A comprehensive approach to safety for highly automated off-road machinery under Regulation 2023/1230

Marea de Koning a,∗, Tyrone Machado a,b, Andrei Ahonen a, Nataliya Strokina c, Morteza Dianatfar a, Francesco De Rosa d, Tatiana Minav a, Reza Ghabcheloo a

a Faculty of Engineering and Natural Sciences, Tampere University, Korkeakoulunkatu 6 33720 Tampere, Finland
b Department of Engineering Systems and Software, Bosch Rexroth AG, Gewerbegebiet Glockerau, Glockeraustraße 2 89275 Elchingen, Germany
c Faculty of Information Technology and Communication, Tampere University, Tampere, Finland
d Department of Functional Safety, ART S.P.A., Turin, Italy

ARTICLE INFO

Keywords:
Software safety
Machinery
Machinery Regulation
Standardization
Decision making
Safe reasoning
Risk management

ABSTRACT

As of January 2027, Machine Safety Regulation 2023/1230 replaces Directive 2006/42/EC, and a proposal for the regulation of Artificial Intelligence (AI) systems is soon to be ratified by the European Council. These legislative changes address hazards from increased automation and AI integration. Original Equipment Manufacturers (OEMs) face compliance challenges due to a lack of standardized guidelines, especially for off-road machinery with level 4 ‘high automation’ capabilities.

At level 4 ‘high automation’, the machine operator assumes a supervisory role, intervening only in situations beyond the designed operational domain. Relying solely on supervisor intervention at this level of automation is unreasonable. Within this context, legislative safety requirements to maintain safety advocate for robust risk management systems, including both a human-in-the-loop safety option and adequate safety-related decision-making by the level 4 ‘high automation’ off-road machine, i.e., the safe reasoning software becomes paramount.

In this research, conducted through a systematic mapping study, we discuss emerging safety trends and integration for the safe reasoning software a level 4 off-road machine requires. Our analyses shows the necessity of safety constraint-based design procedures. These procedures involve constraining the AI system by embedding fault management, reasoning checks, and diagnostic monitoring together with safety-related decision-making, thereby enhancing functional reliability and reducing risks. We consider these trends and patterns through the lens of changing machine safety legislative requirements, and the proposal of the regulation of AI products. Illustrating their implications alongside a case study of a level 4 tree harvesting machine, for a tangible and to industry transferable perspective on our results.

1. Introduction

The European Union (EU) is a world leader in the machinery manufacturing industry, achieving an impressive annual turnover of €740 billion (Gospodinova, 2022). Contributing to its position, the EU guarantees that any machinery bearing the EC declaration of conformity offers high levels of protection and safety for EU workers and citizens (European Parliament, 2022). Nevertheless, a total of 3347 work-related fatalities were recorded in 2021. Nearly two-thirds of these cases occurred in industries where humans interact with off-road machinery (Anon, 2023a). These staggering figures are related to sectors utilizing off-road machinery currently in compliance with essential health and safety requirements (EHSR) established in Directive 2006/42/EC (Anon, 2016a), such as mining, construction, agriculture, and forestry.

Facilitated by the integration of software and AI systems, safety and security risks are expected to exceed the scope of current machine safety requirements (European Commission et al., 2020). Addressing the need for enhanced safety measures outlined in European Commission et al. (2020), a new Machinery Regulation 2023/1230 (Anon, 2016b) becomes mandatory from January 20th, 2027, and a proposal for the regulation of Artificial Intelligence (AI) systems (Anon, 2025) is expected to soon be ratified by the EU Commission. However, standardized guidelines for implementing off-road machine safety requirements

∗ Corresponding author.
E-mail address: marea.dekoning@tuni.fi (M. de Koning).

https://doi.org/10.1016/j.ssci.2024.106517
Received 20 November 2023; Received in revised form 15 March 2024; Accepted 19 March 2024
0925-7535/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Safety, i.e., the freedom of unacceptable risk (Anon, 2010), is typically achieved through proper training of a machine operator who must ensure adequate responses to hazards which emerge at run time during task-execution, paired with the reliable availability of safety functions embedded via the safety-related parts of the control systems (SRP/CS) (Anon, 2023b). The operator understands the capabilities, limitations, and maintenance requirements of the machine and this minimizes the risks of misuse or improper handling, which could result in harm, damage or breakdowns. This, in turn, increases availability, productivity, and safety in the workplace (Anon, 2023c).

SAE J3016 is a widely accepted standard for defining a taxonomy detailing various degrees of driving automation that can exist in automotive vehicles. While, in part, machinery can be matched to this taxonomy, it does not cover the full range of tasks that a machine executes, like the ability to manipulate its environment (e.g., fell a tree, dig a trench, move dirt). For this purpose, we do not reference SAE J3016 instead we follow the definition of degrees of automation as proposed by the taxonomy discussed in Machado et al. (2021). Here, automating a machine up to level 4, in terms of manipulation and/or driving abilities, means that the operator is no longer controlling the machine, and as such adequate risk reduction requires further actions. While, at level 3 ‘conditional automation’, the operator must confirm the safety of each task before execution. At level 4, the role of a supervisor, either remote or in the cabin, intervenes only in situations beyond the scope of the designed operational domain. It is unreasonable to expect an operator to intervene promptly when the operator does not have full control over the highly automated off-road machine (Wickens et al., 2021). Consequently, effective risk management in higher levels of automation requires the off-road machine itself to be capable of initiating the correct safety function to maintain a safe state. A safe state, means sustaining the freedom of unacceptable risk, i.e., safety (Anon, 2010). Preferably, achieving this safe state at level 4 occurs without operator intervention, effectively moving towards removing the operator’s role in the safety chain. (Fig. 1).

Anticipating all potential failure scenarios for highly automated machinery is practically impossible, posing new safety compliance challenges and necessitating a shift in standardized safety practices (Swuste et al., 2020). The previous focus was on ensuring the absence of failure modes through verified and validated design methods. The current emphasis lies in ensuring the capacity to safely handle unforeseen events and safety-related decision making processes which initiate safety-related control output (Rasmussen, 1997).

Legislative safety requirements for highly automated machinery acknowledge this challenge and advocate for robust risk management systems. However, there is a significant absence of adequate guidance for the development of such risk management systems, thereby limiting compliance with Machinery Regulation 2023/1230 and the proposed regulation for AI. This reveals a significant knowledge gap for OEMs, involving the need to ensure safety at higher automation levels and meeting legislative requirements for off-road machinery market placement in the EEA.

1.2. Research questions

In this research, our focus is on establishing the requirements and proposing a design which integrates adequate safety-related decision making logic within a risk management system. Meaning, the Safe reasoning software that provides the cognitive process or decision-making capability that the machinery uses to evaluate data, assess its surroundings, and send safety-related control output, with a central emphasis on ensuring safety at run-time. For this purpose, we centre our search around two main questions:

- RQ1: What are emerging safe reasoning software requirements?
- RQ2: How are safe reasoning elements embedded in a safety system?

We approach this via a Systematic Mapping Study according to Petersen et al. (2015), revisiting the scope defined in our previous study (Aimee M.R. De Koning et al., 2023). This systematic mapping study (SMS) reviews and categorizes emerging safety requirements and integration methods in an effort to identify a common approach to safety in highly automated systems related to off-road machinery. Consequently, addressing the lack of standardized safety guidance for level 4 off-road machinery.

1.3. Structure

This article is organized as follows: Section 2 concretely defines the goal of this article. In Section 3 we discuss the application of our SMS, the results of which are presented in Section 4. We consider these trends and patterns through the lens of changing machine safety legislative requirements, the proposal of the regulation of AI products and illustrate their implications alongside a case study of a highly automated tree harvesting machine in Section 5. This article is concluded and future work is discussed in Section 6.

2. Objective

In this research, conducted through a SMS, a method described in Petersen et al. (2015), we discuss safety trends for embedding AI systems and recently published architectural design patterns relevant for level 4 off-road machinery. After which, we consider these trends and patterns through the lens of changing machine safety legislative requirements, the proposal of the regulation of AI products and illustrate their implications alongside a case study of a highly automated tree harvesting machine. Providing the readers of this manuscript with a more tangible perspective on our results, which can be transferable to other level 4 off-road machinery.

3. Method

A SMS consists of three phases: planning, data collection, and coding, followed by a documentation and discussion of results following the guidelines described in Petersen et al. (2015). During the planning phase, we revisited the scope defined in our previous study (Aimee M.R. De Koning et al., 2023), determined how best to build upon those results and address any limitations identified in our previous iteration. Limitations in our previous study are that it does not evaluate the
Decision making is crucial for ensuring safety at level 4 off-road machinery. It involves minimizing faults and ensuring the reliability of safety-related decision-making at runtime. Formal verification involves validating these safety trends to strengthen the safety case. Additionally, our analysis uncovered novel architectures relevant for level 4 off-road machinery. However, the impact of evolving machine safety regulations, such as Machinery Regulation 2023/1230 (Anon, 2016b) and proposed AI product regulation (Anon, 2025), remains largely unaddressed in analysed references.

4.1. Safety requirements

Decision making Our analysis shows that when the references are evaluated alongside the scope of level 4 off-road machinery, real-time dynamic safety-related-decision making during task execution (Table 2 section “Safety Requirements”), becomes one of the most prominent trends. This decision making does not only account for the capacity of the machine to assess task safety at run time through adequate awareness and reasoning capabilities, but also of arguing the capacity of a supervisor to intervene. While our analysis reveals a gap in the safety assessment of a supervisor’s intervention capacity, this could be considered a limitation stemming from the design of our SMS. Regardless, it presents an interesting future research direction, such as uncovering a formal abstraction concerning various degrees of shared safety-related decision-making, allocation of responsibilities, and minimum requirements for situational awareness and monitoring of human engagement at level 4 off-road machinery.

Safety related decision making embedded within level 4 off-road machinery can be separated in two components. One being the degree of situational awareness that is achieved and the other being the cognitive capacity of the machine to reason on the given input. Typically this is integrated with AI systems enhancing real-time adaptable behaviour. However, challenges with the reliability and trustworthiness of these systems remain, especially when they are applied for risk reduction purposes. AI systems are known to have weaknesses such as false positives/false negatives, unpredictability, the inability to manage unforeseen scenarios, and a high computational strain especially during retraining phases. Uncertainties, both in terms of the automated systems internal state and its external environment, can lead to an increased risk of safety-critical situations. Lastly, while the use of state-of-the-art sensors allows for more efficient data processing at higher speeds, safety considerations often require a reduction in operational speeds due to computational limitations.

Fault management When it comes to fault management (Table 2 section “Faults”), most studies focused on mobile autonomous vehicles. Confirming that when humans share space with highly automated systems, fault identification and management remains critical. A main trend throughout these studies is ensuring fault tolerance through structured methodologies such as cognitive architectures (Dreany and Roncare, 2019), hierarchical designs (Favier et al., 2020), and task-oriented approaches (Rovira-Más, 2010). Modular architectures, and their inherent flexibility, reactivity, and adaptability, remain popular for efficient fault tolerant systems (Ionescu, 2021; Adam et al., 2016). Additionally, fault detection, diagnosis, and recovery remains the foundation of any resilient highly automated system, especially when operating in dynamic and unpredictable environments. We did identify some remaining gaps and challenges, for example, it is evident that challenges remain with algorithm predictability, particularly in managing unforeseen changes, especially in dynamic environments. However, Rovira-Más (2010) suggests there is potential in enhancing accuracy through statistical identification methods and developing architectures specifically adapted to cope with them.

Formal verification Any OEM is required to obtain a CE marking confirming compliance with regulatory requirements before marketing their machinery product within the EEA. Standardized guidelines such
as those referenced in Appendix A are non-mandatory recommendations stemming from generalizations made by industry guiding OEMs in compliance with regulatory requirements. The linkage between academic works in off-road machinery and its transferability into practise shows in the analysis of our references (Table 2 section “Compliance”) that particularly address compliance efforts with existing standards.

The safety consequences of transitioning to level 4, where the machine assumes primary responsibility for safety and the operator shifts to a supervisory role with intervention only when necessary, exceeds the scope of current standardized guidelines, as is detailed in Appendix A. However, OEMs can present a compelling safety case for their off-road machines, demonstrating compliance with regulatory safety requirements to obtain a CE marking under specific circumstances, as discussed in greater detail in Section 4.3.1. Formal verification and validation (V&V) plays a significant role in this argument, showing its crucial importance in arguing a safety case for level 4 off-road machinery. Therefore, while a smaller subset of our analysed references may be because of a limitation in the design of our study, it remains crucial that the topic of Formal V&V is addressed.

Formal verification and validation (V&V) techniques especially for complex automated systems remain necessary in arguing adequate and reliable risk reduction. Progress in formal V&V is made, but it consistently remains a challenging domain (Ingrand, 2019), and the real-world applicability of formal models reveals scalability challenges, especially in multi-agent environments. The most promising studies in formal verification for highly automated systems are those applying specific formal verification tools. For example, Zhang and Zhang (2016) demonstrates the application of the verification tool KeYmaera, showing promising results for the formal verification of their algorithms to satisfy particular safety constraints in dynamic environments. Another used formal verification tool, Isabelle/UTP (Foster et al., 2021), shows promising results for the algorithmic safety constraint based verification of complex automated systems.

4.2. Safety architectures

When a machine lacks safety standards, a risk assessment uncovering hazards is crucial for appropriate risk reduction efforts by OEMs (Chinniah et al., 2017); Appendix B discusses some hazards that emerged from our references applicable to level 4. While this list can provide a starting point, a properly documented risk assessment remains crucial for OEMs to effectively mitigate risks. The authors of Chinniah et al. (2017) discuss hazards related to machinery; in particular, they focus on hazards emerging during the maintenance phase of the machine’s lifecycle. They provide excellent insights into variables for acceptable failure rates, establishing all the necessary safety requirements, and defining safety validation and verification during all phases of the machine’s lifecycle (Chinniah et al., 2017). However, biases in risk estimation tools can lead to flaws in the architecture of the safety control system (Chinniah et al., 2018). The authors of Chinniah et al. (2018) tested different risk assessment tools, uncovering that there is no one-size-fits-all tool, but the importance of selecting the right tool for the job.

The presence of a multitude of hazards requires mitigation through the integration and interaction of multiple safety functions. Placing significant importance on the design of the safety architecture encompassing these functions. Viewing compliance with machine safety legislative requirements for a level 4 off-road machine, a typical approach begins with safety constraint-based design, as discussed by Dreany and Roncace (2019), and expands from there. The authors off (Dreany and Roncace, 2019) apply safety constraint-based design for an unmanned surface vehicle (USV), integrating AI into the architecture and enforcing pre-identified safety constraints. These constraints include mechanisms such as fail-safe designs, self-monitoring, exception handling, and reconfiguration. In Feng et al. (2023b), the focus lies primarily on architectural patterns enabling the utilization of AI in safety-critical applications. Instead of solely reducing the uncertainty of individual elements, these patterns, inspired by literature and standards from the aviation domain, assume structural capabilities to monitor for failures and intervene when necessary.

4.3. Legislative landscape

From January 20th, 2027, onwards, off-road machinery manufacturers must comply with the Essential Health and Safety requirements (EHSR) in Regulation 2023/1230 for their products to be considered safe enough to be marketable within the European Economic Area (EEA) (Anon, 2016b). The definition of off-road machinery is amended as an assembly consisting of linked components, at least one of which moves, and which are joined together for a specific application including off-road machinery that only lacks the upload of software intended for the specific application envisioned by the manufacturer (Anon, 2016b). Furthermore, there is a proposal for the regulation of Artificial Intelligence (AI) (Anon, 2025) that is soon to be ratified by the European Commission. Machinery that embeds AI systems is considered high risk and will have to conform to both Regulation 2023/1230 and the forthcoming proposal of regulation of AI systems (Anon, 2025). The AI proposal defines AI systems as software developed using techniques, including:

- Machine learning approaches: This involves supervised, unsupervised, and reinforcement learning, and methods such as deep learning.
Logic- and knowledge-based approaches: Encompassing knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning, and expert systems.

Statistical approaches: Involving Bayesian estimation, search, and optimization methods.

These AI systems, designed to achieve human-defined objectives, can generate outputs such as content, predictions, recommendations, or decisions, influencing the environments with which they interact. Regulation 2023/1230, refers to behaviour facilitated by AI systems as control systems of off-road machinery or related products with fully or partially self-evolving behaviour or logic that are designed to operate with varying levels of autonomy (Anon, 2016b). In this article we remain in-line with the definition of AI systems as established in the proposal for the regulation of AI products.

4.3.1. A compelling safety case

A compelling off-road machine safety case or argument to obtain a CE marking rests on three core concepts: coverage,1 traceability,2 and explainability3 (Aimée and de Koning, 2023). It can still be possible to establish a compelling safety case/argument without strictly adhering to a standard if an alternative method or approach can prove that the machine meets the Essential Health and Safety requirements (EHSR) in Machinery Regulation 2023/1230 and if AI systems are used the proposal for the regulation of AI products (Anon, 2025). However, to do this, the machinery manufacturer must provide a robust and compelling safety argument that the alternative approach ensures the same level of safety i.e., adheres to an acceptable failure rate.

Finally, it should be noted that in particular cases the involvement of a notified body is required for a declaration of conformity. Regulation 2023/1230 mentions that off-road machinery classified under Appendix A, part A of Anon (2016b) and considered ‘high risk’ is subject to a stricter conformity assessment procedure requiring the participation of a notified body. Additionally, when off-road machinery classified under Appendix A, part B of Anon (2016b) is not developed in accordance with relevant EU harmonized standards then it is subject to a stricter conformity assessment procedure requiring the participation of a notified body (Anon, 2016b). Furthermore, safety components independently placed within the EEA as listed in Appendix B of Anon (2016b), require the participation of a notified body for conformity. This indicative list of safety components has now been amended to include software ensuring safety functions and safety components with fully or partially self-evolving behaviour using machine learning approaches ensuring safety functions (Anon, 2016b).

5. Discussion through a case study

Highly skilled operators of forest machinery perform various tree stem processing tasks. These tasks include potentially hazardous activities such as felling, delimbing, cutting, and measuring tree stems. For our context we consider a level 4 tree harvesting machine see Fig. 2. This means that the operator no longer has full control, and all tree stem processing tasks are executed automatically at the machine’s discretion. Requiring the machine to actively monitor its environment, its state, and respond appropriately to hazards emerging at runtime. Consequently, the human operators role transitions into a supervisory role leading to measurable degradation’s in performance (Wickens et al., 2021), diminishing the reliability of human supervisors’ capacity to intervene when necessary in order to maintain a safe state. Machinery regulation 2023/1230 (Anon, 2016b) and the proposal for the regulating of AI systems (Anon, 2025) advocate for the integration of a risk management system in highly automated machinery, contributing to adequate risk reduction at runtime when it cannot be guaranteed by a supervisor alone. Our case study focuses on the requirements for

---

1 Coverage in the context of conformance to machine safety regulatory requirements refers to the extent to which a risk/hazard analysis covers all reasonably foreseeable safety-critical situations.

2 Traceability in the context of conformance to machine safety regulatory requirements refers to the extent to which a safety requirement can be traced from establishment to integration, to validation and verification of its availability.

3 Explainability in the context of conformance to machine safety regulatory requirements; refers to the extent in which a justifiable and convincing argument can be made towards the reliable availability of safety functions.
safe reasoning software an element of the risk management system addressing one particular hazard: “Erroneous stem-feeding operations”. Safe reasoning software is an aspect of the risk management system which provides the cognitive process or decision-making capability that the machinery uses to evaluate data, assess its surroundings, and send safety-related control output, with a central emphasis on ensuring safety at run-time.

5.1. Introducing the hazard

In one potential hazardous scenario, the level 4 tree harvesting machine exhibits rapid pulling of tree stems, erroneously feeding them in the wrong direction. In this situation, the stems lack proper alignment and/or guidance, posing a substantial risk of collision with the cabin where the operator is stationed. The operator, unable to intervene in this faulty feeding process, faces the potential for severe injury, and there is a heightened risk of the machine tipping over due to unbalanced loading.

5.2. Risk assessment

A major concern within this context is the potential collision between a tree stem and the operator’s cabin, carrying the severe risk of serious injury or even fatality. Although such collisions are deemed infrequent, occurring in less than 1/20 of the overall operating time, larger stems alone hold the potential to breach the protective cabin structure when the machine is operated as instructed. Erroneous stem feeding may occur, primarily attributed to the high feeding speed of 6 m/s, and a maximum feeding distance of 8 m. The third parameter for risk estimation, the possibility of avoiding the hazard or limiting harm, indicates limited options for the supervisor in the analysed scenarios.

5.3. Risk estimation

Following our risk assessment, guided by a risk estimation graph in ISO EN 13849-1:2023 (Fig. 3), we identify that the overall required performance level is at d (PLd (Anon, 2023b)), equivalent to system integrity level 2 (SIL2 (Anon, 2010)). Based on the parameters identified during our risk assessment where the Severity ‘S’ of the hazard is considered serious (S2), Frequency ‘F’ is seldom (F1), and Probability ‘P’ of avoidance is scarcely possible (P2).

5.4. Risk evaluation and reduction

Our claim of adequate risk reduction in the safe reasoning software faces significant challenges due to the presence of an AI/ML element and the requirement for complex kinematic pose deduction algorithms, both of which are susceptible to hazards leading to failures. Such failures can result in pose estimation faults, consequently elevating the risk of collisions between the stem and the cabin. These hazards may arise from both classical deterministic software elements and AI/ML models, encompassing issues like inaccurate real-time pose estimation, sensor failures, computational bottlenecks, latency, security vulnerabilities, and a lack of redundancy.

Inspired by guidance provided in existing standards and what we learned from our SMS, we propose a design for this safe reasoning software, including AI/ML. We break it down into six distinct elements, as visualized in Fig. 4. These elements work together to manage potential failures, autonomously triggering safety stops or delivering timely alerts when human intervention is still possible. The target is to achieve a risk reduction level up to Pd, as determined by EN ISO 13849-1:2023 (Anon, 2023b).

In EN ISO 13849-1:2023, Categories 2 and 3 mainly differ in their structural configurations and the quality of their hardware components. In Category 2, safety functions are periodically checked, while in Category 3, a redundant setup ensures that a fault does not immediately result in a loss of the safety function. Similar patterns can be observed in aviation safety systems, and the implications of incorporating AI/ML elements are discussed in Fenn et al. (2023a). The most promising design that can be adapted to our context (proposed in Fenn et al., 2023a) is the Run-time Assurance (RTA) pattern, which has proven effective in aviation (Burns et al., 2011). Managing the behaviour of the AI/ML element in the safe reasoning element through an RTA or a ‘wrapper’ is not an entirely new concept and has also been recommended in recent studies, such as those by Mikkonen et al. (2021) and Kuwajima et al. (2020), which discuss safety constraint-based design principles. However, this does mean that the wrapper itself must be able to claim a higher risk reduction than the AI/ML element it encompasses, meaning that all failure conditions of the AI/ML element need to be correctly detected under all reasonably foreseeable operating conditions and artifacts must be in place so that an AI/ML failure does not cause an increase of risks, as advised by Fenn et al. (2023a), Mikkonen et al. (2021), Kuwajima et al. (2020). In our design proposal the elements wrapping the AI/ML are the back-up logic and reasoning output (Fig. 4), supported by diagnostic and monitoring elements such as input monitoring, run-time monitor and a timer (Fig. 4).

5.4.1. Structural design

Here we discuss the elements in Fig. 4 that altogether contribute to overall adequate risk reduction. With respect to the AI/ML element, the authors of Kurzidem et al. (2022), discuss a promising method where the AI/ML model, such as a Deep Neural Network (DNN), represents the generic black-box model. Adapting their methodology to our context requires us to train and optimize a DNN for accurate safe pose estimation at run-time. A data set with labelled safety-relevant features is used for testing the DNN. The predictions, along with their evaluation (including successful and unsuccessful detections), and the safety features are used to train an explainable model. Now, the trained, explainable model could be used to predict the black-box behaviour and, consequently, assess the safety of the baseline DNN. In essence, the white-box model predicts the outcome of the black-box for a reduced input space while also providing an explainable representation.

The Back-up Logic element, is specifically designed to continuously approximate through kinematic algorithms the pose of the harvester head derived from the given joint angles after which the Reasoning Output element assesses if a given pose can be considered safe. Reasoning Output This element is essentially a final check for the correctness of the AI/ML and either send the signal to automatically activate the safety stop or makes a request for Human intervention.

A Run-time monitor element verifies inputs for accuracy and consistency, like joint angles and lengths, to prevent calculations from reaching an invalid solution. In cases where the module cannot ensure trustworthiness, human intervention becomes necessary. Depending on the specific situation and error propagation within the system, insufficient outputs can lead to hazardous behaviour at the system level. To address this, one option is the real-time monitoring of unforeseen data during operation, as discussed in the context of the SOTIF cause and effect model.

A Watchdog, as its name suggests, monitors for bugs or other errors, which may for example trigger an infinite loop or introduce dangerous delay. To prevent this, a timeout mechanism must be implemented capable of sending a signal leading the highly automated machine into a safe state.

Lastly Human intervention to bring the system to a safe state is permitted by EN ISO 13849-1:2023 and further detailed in the standard. EN ISO 13849-1:2023 introduces the concept of an ‘immediate action warning indicator’ that informs the operator (or supervisor, in the case of highly automated machines) of a fault or malfunction, prompting them to take necessary safety actions. This concept emphasizes the interaction between the operator/supervisor and the safety system, which is particularly important in the context of highly automated machines.
5.5. Formal V&V

In our case study, we advocate achieving adequate risk reduction, in-line with Machinery Regulation 2023/1230 (Anon, 2016b) and the AI proposal (Anon, 2025), by integrating a risk management system that includes safe reasoning software encapsulating the AI system. Ensuring the reliability of this safe reasoning element to meet $\text{PL}_d/\text{SIL}2$ requirements necessitates the quantification of a failure rate. This involves evaluating our real-time safe pose estimator and assessing its performance, as discussed in Hutchinson et al. (2022).

This design of a safe reasoning element, incorporating AI/ML (Fig. 4), is designed to meet $\text{PL}_d/\text{SIL}2$ requirements. Achieving this through the combination of a category 2 structure and the integration of additional elements, ensuring a high level of diagnostic coverage for safety-related faults and failures, which is another crucial parameter for $\text{PL}d$ compliance. The use of cross-monitoring mechanisms, such as the Back-up-logic, Watchdog, and a Run-time monitor, enables us to achieve a diagnostic coverage within the medium range (i.e., from 90% to 99%).

6. Conclusion

Starting from January 20th, 2027, any machine marketed within the EEA must conform to the Essential Health and Safety Requirements (EHSR) established in Machinery Regulation 2023/1230. If the machine is considered high risk, meaning it is classified under Appendix A, part A of Regulation 2023/1230, it must also adhere to the proposal for the regulation of AI, soon to be ratified by the EU Commission. Anticipating all potential failure scenarios for level 4 machinery is practically impossible, posing new safety compliance challenges and necessitating a shift in standardized safety practices. At present, Machine safety standards do not adequately address this shift as detailed in Appendix A, revealing a significant knowledge gap for OEMs. This gap exists between developing safe, level 4 off-road machinery and achieving the legislative conformance required for placement in the EEA market. We discuss three prominent safety trends crucial for ensuring safety at level 4 off-road machinery: decision making, fault management, and formal verification. Decision making encompasses the machine’s ability to assess task safety and the supervisor’s intervention capabilities. Fault management focuses on minimizing faults and ensuring the reliability of safety-related decision-making at runtime. Formal verification involves validating these safety trends to strengthen the safety case. It is evident that emerging safety requirements prioritize encapsulating AI systems through safety constraint-based design managing failure rates and ensure safety function availability for risk minimization. This emphasis on safety is reinforced by discussions.
on formal verification and validation of integrated fault management and reasoning capabilities. In this work we continue by considering uncovered trends and patterns through the lens of changing machine safety legislative requirements, the proposal of the regulation of AI products and illustrate their implications alongside a case study of a highly automated tree harvesting machine. A discussion alongside a real-world case study of a level 4 tree harvesting machine provides tangible perspective on our results, which can be transferable to other level 4 off-road machinery. We consider the evolving legislative landscape and exceed the scope set by existing Machine Safety Standards such as IEC 62061:2005 and ISO 13849-1:2023 for assuring safety at level 4 building a foundation for future safety considerations in highly automated off-road machinery.

6.1. Future research

In the next phase of our investigation into compelling safety compliance arguments for highly automated off-road machinery, our primary emphasis will be on the integration, validation, and verification of safe reasoning software for a level 4 tree harvesting machine. This safe reasoning software not only manages the risk of the AI system integrated for a safety purpose it also considers the capacity of a supervisor’s ability to timely intervene, assessing the availability of adequate risk reduction efforts at run-time. Furthermore, legislative safety requirements for level 4 off-road machinery continue to advocate for a robust risk management system, incorporating a human-in-the-loop safety option. Our findings show the absence of any formal abstraction concerning various degrees of shared safety-related decision-making, allocation of responsibilities, minimum requirements for situational awareness and monitoring of human engagement at level 4 off-road machinery. Consequently, we intend to further explore this topic.

CRediT authorship contribution statement

Marea de Koning: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Tyrone Machado: Writing – review & editing, Writing – original draft, Validation, Methodology. Andrei Ahonen: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation. Nataliya Strokina: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation. Morteza Dianatfar: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation. Francesco De Rosa: Writing – review & editing, Validation. Tatiana Minav: Writing – review & editing, Supervision. Reza Ghachbeloo: Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Marea de Koning reports financial support was provided by Tampere University and Epec oy. Andrei Ahonen reports financial support was provided by Tampere University and FIMA Forum for Intelligent Machines ry. Francesco De Rosa reports was provided by ART S.P.A. Tyrone Machado reports was provided by Bosch Rexroth AG & The European Union s Horizon 2020 research and innovation programme.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Scope</th>
<th>Limitations</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO/TR 22100-5:2022 Safety of machinery</td>
<td>Addresses the implications of artificial intelligence and machine learning in machinery in relation to ISO 12100, offers guidance on risk assessment.</td>
<td>Does not adequately cover safety mitigation strategies or encompass AI-based safety systems, including safety-related (smart) sensors and control components.</td>
<td>Anon (2022a)</td>
</tr>
<tr>
<td>EN ISO 13849-1:2023 safety of machinery; Safety-related parts of control systems</td>
<td>Amended with an acknowledgement that AI can be integrated as part of a safety function.</td>
<td>Does not provide additional support or guidelines for machinery manufacturers to demonstrate adequate risk reduction, i.e., traceability &amp; explainability.</td>
<td>Anon (2023b)</td>
</tr>
<tr>
<td>EN 61508:2010 Functional Safety of Electrical/ Electronic/ Programmable Electronic Safety-Related Systems</td>
<td>Advises against, but does not prohibit, AI usage in systems targeting risk reduction efforts at SIL 2 and above.</td>
<td>IEC 61508 is criticized for its ambiguities in failure classification, safety loss quantification methods, and the CCF model it presents (Holstad and Corneliussen, 2004). These challenges intensify as automation levels in systems rise.</td>
<td>Anon (2010)</td>
</tr>
<tr>
<td>ISO 21448:2022 Road vehicles Safety of the intended functionality</td>
<td>Expands functional safety by providing a framework and guidance for ensuring Safety Of Intended Functionality (SOTIF) to mitigate risks from functional insufficiencies.</td>
<td>Does not specifically delve into embedding AI for safety in road vehicles.</td>
<td>Anon (2022b)</td>
</tr>
<tr>
<td>EN ISO 17757-2019 Earth-moving machinery and mining Autonomous and semi-autonomous machine system safety</td>
<td>Provides safety design guidelines for self-driving machines in earth-moving and mining industries, it primarily addresses safety requirements for these machines.</td>
<td>Excluding necessary guidance on the integration of such requirements i.e., traceability &amp; explainability.</td>
<td>Anon (2019)</td>
</tr>
<tr>
<td>EN ISO 18497:2018 Agricultural machinery and tractors. Safety of highly automated agricultural machines. Principles for design</td>
<td>Essential safety requirements at specific levels of automation and task handling, as well as means of verification and required usage information for highly automated operations.</td>
<td>Does not go into specific architectural design considerations for enhancing the reliability.</td>
<td>Anon (2018b)</td>
</tr>
<tr>
<td>EN ISO 19014:2018 Earth-moving machinery</td>
<td>Discusses the functional safety of earth-moving machinery and considers autonomous machine control systems as part of the safety control system.</td>
<td>Does not provide recommendations towards integration of AI in the safety control system.</td>
<td>Anon (2018a)</td>
</tr>
</tbody>
</table>
under the Marie Skłodowska-Curie grant agreement No 858101. Nataliya Strokina was supported by the Academy of Finland (project no. 336357, PROFI 6 - TAU Imaging Research Platform) If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to help the editing and writing process, particularly in terms of grammar, flow and use of the English language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgements

We would like to thank Jyrki Sauramäki and Petri Aaltonen for their thorough discussion on the topic. This research is funded through the Doctoral School Of Industry Innovation (DSII), meaning it is a joined funding between EPEC oy and Tampere University, Finland.

Appendix A. Overview of scope and limitations of machine safety standards

See Table A.3.

Appendix B. Emerging hazards for level 4 off-road machinery

See Table B.4.

Table B.4

<table>
<thead>
<tr>
<th>Hazard Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware resource constraints</td>
<td>Running computational heavy algorithms for on resource-constrained devices may lead to computational bottlenecks and reduced accuracy.</td>
</tr>
<tr>
<td>Limited Generalization</td>
<td>AI/ML model could struggle to accurately infer on data it has not been adequately trained on, leading to inaccurate results.</td>
</tr>
<tr>
<td>Security Vulnerabilities</td>
<td>Software systems especially those connected to a network may be susceptible to cyber-attacks or adversarial manipulations, potentially leading to compromised results, as discussed in Kaneko et al. (2023).</td>
</tr>
<tr>
<td>Model Drift</td>
<td>Over time, the AI/ML performance can degrade due to changes in the environment, necessitating regular monitoring and re-training.</td>
</tr>
<tr>
<td>Failure to Account for Dynamic Objects or Movements</td>
<td>AI/ML model can struggle to accurately estimate poses in scenarios with rapid or unexpected movements.</td>
</tr>
<tr>
<td>Lack of Redundancy or Backup Systems</td>
<td>Failing to implement redundancy or backup systems could lead to a complete breakdown in capabilities in the event of a failure.</td>
</tr>
<tr>
<td>Systematic failures</td>
<td>A systematic failure is a recurring or consistent flaw or deficiency in a software, leading to consistent and predictable errors rather than isolated incidents (Anon, 2010).</td>
</tr>
<tr>
<td>Random failures</td>
<td>those failures that occur unpredictably due to inherent physical or statistical processes, rather than systematic or design-related issues (Anon, 2010).</td>
</tr>
</tbody>
</table>

References


