Towards Model-Based Assessment of Trustworthiness in Autonomous Cyber-Physical Production Systems

Maryam Zahid[1(✉)], Alessio Bucaioni[2], and Francesco Flammini[1]

[1] School of Innovation, Design and Engineering, Division of Product Realisation, Mälardalens Universitet, Västerås, Sweden
{maryam.zahid,francesco.flammini}@mdu.se
[2] School of Innovation, Design and Engineering, Division of Computer Science and Software Engineering, Mälardalens Universitet, Västerås, Sweden
alessio.bucaioni@mdu.se

Abstract. The latest industrial revolution has introduced autonomous cyber-physical production systems, integrating machine learning into smart manufacturing to optimize production and resource management. However, this integration impacts trustworthiness due to less predictable and explainable behaviors. This paper presents a novel model-based methodology for evaluating the trustworthiness of such systems. A study was conducted to explore the potential and limitations of model-based assessment, categorizing limitations into structural, behavioral, and resource-related aspects. The findings highlight inadequate risk identification and assessment of ML components in these systems and the constraints of single modeling approaches. Based on these insights, we propose a new methodology to address these limitations and improve the risk assessment of ML components in autonomous production systems.

Keywords: Model-based risk assessment · Limitations · Machine Learning · ISO/IEC 61508 Standard · ISO 31010 Standard · Autonomous Cyber Physical Production Systems · Autonomous Cyber Physical Manufacturing Systems · Architecture · Software System Behavior · Trustworthiness

1 Introduction

Autonomous Cyber-Physical Production Systems (ACPPS) is a product of having artificial intelligence (AI) based solutions such as the machine learning (ML) algorithms integrated into the cyber layer of the traditional cyber-physical production/manufacturing systems (CPPS/CPMS), or the involvement of Collaborative Robots (Cobots) [6], making it intelligent and independent [23].

A. Bucaioni and F. Flammini—These authors contributed equally to this work.

S. Latifi (Ed.): ITNG 2025, AISC 1463, pp. 583–594, 2025.
https://doi.org/10.1007/978-3-031-89063-5_50

Utilizing ML algorithms, ACPPS autonomously refines its predictive models, optimizing decision-making for parameters, and conducts thorough root cause analysis on detected anomalies [20].

Although attractive, such autonomy in ACPPS also raises concerns regarding its trustworthiness [13]. According to a taxonomy, the word trustworthiness here can be defined as a collection of attributes such as robustness, data governance, lawful, ethics, human agency and oversight, and societal and environmental well-being [14]. The Artificial Intelligence Act (AI-ACT) proposed by the European Union classifies AI-based systems by the defined risks and, therefore, mandates the implementation of trustworthiness-related requirements [32]. The scope of such requirements spans the entire ACPPS where, e.g., robustness is crucial for maintaining the system's overall stability, efficient and safe decision-making requires trust, human agency, and oversight is essential to gain trust among its human operators [33]. Moreover, for a product such as vehicles, medical equipment or even cobots to be marketed, it must undergo an extensive risk assessment process per the International Standardization Organization (ISO) standards. Model-based risk assessment (MBRA) techniques have been preferred for detection, and accurate and comprehensive assessment of risks at an early stage of the development life-cycle [29]. However, existing techniques focus on general/traditional CPS or evaluate specific risks related to single trustworthiness attributes [35]. Moreover, since ACPPS is a new concept introduced, no studies have been conducted on assessing trustworthiness-related risks, especially the ones originating from the dynamic behavior of the ML component of ACPPS.

As a main contribution, this paper presents an overview of the proposed risk assessment methodology titled *"Model-based Assessment for Trustworthiness of Industrial Cyber-physical Systems"* (MATrICS), aiming to identify and evaluate trustworthiness-related risks originating from the dynamic behavior of ML component in ACPPS [25]. To develop the new methodology, we conducted a preliminary literature review of the existing MBRA techniques used to assess trustworthiness-related risks in ACPPS and their limitations. The limitation identified became the basis for the developed methodology.

The rest of the paper is structured as follows. Section 2 presents a preliminary review of the existing state-of-the-art MBRA techniques and their limitations, and Sect. 2.1 discusses the adopted research process. Section 3 presents the new methodology proposed to overcome the limitations highlighted in the preliminary review. Finally, Sect. 4 summarizes the study's findings and presents plans for future work.

2 Preliminary Review

This section discusses the review process and presents the findings of a preliminary literature review on the model-based techniques used to assess the trustworthiness-related risks in ACPPS and their reported limitations. To have a discussion, the techniques studied were categorized as shown below for presentation purposes and, therefore, are not semantic in any way.

2.1 Preliminary Review Process

The MATrICS methodology was developed using a cyclic research process following Crnkovic's guidelines [7].

It began with a systematic literature review of MBRA techniques for assessing trustworthiness-related risks in ACPPS, aiming to identify their limitations concerning ML components. The preliminary studies followed Kitchenham et al.'s guidelines [19], using a search string with five parts to include all possible term synonyms and ensure comprehensive coverage, as follows: *("Cyber-Physical System" OR CPS) AND ("Manufacturing" OR "Industry 4" OR "Production") OR (CPPS OR CPMS) AND ("Smart" OR "Artificial Intelligent" OR "Self-Sustain" OR "Autonomous") AND ("Safety" OR "Security" OR "Trust" OR "Dependability" OR "Resilience" OR "Robust" OR "Self-Heal" OR "Self-Repair") AND ("Risk") AND ("Model-Based") AND ("Assessment" OR "Analysis").* The search was conducted on four major databases in software engineering and computer science [19], with details and findings available in Sect. 2.1 and the replication package[1]. The initial search yielded 959 articles, but only peer-reviewed, English-language studies focusing on model-based risk techniques for trustworthiness in ACPPS ML components were included. Studies addressing MBRA limitations were prioritized, while tertiary studies were excluded, resulting in a final set of 44 relevant articles. A data extraction form was created to address the research objective. Extracted data was analyzed, stored in the form, and clustered into two groups: MBRA techniques (Sect. 2.2) and their limitations (Sect. 2.3). Relevant information not initially captured was reviewed and, if necessary, added to the form after re-analysis against the updated requirements.

During the problem formulation step, the literature review data was analyzed and synthesized following Crtuzes et al.'s guidelines [8]. Initially, we identified techniques for assessing trustworthiness-related risks in ACPPS, evaluating their coverage, benefits, and challenges. Vertical and orthogonal analyses were conducted to understand the relationship between the techniques and their characteristics. ACPPS, like traditional CPPS, includes three layers: cyber, network, and physical. However, ACPPS differs due to the ML component in its cyber layer. Findings revealed vulnerabilities from integrating heterogeneous components without proper configuration and risk mitigation, particularly with software patches. Behavioral and structural models were commonly used for risk assessment. Further analysis highlighted that while ACPPS systems rely on intelligent components, risks associated with the ML component and its integration with other ACPPS elements are largely overlooked. These insights informed the formulated problem: *Existing solutions proposed lacked focus on the trustworthiness-related risks, possibly resulting due to the dynamic nature of the machine learning component of ACPPS.*

In the solution proposal step, an iterative approach was used to develop the solution, starting with defining requirements to address the formulated problem. This involved further analysis and a survey of ISO recommendations for risk management in ACPPS, ensuring the proposed solution aligned with these standards. Due to space constraints, process details are omitted. Using the survey data as a foundation, existing techniques were re-examined to identify those

[1] https://github.com/PhDResearcher/ACM2025.

meeting the defined requirements and ISO standards. Based on this analysis, the proposed solution, the MATrICS methodology, is detailed in Sect. 3.

2.2 Model-Based Risk Assessment Techniques

Integrating digital twins in production sections has introduced many issues as they are heavily reliant on data and, therefore, prone to attacks. Based on the review results, identified MBRA techniques used can be clustered into two main categories, namely: *quantitative techniques* providing numerically the probability of occurrence and the severity of the identified risk, and *qualitative techniques* highlighting the risks along with their propagation paths, control strategies and seriousness in terms of high, medium, and low [37]. Among the most reported MBRA techniques, Attack Trees, Fault Tree Analysis (FTA), Event Tree Analysis (ETA), Probability Risk Assessment (PRA), stochastic modeling, Graphs, Decision Trees and Binary Trees, and Bayesian Networks (BN) were the quantitative techniques. Similarly, most reported qualitative techniques included the bow-tie Diagram (BTD). BTD, HAZOP, Petri Nets, and Markov Chain Model (MCM) [36].

In addition, model-checking techniques have been reported for verifying the ML models in ACPPS, ensuring they meet safety and performance requirements [11]. For instance, a model-checking approach was applied using an ML-based fault-diagnostic tool to verify reliability requirements in an automated hydraulic press [34]. Unlike MBRA techniques, which focus on identifying and managing risks [16], model-checking is used for formal system verification [12].

2.3 Limitations of Existing MBRA Techniques

Limitations reported against the proposed MBRA techniques can be categorized into three main types (see Fig. 1).

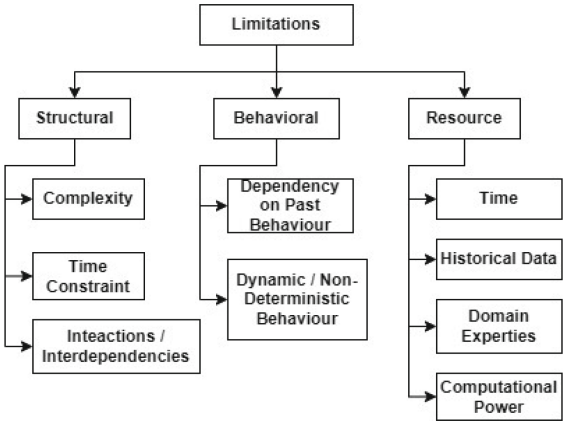

Fig. 1. Identified limitations of MBRA techniques

Structural limitations is associated with the ACPPS's architecture, including complex models with tightly coupled components, numerous interdependencies, and resource-intensive requirements. According to the findings obtained, such limitations are mainly due to the networked nature of ACPPS, where MBRA models fail to capture the tight interactions between the components within a single layer or between multiple layers of ACPPS [10]. This is evident in probabilistic models used for risk assessment, often overlooking real-time requirements and the heterogeneous nature of ACPPS, leading to incomplete risk assessments [11].

ML components' dynamic and periodically evolving nature poses additional challenges for the existing MBRA techniques. In the case of ML components within ACPPS, the traditional MBRA techniques may not account for continuous learning and adaptation capabilities, leading to risk evaluation gaps such as overlooking data quality risks, model bias, and adversarial attacks [22].

Scalability is another significant limitation. The increase in system size leads to an increase in the complexity of the probabilistic model, as evident with the use of probabilistic models. To manage this, studies often focus on a single MBRA technique or apply it to a specific layer of the ACPPS rather than the entire system [28].

Behavioral limitations are tied to the dynamic nature of systems, such as complex behaviors or dependency on past actions [9]. Influenced by concurrency, asynchrony, and distribution, ACPPS operates in variable environments [33]. The reliance of ACPPS on real-time adaptation, data processing, high-speed communication, autonomous decision-making, decentralized control, scalability, flexibility, and robust fault tolerance complicates implementing these requirements [22]. Failing to adhere to such dynamic behavior is why most MBRA techniques, like FTAs, ETAs, BN, Petri Nets, and HAZOP, provide incomplete or incorrect risk assessment of ML models in ACPPS. This can further result in poor decisions, affecting production. Complex behaviors, such as accident sequences or resilience requirements, complicate risk assessment model construction [1]. Studies often overlook device dependencies and complex network topologies in ACPPS [9]. Techniques treating cyber and physical layers separately fail to model risk propagation accurately. MBRAs are generally unsuitable for dynamic risks in ACPPS [15]. As systems grow, techniques like PNs struggle with concurrent processing [3]. Dependence on past behavior documentation limits current MBRA techniques' accuracy [27]. Assessment of Trustworthiness-related risks is even more complex due to rapid changes in the trust metric, complex interdependencies, lack of real-time modeling, overlooked interactions, scalability issues, and reliance on historical data. These limitations highlight the need for advanced methodologies to better address dynamic, interdependent, and real-time trustworthiness risks in ACPPS.

Resource-related limitations refer to the need for high computational power, extensive historical data, and specialized expertise [4]. MBRA techniques such as BN, FTA, ETA, and MCM's dependency on domain-specific historical data for model construction and risk calculations can be challenging and expen-

sive due to privacy concerns and system size [38]. Real-time risk assessment demands significant computational power, often unavailable [27]. For example, continuously monitoring nodes to identify malicious activity increases complexity and requires high computational power and memory [2]. MBRA techniques (e.g., FTA, HAZOP, PRA, ETA) need deep domain knowledge, adding to development and maintenance costs [30]. Domain-specific data may include complete system source code, requiring domain experts [5].

Evaluating trustworthiness-related risks of ML components within ACPPS adds further limitations as it requires extensive datasets for training and validation, which can be challenging to compile. Moreover, training complex models requires substantial computational power, often needing specialized hardware like GPUs or TPUs, leading to limitations associated with the availability of adequate resources. Developing, training, and validating these models requires highly specialized expertise, adding to the overall cost and complexity [24].

2.4 Mapping of Limitations Onto MBRA Techniques for ACCPS

Mapping of the identified MBRA techniques and the constraints limiting their applicability can be seen in Table 1. Looking into the limitations of the model-based techniques, studies have reported overlapping of these limitations such that some techniques showed structural and behavioral limitations. At the same time, some had structural and resource-related limitations or, in other cases, had reported limitations falling into all three of these categories. Although Stochastic models allow the identification of critical vulnerabilities by graphically modeling the system behavior, it is assumed that the system is secure from the start, leading to possible underestimation of risks and misallocation of resources. As for Model-checking techniques, they are more prone to all three types of limitations due to the possibility of a state explosion or a scenario explosion in autonomous systems, making it challenging to handle concurrency or other complex interactions within ACPPS [26].

Table 1. Limitations of Model-Based Assessment Techniques

Type of Limitations	MBRA Technique
Structural	BTD, FTA, ETA, BN, PNs, Graphs, Trees, Stochastic Modelling
Resource Consumption	BTD, FTA, PRA, ETA, BN, HAZOP, MCM
Behavioral	BTD, Stochastic Modelling, FTA, ETA, HAZOP, BN, MCM, HRD, safety use cases, models constructed using OUM, Component Diagram

Current studies focus more on limited system areas, lacking discussions on probable scalability and reliability-related issues of the applied MBRA techniques. A significant gap identified is the lack of focus on trustworthiness-related

risks possibly emerging from the evolving nature of ML components within ACPPS and how they can impact the overall system. Moreover, there exists a lack of information on the mitigation measures to counter risks associated with data availability, interpretability, model simplification, expert reliance, and computational costs.

3 The MATrICS Methodology

This section presents an overview of the proposed MATrICS methodology and how it implements the findings of the review.

Although proven helpful in assessing risks, MBRA techniques such as Fault-/Attack Trees, variants of HAZOP, and Markov Models have limitations with multiple layers of ACPPS, ML components, and various aspects of trustworthiness. Researchers, therefore, often focus on a specific set of trustworthiness-related attributes (e.g., safety and security) on a single layer of ACPPS, mainly

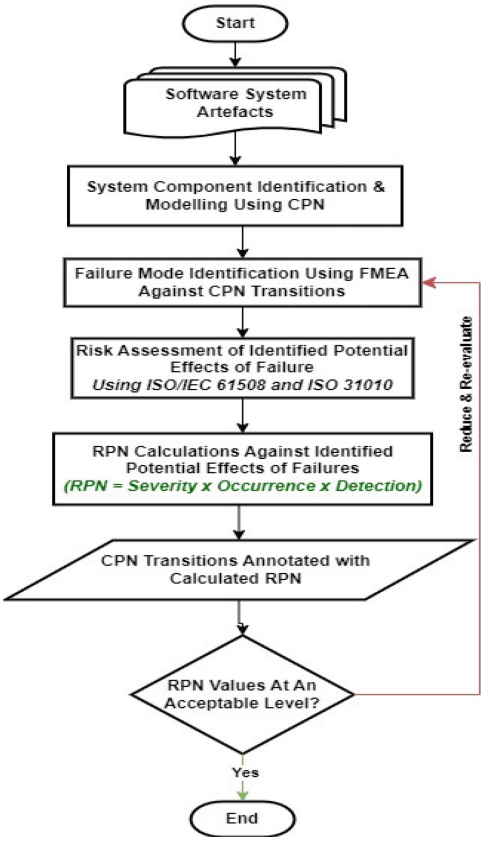

Fig. 2. The MATrICS methodology

the network layer. In the case of the cyber layer, the risk assessment has been neglecting ML-related risks like opacity and bias, which require a multi-faceted evaluation. Here, a hybrid MBRA technique can help capture system complexities, provide a more comprehensive risk assessment, and optimize resources.

The proposed hybrid MBRA methodology, named MATrICS (see Fig. 2), uses a combination of Colored Petri Nets (CPN) with Failure Model and Effect Analysis (FMEA) for better trustworthiness-related risk evaluation of ML components within ACPPS. MATrICS, when applied incrementally, makes the risk evaluation process more suitable and scalable for complex systems. Here, CPN's dynamic modeling capabilities can help capture the concurrent, asynchronous, distributed, non-deterministic, and stochastic behavior of ML components in ACPPS [18]. At the same time, FMEA applied to generated CPN can systematically identify and address multiple failure modes [21], offering a comprehensive, systematic, and practical framework for assessing trustworthiness-related risks in ACPPS. Moreover, the outcomes of mitigation measures can be observed by looping the proposed risk assessment technique.

3.1 MATrICS Requirements

The requirements for the proposed MATrICS methodology were driven based on the findings obtained from the review (as seen in Table 2). CPNs dynamically illustrate the interactions between components and how a failure propagates through the system, providing a clear structural representation of the system. Combining the FMEA analysis with the CPN's graphical depiction can allow for a better understanding of how the structural changes affect the system and help identify the critical areas for improvement. CPNs can model the flow of processes, highlighting possible bottlenecks or inefficiencies. At the same time, a targeted FMEA analysis identifies the potential failure points within the constructed model, improving the overall accuracy of the risk assessment process and optimizing the resource allocation. Moreover, CPNs simulating scenarios can help visualize the impact of identified failure modes, offering a more com-

Table 2. Driven Requirements for MATrICS Methodology

Type of Requirement	Requirement
Structural	Risk Assessment at Design Stage
Structural	Coverage of Tight Interactions or Inter-Dependencies Between Components
Behavioral	Coverage of System's Dynamic Behavior
Behavioral	Coverage of System's Asynchronous, Concurrent, and Dynamic Requirements
Resources	Optimal Resource Allocation
Resources	Identification of Bottlenecks or Inefficiencies

prehensive view of the system's performance and the impact risks can have on it.

3.2 MATrICS Steps

MATrICS methodology is composed of the following seven steps (see Fig. 2):

1. Take as an input the available software system design artifacts to construct a CPN.
2. For each Transition in CPN, conduct a Process-focused risk assessment using FMEA, and the ISO/IEC 61508 standard [31] and ISO 31010 standard [17].
3. Calculate Risk Priority Number (RPN) for each of the identified and assessed potential effects of failure (see Eq. 1).
4. Annotate transitions in CPN associated with relevant failure modes using the calculated RPN values.
5. Check if the calculated RPN value for a particular CPN transition is acceptable.
   - If acceptable, end execution of the methodology.
   - If unacceptable, implement required mitigation measures to reduce the severity of the identified risks and repeat the process starting from step 2.

Here, step 1 is responsible for mapping the components of the system under evaluation, and steps 2–5 evaluate the system for possible risks, their severity, possibility of occurrence, and probability of detection.

$$RPN = \text{Severity} \times \text{Occurrence} \times \text{Detection} \tag{1}$$

Here, RPN is a product of three factors associated with the identified risk [21]: the seriousness of the consequences if the failure occurs *(Severity)*, the likelihood of a failure to occur *(Occurance)* and probability of the failure being identified before the occurrence of the resultant problem *(Detection)*.

Values for RPN's above-stated factors are ranked on a scale of *1–10* with '1' being the lowest and '10' being the highest. Overall, the value of a calculated RPN can range between *1–1000* [21]. Once the RPN values are calculated against each failure mode, they are summed up based on the failure modes involved against each transition of the CPN. The higher the accumulated RPN (ARPN) value, the higher the priority of the studied risk to be mitigated.

Moreover, the methodology allows for evaluating the effects an applied risk mitigation measure can have on the system via the implementation of step 7. Since a threshold RPN value generally does not exist, the risk management team will decide the threshold RPN value [21]. However, the scope of this study is limited only to assessing the identified risks.

Considering the ML components, the proposed methodology aims to provide a structured and systematic risk assessment methodology addressing all

trustworthiness-relevant aspects of ACPPS. It supports a detailed understanding of ML components' dynamic and adaptive aspects and their failure modes. This enables prioritization and mitigation of trustworthiness-related risks during ACPPS development and validation.

4 Conclusion and Future Work

Considering the new risks and complexity introduced by ML components, we moved further toward model-based assessment of trustworthiness in ACPPS in this paper. We developed an abstract of the novel hybrid methodology combining several techniques. Based on the results of the preliminary studies, we found that stochastic modeling formalisms, such as extensions of Petri Nets and Fault Trees, are the most reported techniques for risk assessment, while metrics such as MTTF and intrusion probabilities are the most commonly used for evaluating and prioritizing the identified risks. Model-checking has also been reported as a means to evaluate trustworthiness in ACPPS. However, none of the existing studies addresses the application of those techniques on the ML component of ACPPS. The reported limitations of the existing techniques can be categorized as structural, behavioral, and resource-related. The most reported limitations were scalability-related issues, failure to capture the system's dynamic behavior, and a lack of resources such as documented data, domain experts, and hardware capable of handling complex computations. Based on the current limitations of model-based techniques and tools in assessing the risks associated with the ML components of ACPPS, our methodology – named MATrICS – aims to comprehensively and systematically evaluate relevant trustworthiness attributes. The MATrICS methodology is hybrid since it combines Colored Petri Nets to ensure high expressive power to model any dynamic and adaptive behavior due to ML and FMEA for modular, iterative, and incremental bottom-up modeling to ensure scalability to complex (i.e., large, distributed, and heterogeneous) production systems.

In the future, we plan to further develop the MATrICS methodology and empirically evaluate its effectiveness and performance regarding accurate ACPPS risk identification.

Acknowledgements. This work has been partially funded by the European HORIZON-KDT-JU research project MATISSE "Model-based engineering of Digital Twins for early verification and validation of Industrial Systems", HORIZON-KDT-JU-2023-2-RIA, Proposal number: 101140216-2, KDT232RIA_00017, the Swedish Knowledge Foundation through the MoDEV project (20200234), and the Sweden's innovation agency Vinnova through the project iSecure (202301899).

References

1. Evrin, E., et al.: Risk assessment and analysis methods: qualitative and quantitative. ISACA J. **28** (2021)

2. Ambore, B., et al.: Novel model for boosting security strength and energy efficiency in Internet-of-Things using multi-staged game. Int. J. Electri. Comput. Eng. (2088-8708) **9**(5) (2019)
3. Berger, S., et al.: Modelling availability risks of it threats in smart factory networks–a modular petri net approach. In: European Conference on Information Systems (2021)
4. Burkart, N., Huber, M.F.: A survey on the explainability of supervised machine learning. J. Artif. Intell. Res. **70**, 245–317 (2021)
5. Castellanos, J.H., et al.: Finding dependencies between cyber-physical domains for security testing of industrial control systems. In: Proceedings of the 34th Annual Computer Security Applications Conference, pp. 582–594 (2018)
6. Čelebić, V., Bucaioni, A.: A systematic mapping study on the role of software engineering in enabling society 5.0. In: 2023 IEEE International Smart Cities Conference (ISC2), pp. 1–8. IEEE (2023)
7. Crnkovic, G.D.: Constructive research and info-computational knowledge generation. In: Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery, pp. 359–380. Springer (2010)
8. Cruzes, D.S., Dyba, T.: Recommended steps for thematic synthesis in software engineering. In: 2011 International Symposium on Empirical Software Engineering and Measurement, pp. 275–284 (2011)
9. Darandale, S., Mehta, R.: Risk assessment and management using machine learning approaches. In: 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 663–667 (2022)
10. Deng, S., et al.: Security risk assessment of cyber physical power system based on rough set and gene expression programming. IEEE/CAA J. Autom. Sin. **2**(4), 431–439 (2015)
11. DeSmit, Z., et al.: An approach to cyber-physical vulnerability assessment for intelligent manufacturing systems. J. Manuf. Syst. **43**, 339–351 (2017)
12. Embedded.com: An introduction to model checking (2023). https://www.embedded.com/an-introduction-to-model-checking
13. Euro News: The enemy is false information': world leaders and businesses take on cybersecurity in riyadh (2023). https://www.euronews.com/2023/11/09/
14. Flammini, F., Alcaraz, C., Bellin, E., Marrone, S., Lopez, J., Bondavalli, A.: Towards trustworthy autonomous systems: a survey of taxonomies and future perspectives. J. Artif. Intell. Res. **70**, 245–317 (2021)
15. Geng, S., et al.: Web application architecture security evaluation method based on AADL. In: 2015 20th International Conference on Engineering of Complex Computer Systems (ICECCS), pp. 186–189 (2015)
16. Gran, B.A., Fredriksen, R., Thunem, A.P.J.: An approach for model-based risk assessment. In: International Conference on Computer Safety, Reliability, and Security, pp. 311–324 (2004)
17. ISO International Organization for Standardization: ISO 31010 - risk management (2020). https://practicalrisktraining.com/iso31010
18. Jitmit, C., Vatanawood, W.: Simulating artificial neural network using hierarchical coloured petri nets. In: Proceedings of the 2021 6th International Conference on Machine Learning Technologies, pp. 127–131 (2021)
19. Kitchenham, B., Brereton, P.: A systematic review of systematic review process research in software engineering. Inf. Softw. Technol. **55**(12), 2049–2075 (2013)
20. Lazaroiu, G., Androniceanu, A., Grecu, I., Grecu, G., Neguriță, O.: Artificial intelligence-based decision-making algorithms, Internet of Things sensing networks,

and sustainable cyber-physical management systems in big data-driven cognitive manufacturing. Oeconomia Copernicana **13**(4), 1047–1080 (2022)
21. Li, J., Chignell, M.: FMEA-AI: AI fairness impact assessment using failure mode and effects analysis. AI Ethics **2**(4), 837–850 (2022)
22. Li, X., Kang, K., Hu, S.: Real-time requirements for cyber-physical systems: a survey. IEEE Trans. Industr. Inf. **13**(6), 2817–2826 (2017)
23. Lichte, D., Wolf, K.D.: Use case-based consideration of safety and security in cyber physical production systems applied to a collaborative robot system. In: Safety and Reliability–Safe Societies in a Changing World, pp. 1395–1401. CRC Press (2018)
24. Mattioli, J., et al.: An overview of key trustworthiness attributes and KPIs for trusted ML-based systems engineering. AI Ethics **4**(1), 15–25 (2024)
25. Nazir, R., Bucaioni, A., Pelliccione, P.: Architecting ml-enabled systems: challenges, best practices, and design decisions. J. Syst. Softw. **207**, 111860 (2024)
26. Pappagallo, A., et al.: Monte Carlo based statistical model checking of cyber-physical systems: a review. Information **11**(12), 588 (2020)
27. Patel, A.R., Liggesmeyer, P.: Machine learning based dynamic risk assessment for autonomous vehicles. In: 2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC), pp. 73–77 (2021)
28. Peng, Y., et al.: Cyber-physical system risk assessment. In: 2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 442–447 (2013)
29. Rathore, L.K., Sao, N.: An integrated model based test case prioritization using UML sequence and activity diagram. Int. J. Res. Comput. Appl. Robot. **3**(12), 31–41 (2015)
30. Tam, K., Jones, K.: Cyber-risk assessment for autonomous ships. In: 2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), pp. 1–8 (2018)
31. TÜV SÜD: Functional safety - IEC 61508. https://www.tuvsud.com/en-us/services/functional-safety/iec-61508. Accessed Apr 2024
32. Union, E.: The enemy is false information': world leaders and businesses take on cybersecurity in riyadh (2023). https://www.weforum.org/agenda/2023/06/european-union-ai-act-explained/
33. Wan, J., Yi, M., Li, D., Zhang, C., Wang, S.: Data-driven and autonomous manufacturing control in cyber-physical production systems. J. Manuf. Syst. **54**, 97–107 (2020)
34. Yang, R., Zhong, M.: Machine Learning-Based Fault Diagnosis for Industrial Engineering Systems. CRC Press (2022)
35. Zahid, M., Bucaioni, A., Flammini, F.: Trustworthiness-related risks in autonomous cyber-physical production systems-a survey. In: 2023 IEEE International Conference on Cyber Security and Resilience (CSR), pp. 440–445. IEEE (2023)
36. Zahid, M., Bucaioni, A., Flammini, F.: Model based trustworthiness evaluation of autonomous cyber-physical production systems: a systematic mapping study. ACM Comput. Surv. (2024)
37. Zahid, M., Inayat, I., Deneva, M.: Security risks in cyber physical systems-a systematic mapping study. J. Softw. Evol. Process **33**(9), e2346 (2021)
38. Zografopoulos, I., et al.: Cyber-physical energy systems security: threat modeling, risk assessment, resources, metrics, and case studies. IEEE Access **9**, 29775–29818 (2021)