Arslan, A.K. (2020) *A Design Framework For Auditing AI*. Available at: https://www.jmest.org/wp-content/uploads/JMESTN42353353.pdf.

Balamurugan, E., Laith, R.F., Yuvaraj, D., Sangeetha, K., Jayanthiladevi, A. and Senthil, K. (2019) 'Use Case of Artificial Intelligence in Machine Learning Manufacturing 4.0', *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, Dubai, United Arab Emirates, 11-12 December: IEEE, pp. 656–659. doi: 10.1109/ICCIKE47802.2019.9004327.

Bannister, F. and Connolly, R. (2020) 'Administration by algorithm: A Risk Management Framework', *Information Polity*, 25(4), pp. 471–490. doi: 10.3233/IP-200249.

Barrett, A.M., Hendrycks, D., Newman, J. and Nonnecke, B. (2022) *Actionable Guidance for High-Consequence AI Risk Management: Towards Standards Addressing AI Catastrophic Risks*. Available at: http://arxiv.org/pdf/2206.08966v3.

Bartodziej, C.J. (ed.) (2017) *The Concept Industry 4.0*. Wiesbaden: Springer Fachmedien Wiesbaden. Available at: https://doi.org/10.1007/978-3-658-16502-4 (Accessed: 20 March 2022).

Bathaee, Y. (2018) 'The Artificial Intelligence Black Box and the Failure of Intent and Causation', *Harvard Journal of Law & Technology*, 2018(Volume 31), pp. 890–938.

Battistoni, E., Gitto, S., Murgia, G. and Campisi, D. (2023) 'Adoption paths of digital transformation in manufacturing SME', *International Journal of Production Economics*, 255, p. 108675. doi: 10.1016/j.ijpe.2022.108675.

Bécue, A., Praça, I. and Gama, J. (2021) 'Artificial Intelligence, cyber-threats and Industry 4.0: challenges and opportunities', *Artificial Intelligence Review*, 54(5), pp. 3849–3886. doi: 10.1007/s10462-020-09942-2.

Brunnbauer, M., Piller, G. and Rothlauf, F. (2021) 'idea-AI: Developing a Method for the Systematic Identification of AI Use Cases'.

Bueno, R.E., Freitas Junior, M. de, Lombardi, I., Da Silva, M.A., Bueno, J.V. and Toloi, R.C. (2022) 'PROCUREMENT 4.0: MODEL-BASED EVOLUTION', in Digital, E.C. (ed.) *Open Science Research:* Editora Científica Digital, pp. 2766–2786.

Buiten, M.C. (2019) 'Towards Intelligent Regulation of Artificial Intelligence', *European Journal of Risk Regulation*, 10(1), pp. 41–59. doi: 10.1017/err.2019.8.

Buntak, K., Kovacic, M. and Mutavdzija, M. (2021) 'The Influence Of Industry 4.0 On Transport And Logistics In Context Of Supply Chains'. *Business Logistics in Modern Management*. Available at: http://www.efos.unios.hr/repec/osi/bulimm/PDF/BusinessLogisticsinModernManagement21/blimm2124.pdf.

Butcher, J. and Beridze, I. (2019) 'What is the State of Artificial Intelligence Governance Globally?' *The RUSI Journal*, 164(5-6), pp. 88–96. doi: 10.1080/03071847.2019.1694260.

Cadavid, J.P.U., Lamouri, S., Grabot, B. and Fortin, A. (2019) 'Machine Learning in Production Planning and Control: A Review of Empirical Literature', *IFAC-PapersOnLine*, 52(13), pp. 385–390. doi: 10.1016/j.ifacol.2019.11.155.

Cath, C. (2018) 'Governing artificial intelligence: ethical, legal and technical opportunities and challenges', *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 376(2133). doi: 10.1098/rsta.2018.0080.

Chambers, A. (2021) *Artificial Intelligence Risk Management Framework*. Available at: https://www.federalregister.gov/documents/2021/07/29/2021-16176/artificial-intelligence-risk-management-framework.

Cheatham, B., Javanmardian, K. and Samandari, H. (2019) *Confronting the risks of Artificial Intelligence*. Available at: https://www.cognitivescale.com/wp-content/uploads/2019/06/confronting_ai_risks_-_mckinsey.pdf (Accessed: 19 September 2022).

Chesterman, S. (2019) 'Should We Regulate A.I.? Can We?' *SSRN Electronic Journal*. doi: 10.2139/ssrn.3357618.

Cubric, M. (2020) 'Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study', *Technology in Society*, 62, p. 101257. doi: 10.1016/j.techsoc.2020.101257.

Demary, V. and Goecke, H. (2019) 'Künstliche Intelligenz: Deutsche Unternehmen zwischen Risiko und Chance', *IW-Trends - Vierteljahresschrift zur empirischen Wirtschaftsforschung* (Volume 46 (Issue 4)), pp. 3–18. Available at: https://doi.org/10.2373/1864-810X.19-04-01.

Deshpande, S.N. and Jogdand, R.M. (2020) 'A Survey on Internet of Things (IoT), Industrial IoT (IIoT) and Industry 4.0'.

Divya, D., Marath, B. and Santosh Kumar, M.B. (2023) 'Review of fault detection techniques for predictive maintenance', *Journal of Quality in Maintenance Engineering*, 29(2), pp. 420–441. doi: 10.1108/JQME-10-2020-0107.

Dohale, V., Verma, P., Gunasekaran, A. and Akarte, M. (2023) 'Manufacturing strategy 4.0: a framework to usher towards industry 4.0 implementation for digital transformation', *Industrial Management & Data Systems*, 123(1), pp. 10–40. doi: 10.1108/IMDS-12-2021-0790.

Dossou, P.-E. (2019) 'Development of a new framework for implementing Industry 4.0 in companies', *Procedia Manufacturing*, 38, pp. 573–580. doi: 10.1016/j.promfg.2020.01.072.

Ellul, J., Pace, G., McCarthy, S., Sammut, T., Brockdorff, J. and Scerri, M. (2021) 'Regulating Artificial Intelligence: A Technology Regulator´s Perspective', *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law, ICAIL '21: Eighteenth International Conference for Artificial Intelligence and Law*. ACM Special Interest Group on Artificial Intelligence, São Paulo Brazil, 21 06 2021 25 06 2021. New York,NY,United States: Association for Computing Machinery, pp. 190–194. doi: 10.1145/3462757.3466093.

Erdélyi, O.J. and Goldsmith, J. (2018) 'Regulating Artificial Intelligence: Proposal for a Global Solution', *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, AIES '18: AAAI/ACM Conference on AI, Ethics, and Society*, New Orleans LA USA, 02 02 2018 03 02 2018. New York, NY, USA: ACM, pp. 95–101. doi: 10.1145/3278721.3278731.

Eschenbach, W.J. von (2021) 'Transparency and the Black Box Problem: Why We Do Not Trust AI', *Philosophy & Technology*, 34(4), pp. 1607–1622. doi: 10.1007/s13347-021-00477-0.

EU Commission (2023) *The AI Act*. Available at: https://artificialintelligenceact.eu/.

Gashi, M., Gursch, H., Hinterbichler, H., Pichler, S., Lindstaedt, S. and Thalmann, S. (2022) 'MEDEP: Maintenance Event Detection for Multivariate Time Series Based on the PELT Approach', *Sensors (Basel, Switzerland)*, 22(8). doi: 10.3390/s22082837.

Gashi, M., Mutlu, B., Lindstaedt, S. and Thalmann, S. (2022) *No Time to Crash: Visualizing Interdependencies for Optimal Maintenance Scheduling*. Available at: https://graz.pure.elsevier.com/en/publications/no-time-to-crash-visualizing-interdependencies-for-optimal-mainte.

Gashi, M., Vuković, M., Jekic, N., Thalmann, S., Holzinger, A., Jean-Quartier, C. and Jeanquartier, F. (2022) 'State-of-the-Art Explainability Methods with Focus on Visual Analytics Showcased by Glioma Classification', *BioMedInformatics*, 2(1), pp. 139–158. doi: 10.3390/biomedinformatics2010009.

Gopal, P.R.C., Rana, N.P., Krishna, T.V. and Ramkumar, M. (2022) 'Impact of big data analytics on supply chain performance: an analysis of influencing factors', *Annals of Operations Research*. doi: 10.1007/s10479-022-04749-6.

Greiner, R., Berger, D. and Böck, M. (2022) 'Design Thinking und Data Thinking', in Greiner, R., Berger, D. and Böck, M. (eds.) *Analytics und Artificial Intelligence*. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 37–66.

Henríquez-Alvarado, F., Luque-Ojeda, V., Macassi-Jauregui, I., Alvarez, J.M. and Raymundo-Ibañez, C. (2019) 'Process Optimization Using Lean Manufacturing to Reduce Downtime', *Proceedings of the 2019 5th International Conference on Industrial and Business Engineering, ICIBE 2019: 2019 The 5th International Conference on Industrial and Business Engineering*, Hong Kong Hong Kong, 27 09 2019 29 09 2019. New York, NY, USA: ACM, pp. 261–265. doi: 10.1145/3364335.3364383.

Hess, T. (2019) *Digitale Transformation strategisch steuern*. Wiesbaden: Springer Fachmedien Wiesbaden. Available at: https://doi.org/10.1007/978-3-658-24475-0.

Howard, J. (2019) 'Artificial intelligence: Implications for the future of work', *American Journal of Industrial Medicine*, 62(11), pp. 917–926. doi: 10.1002/ajim.23037.

Iafrate, F. (2018) *Artificial Intelligence and Big Data: The birth of a new intelligence*. (Advances in information systems set, Volume 8). Hoboken, NJ: Wiley; ISTE Ltd.

Ibarra, D., Ganzarain, J. and Igartua, J.I. (2018) 'Business model innovation through Industry 4.0: A review', *Procedia Manufacturing*, 22, pp. 4–10. doi: 10.1016/j.promfg.2018.03.002.

Iphofen, R. and Kritikos, M. (2021) 'Regulating artificial intelligence and robotics: ethics by design in a digital society', *Contemporary Social Science*, 16(2), pp. 170–184. doi: 10.1080/21582041.2018.1563803.

Jaekel, F.-W., Nieto, M.T.A., Scholz, J.-A. and Bode, D. (2022) *Risk Assessment for Artificial Intelligence Applications in Manufacturing*.

Jakhar, D. and Kaur, I. (2020) 'Artificial intelligence, machine learning and deep learning: definitions and differences', *Clinical and Experimental Dermatology*, 45(1), pp. 131–132. doi: 10.1111/ced.14029.

Jarrahi, M.H. (2018) 'Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making', *Business Horizons*, 61(4), pp. 577–586. doi: 10.1016/j.bushor.2018.03.007.

Javaid, M., Haleem, A., Singh, R.P. and Suman, R. (2022) 'Artificial Intelligence Applications for Industry 4.0: A Literature-Based Study', *Journal of Industrial Integration and Management*, 07(01), pp. 83–111. doi: 10.1142/S2424862221300040.

Kehayov, M., Holder, L. and Koch, V. (2022) 'Application of artificial intelligence technology in the manufacturing process and purchasing and supply management', *Procedia Computer Science*, 200, pp. 1209–1217. doi: 10.1016/j.procs.2022.01.321.

Kersting, K. (2018) 'Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines', *Frontiers in Big Data*, 1, p. 6. doi: 10.3389/fdata.2018.00006.

Kim, S.W., Kong, J.H., Lee, S.W. and Lee, S. (2022) 'Recent Advances of Artificial Intelligence in Manufacturing Industrial Sectors: A Review', *International Journal of Precision Engineering and Manufacturing*, 23(1), pp. 111–129. doi: 10.1007/s12541-021-00600-3.

Kloker, A., Fleiß, J., Koeth, C., Kloiber, T., Ratheiser, P. and Thalmann, S. (2022) 'Caution or Trust in AI? How to design XAI in sensitive Use Cases?'. *AMCIS 2022 Proceedings*. Available at: https://aisel.aisnet.org/amcis2022/sig_dsa/sig_dsa/16.

Königstorfer, F. and Thalmann, S. (2020) 'Applications of Artificial Intelligence in commercial banks – A research agenda for behavioral finance', *Journal of Behavioral and Experimental Finance*, 27, p. 100352. doi: 10.1016/j.jbef.2020.100352.

Königstorfer, F. and Thalmann, S. (2022) 'AI Documentation: A path to accountability', *Journal of Responsible Technology*, 11, p. 100043. doi: 10.1016/j.jrt.2022.100043.

Kurshan, E., Shen, H. and Chen, J. (2020) 'Towards self-regulating AI', *Proceedings of the First ACM International Conference on AI in Finance, ICAIF '20: ACM International Conference on AI in Finance*, New York New York, 15 10 2020 16 10 2020. New York, NY, USA: ACM, pp. 1–8. doi: 10.1145/3383455.3422564.

Lauterbach, A. (2019) 'Artificial intelligence and policy: quo vadis?' *Digital Policy, Regulation and Governance*, 21(3), pp. 238–263. doi: 10.1108/DPRG-09-2018-0054.

Liu, L., Song, W. and Liu, Y. (2023) 'Leveraging digital capabilities toward a circular economy: Reinforcing sustainable supply chain management with Industry 4.0 technologies', *Computers & Industrial Engineering*, 178, p. 109113. doi: 10.1016/j.cie.2023.109113.

Mäntymäki, M., Minkkinen, M., Birkstedt, T. and Viljanen, M. (2022) 'Defining organizational AI governance', *AI and Ethics*, 2(4), pp. 603–609. doi: 10.1007/s43681-022-00143-x.

Masood, A. and Hashmi, A. (2019) 'AI Use Cases in the Industry', in Masood, A. and Hashmi, A. (eds.) *Cognitive Computing Recipes.* Berkeley, CA: Apress, pp. 383–396.

Mayring, P. (2015) 'Qualitative Content Analysis: Theoretical Background and Procedures', in Bikner-Ahsbahs, A., Knipping, C. and Presmeg, N. (eds.) *Approaches to Qualitative Research in Mathematics Education.* (Advances in Mathematics Education). Dordrecht: Springer Netherlands, pp. 365–380. Available at: http://dx.doi.org/10.1007/978-94-017-9181-6_13 (Accessed: 1 April 2022).

Monett, D. and Lewis, C.W.P. (2018) 'Getting Clarity by Defining Artificial Intelligence—A Survey', in Müller, V.C. (ed.) *Philosophy and Theory of Artificial Intelligence 2017.* (Studies in Applied Philosophy, Epistemology and Rational Ethics). Cham: Springer International Publishing, pp. 212–214.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J. and Fernández-Leal, Á. (2023) 'Human-in-the-loop machine learning: a state of the art', *Artificial Intelligence Review*, 56(4), pp. 3005–3054. doi: 10.1007/s10462-022-10246-w.

My, C.A. (2021) 'The Role of Big Data Analytics and AI in Smart Manufacturing: An Overview', in Kumar, R., Quang, N.H., Kumar Solanki, V., Cardona, M. and Pattnaik, P.K. (eds.) *Research in Intelligent and Computing in Engineering.* (Advances in Intelligent Systems and Computing). Singapore: Springer Singapore, pp. 911–921.

Na, S., Heo, S., Han, S., Shin, Y. and Roh, Y. (2022) 'Acceptance Model of Artificial Intelligence (AI)-Based Technologies in Construction Firms: Applying the Technology Acceptance Model (TAM) in Combination with the Technology–Organisation–Environment (TOE) Framework', *Buildings*, 12(2), p. 90. doi: 10.3390/buildings12020090.

Nikitaeva, A.Y. and Salem, A.-B.M. (2022) 'Institutional Framework for The Development of Artificial Intelligence in The Industry', *Journal of Institutional Studies*, 14(1), pp. 108–126. doi: 10.17835/2076-6297.2022.14.1.108-126.

Olson, D.L. and Wu, D. (2020) *Enterprise Risk Management Models*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Pagallo, U., Ciani Sciolla, J. and Durante, M. (2022) 'The environmental challenges of AI in EU law: lessons learned from the Artificial Intelligence Act (AIA) with its drawbacks', *Transforming Government: People, Process and Policy*, 16(3), pp. 359–376. doi: 10.1108/TG-07-2021-0121.

Pangsuban, P., Nilsook, P. and Wannapiroon, P. (2020) 'A Real-time Risk Assessment for Information System with CICIDS2017 Dataset Using Machine Learning', *International Journal of Machine Learning and Computing*, 10(3), pp. 465–470. doi: 10.18178/ijmlc.2020.10.3.958.

Papagiannidis, E., Enholm, I.M., Dremel, C., Mikalef, P. and Krogstie, J. (2023) 'Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes', *Information Systems Frontiers : a Journal of Research and Innovation*, 25(1), pp. 123–141. doi: 10.1007/s10796-022-10251-y.

Polzer, A., Fleiß, J., Ebner, T., Kainz, P., Koeth, C. and Thalmann, S. (2022) 'Validation of AI-based Information Systems for Sensitive Use Cases: Using an XAI Approach in Pharmaceutical Engineering', *Proceedings of the 55th Hawaii International Conference on System Sciences, Hawaii International Conference on System Sciences:* Hawaii International Conference on System Sciences. doi: 10.24251/HICSS.2022.186.

Quest, H., Cauz, M., Heymann, F., Rod, C., Perret, L., Ballif, C., Virtuani, A. and Wyrsch, N. (2022) 'A 3D indicator for guiding AI applications in the energy sector', *Energy and AI*, 9, p. 100167. doi: 10.1016/j.egyai.2022.100167.

Ralph, B.J. and Stockinger, M. (2020) *Digitalization and Digital Transformation in Metal Forming: Key Technologies, Challenges and current Developments of Industry 4.0 Applications*. Available at: https://www.researchgate.net/publication/343826977.

Reed, C. (2018) 'How should we regulate artificial intelligence?' *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 376(2128). doi: 10.1098/rsta.2017.0360.

Regona, M., Yigitcanlar, T., Xia, B. and Li, R.Y.M. (2022) 'Opportunities and Adoption Challenges of AI in the Construction Industry: A PRISMA Review', *Journal of Open Innovation: Technology, Market, and Complexity*, 8(1), p. 45. doi: 10.3390/joitmc8010045.

Salkin, C., Oner, M., Ustundag, A. and Cevikcan, E. (2018) 'A Conceptual Framework for Industry 4.0', in Ustundag, A. and Cevikcan, E. (eds.) *Industry 4.0: Managing The Digital Transformation.* (Springer Series in Advanced Manufacturing). Cham: Springer International Publishing, pp. 3–23.

Sanchez, M., Exposito, E. and Aguilar, J. (2020) 'Autonomic computing in manufacturing process coordination in industry 4.0 context', *Journal of Industrial Information Integration*, 19, p. 100159. doi: 10.1016/j.jii.2020.100159.

Savage, N. (2022) 'Breaking into the black box of artificial intelligence', *Nature*. doi: 10.1038/d41586-022-00858-1.

Schellinger, J., Tokarski, K.O. and Kissling-Näf, I. (2020) *Digitale Transformation und Unternehmensführung*. Wiesbaden: Springer Fachmedien Wiesbaden.

Schneider, J., Abraham, R., Meske, C. and vom Brocke, J. (2022) 'Artificial Intelligence Governance For Businesses', *Information Systems Management*, pp. 1–21. doi: 10.1080/10580530.2022.2085825.

Schuetz, C.G., Selway, M., Thalmann, S. and Schrefl, M. (2023) 'Discovering Actionable Knowledge for Industry 4.0: From Data Mining to Predictive and Prescriptive Analytics', in Vogel-Heuser, B. and Wimmer, M. (eds.) *Digital Transformation.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 337–362.

Senders, J.W. and Moray, N.P. (2020) *Human Error: Cause, Prediction, and Reduction*: CRC Press.

Shivam and Gupta, M. (2023) 'Quality process reengineering in industry 4.0: A BPR perspective', *Quality Engineering*, 35(1), pp. 110–129. doi: 10.1080/08982112.2022.2098044.

Sorger, M., Ralph, B.J., Hartl, K., Woschank, M. and Stockinger, M. (2021) 'Big Data in the Metal Processing Value Chain: A Systematic Digitalization Approach under Special Consideration of Standardization and SMEs', *Applied Sciences*, 11(19), p. 9021. doi: 10.3390/app11199021.

Stuurman, K. and Lachaud, E. (2022) 'Regulating AI. A label to complete the proposed Act on Artificial Intelligence', *Computer Law & Security Review*, 44, p. 105657. doi: 10.1016/j.clsr.2022.105657.

Thalmann, S., Mangler, J., Schreck, T., Huemer, C., Streit, M., Pauker, F., Weichhart, G., Schulte, S., Kittl, C., Pollak, C., Vukovic, M., Kappel, G., Gashi, M., Rinderle-Ma, S., Suschnigg, J., Jekic, N. and Lindstaedt, S. (2018) 'Data Analytics for Industrial Process Improvement A Vision Paper', *2018 IEEE 20th Conference on Business Informatics (CBI), 2018 IEEE 20th Conference on Business Informatics (CBI)*, Vienna, 11-14 July: IEEE, pp. 92–96. doi: 10.1109/CBI.2018.10051.

Tupa, J., Simota, J. and Steiner, F. (2017) 'Aspects of Risk Management Implementation for Industry 4.0', *Procedia Manufacturing*, 11, pp. 1223–1230. doi: 10.1016/j.promfg.2017.07.248.

Veale, M. and Zuiderveen Borgesius, 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*, 22(4), pp. 97–112. doi: 10.9785/cri-2021-220402.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A. and Trichina, E. (2022) 'Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review', *The International Journal of Human Resource Management*, 33(6), pp. 1237–1266. doi: 10.1080/09585192.2020.1871398.

Vyhmeister, E., Gonzalez-Castane, G. and Östbergy, P.-O. (2023) 'Risk as a driver for AI framework development on manufacturing', *AI and Ethics*, 3(1), pp. 155–174. doi: 10.1007/s43681-022-00159-3.

Wan, J., Yang, J., Wang, Z. and Hua, Q. (2018) 'Artificial Intelligence for Cloud-Assisted Smart Factory', *IEEE Access*, 6, pp. 55419–55430. doi: 10.1109/ACCESS.2018.2871724.

Wang, P. (2019) 'On Defining Artificial Intelligence', *Journal of Artificial General Intelligence*, 10(2), pp. 1–37. doi: 10.2478/jagi-2019-0002.

Webster, J. and Watson, R.T. (2002) *Analyzing the Past to Prepare for the Future: Writing a Literature Review*. Available at: http://www.jstor.org/stable/4132319.

Wirtz, B.W., Weyerer, J.C. and Kehl, I. (2022) 'Governance of artificial intelligence: A risk and guideline-based integrative framework', *Government Information Quarterly*, 39(4), p. 101685. doi: 10.1016/j.giq.2022.101685.

Wirtz, B.W., Weyerer, J.C. and Sturm, B.J. (2020) 'The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration', *International Journal of Public Administration*, 43(9), pp. 818–829. doi: 10.1080/01900692.2020.1749851.

Zhang, X., Chan, F.T.S., Yan, C. and Bose, I. (2022) 'Towards risk-aware artificial intelligence and machine learning systems: An overview', *Decision Support Systems*, 159, p. 113800. doi: 10.1016/j.dss.2022.113800.

Zhang, Z., Singh, J., Gadiraju, U. and Anand, A. (2021) 'Dissonance Between Human and Machine Understanding'. doi: 10.48550/arXiv.2101.07337.

Appendix

DOMAIN		PROD*	BI*	TIE*	FURTHER AREAS
Journal/Conference		*Elsevier (Artificial Intelligence)* *International Journal of Production Research* *Journal of Intelligent Manufacturing* *Science Robotics* *Springer (AI and Ethics)*	*ACM Computing Survey* *Big Data and Cognitive Computing* *Decision Support Systems* *IEEE Computational Intelligence Magazine* *Information Systems Management* *International Journal of Computer Vision* *International Journal of Information Management* *IOS Press (Information Polity)* *Springer (Ethics and Information Technology)* *Science Direct (Computer Law & Security Review)*	*Artificial Intelligence Review* *EAI Endorsed Transactions on Creative Technologies* *Engineering Applications of Artificial Intelligence* *Harvard Journal of Law & Technology* *ICAIL '21 (International Conference on AI and Law)* *International Journal of Artificial Intelligence* *Journal of Multidisciplinary Engineering Science & Technology* *MDPI (Sustainability)* *Science Direct (Energy and AI)*	*AIES'18 (Conference on AI, Ethics, and Society)* *BSA Software Alliance* *Emerald (Digital Policy, Regulation and Governance)* *European Journal of Risk Regulation* *Government Information Quarterly* *ICAIF '20 (International Conference on AI in Finance)* *IEEE Xplore* *International Journal of Public Administration* *Journal of the Academy of Social Sciences* *McKinsey Analytics (Quarterly)* *Network Industries Quarterly* *Royal Society Publishing* *The RUSI Journal* *UUM Journal of Legal Studies*
Primary Keywords	**Risk Management (RM)**	3	9		4
	Artificial Intelligence (AI)	7	2	5	15
	Industry 4.0	6	9	3	2
	Supply Chain (SC)	4	17	1	3
Secondary Keywords	**Framework**		3	5	14
	Regulation				3
	Governance		2		1
	Auditing		1		1
	Certification		1		
TOTAL		**20**	**44**	**14**	**43**

***Abbreviations:** PROD: Production Economy, BI: Business Informatics, TIE: Technology, Innovation & Entrepreneurship*