Towards Safer Supervisor-Machine Shared Decision-Making for Runtime Hazard Mitigation in Highly Automated Off-road Machinery

Marea de Koning[1], Tatiana Minav[1], Willie L. Brown[2], Reza Ghabcheloo[1]

1 Faculty of Engineering and Natural Sciences, IHA, Tampere University, Finland
2 Faculty of Engineering and Aviation Sciences, University of Maryland Eastern Shore, United States of America

KEYWORDS: Automation, Machinery, Safety, Legislation, AI proposal, Machinery regulation, Risk management, Decision-making, Runtime Hazard Mitigation

ABSTRACT

In light of evolving safety risks associated with automating off-road mobile machinery, the European Union (EU) introduced Machinery Regulation 2023/1230 and the world's first Artificial Intelligence (AI) regulation the '*AI Act'*. Challenges persist in aligning EU harmonized standards with these legislative changes, presenting compliance difficulties for Original Equipment Manufacturers (OEMs), suppliers, and regulatory authorities. Transitioning towards highly automated off-road mobile machinery places humans into a supervisor role, where they are not actively engaged with the control of the equipment. This introduces safety assurance complexities as placing a human in a supervisory role is known to be impractical, due to measurable costs in human performance, such as decreased situational awareness and automation bias. Addressing this, we propose extending a safety risk assessment process with hazard awareness and a machine's capacity for timely reactions. Our proposed conceptual risk assessment framework integrates Machine Awareness of the Hazard (MAH) and Degree of Decision Making (DoDM) to determine Levels of Controllability (LoC). Additionally, it emphasizes the importance of considering human engagement levels (HEL) alongside machine capabilities. This conceptual framework aims to bridge the gap between highly automated machinery and effective supervision, laying the groundwork for future safety compliance within the EU.

1 INTRODUCTION

The European Union (EU) dominates the machinery manufacturing sector, boasting an impressive annual turnover of €740 billion [1]. Committed to ensuring high levels of protection and safety for workers and citizens, the EU mandates that Original Equipment Manufacturers (OEMs), and suppliers obtain a 'Conformité Européenne' (CE) declaration before placing off-road machinery within the European Economic Area (EEA) [2]. In response to evolving risks associated with automation in off-road mobile machinery, the EU introduced Machinery Regulation 2023/1230 [3] and drafted the world's first regulation of Artificial Intelligence (AI), the '*AI Act'* [4]. While current legislative gaps are addressed, EU-harmonized standards for OEMs have yet to fully align with these changes. This discrepancy poses challenges for both OEMs and regulatory authorities in ensuring consistent compliance. Furthermore, regulations continue to advocate for robust risk management systems, often incorporating a human-in-the-loop safety option [2, 5], which is known to be impractical [6].

Advances in automation show that highly automated off-road mobile machinery, where the operator assumes the role of a supervisor is technologically feasible, as demonstrated by the Volvo LX03 research concept in smart construction technology [7], and the BOMAG fully autonomous tandem roller [8]. Unfortunately, while it often appears that autonomous off-road machinery, where a human presence is no longer required, is just around the corner, the reality is that edge cases are endless. As discussed in [9], no matter how sophisticated the computation, how fast the Central Processing Unit (CPU) or storage capacity is, there remains, perhaps for now, an unbridgeable gap (a "humanity gap") where we still require a human presence [9].

OEMs and suppliers must ensure that under the intended conditions of use, operator discomfort, fatigue, and physical and psychological stress are eliminated or reduced to the minimum possible to prevent accidents [3]. Therefore, when an accident does occur, and it can be determined it was not caused by a faulty highly automated off-road mobile machine or the employer being negligent in providing a safe work environment, the operator could be held liable i.e., responsible. However, placing a human in a supervisory role without designing additional measures is unreliable [6, 10, 11] and morally questionable [12]. There is a myriad of examples where impairments in human-machine interaction are known to have led to disastrous consequences some might recall the Three Mile

Island nuclear accident from 1979 [13], or more recently the tragic loss of Air France Flight 447 in 2009, resulting in the loss of life of 228 people [14], the unintended acceleration of a Tesla Model 3 in Paris France in 2021, resulting in one fatality and 20 people injured, 3 of which severely [15], and a fatal collision involving a pedestrian and a self-driving Uber Technologies test vehicle where a back-up safety driver was present [16]. These examples of impairments in human-machine interaction leading to disastrous consequences are where what is known as a '*Responsibility gap*' emerges, meaning that normal conditions for holding human actors morally responsible for harm are not present [17]. This concept relates to the *'Moral crumple zone,*' coined by [12], it emerges when responsibility for an action could be misattributed to a human actor who has limited control or understanding over the behaviour of a highly automated or autonomous system.

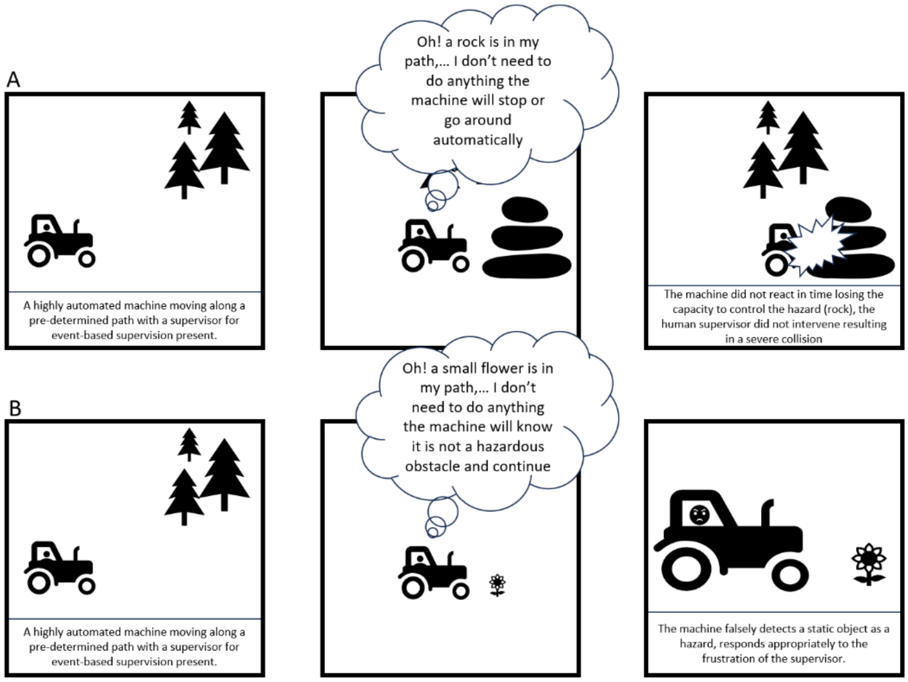
Ultimately, it still is the supervisor's capacity for event-based intervention which determines the avoidance of unexpected hazards. However, already back in 1948, Mackworth demonstrated that humans lack the cognitive capacity to effectively supervise automation, stemming from a significant deterioration in perceptual efficiency (i.e., situational awareness) eventually resulting in psychological problems [11]. Such deterioration can lead to the occurrence of '*Mode confusion*', a term originating from aviation psychology. Mode confusion occurs when the human(s) believes they are in a mode different than the one they are actually in and consequently make inappropriate requests or responses to the automated control system [18]. Measurable costs of placing a human into a supervisory role continue to be a limiting factor stemming from the onset of boredom, complacency, confusion on control goals/responsibilities, reaching the limitations of a supervisor's cognitive capacities, and automation bias [6]. This all comes down to a loss of awareness, resulting in ineffective and in some cases even harmful decision-making during safety-critical event-based intervention scenarios. In consideration of aviation psychology principles, this study examines how one could assess the risks of a highly automated off-road mobile machine by considering the relationships and roles associated with human or machine situational awareness in the design.

1.1 Supervisor awareness impact on event-based intervention

The dramatic decrease in awareness resulting in longer event-based response times coming from a lack of engagement because a human transitioned into a supervisory role is shown by the authors of [10]. They discuss the relationship between engagement in a non-driving task and driver response time when intervention is requested. Their findings suggest that the degree of engagement in a non-driving task could affect the drivers' information processing and recognition after a Take-Over Request (TOR). Indicating an average drop-in driver response time to 2 s once a TOR has been given. These findings are corroborated by the authors of [19]. In their extensive systematic review and meta-analysis of academic discourse on TOR and performance, they also analysed an average drop-in driver response time of 2.0-2.5 s. Interestingly enough, this drop-in response time is similar to that found in drivers who are considered indisposed (alcohol, illness, fatigue, etc.) [20, 21]. Regardless of disengagement due to a driver being placed in a supervisory role or other reasons for being indisposed, this is a significant drop from the normal driver response time when actively engaged, which is said to be about 0.5-0.55 s [22].

Naturally, these findings on event-based response times were based on passenger car drivers, limiting their applicability. It is challenging to determine if highly trained off-road mobile machine operators transitioning to supervisory roles experience a similar decline. However, considering known issues with human performance in supervisory roles, some decrease is anticipated. Figures 1A and B, illustrate examples of a gap that emerges by overly relying on event-based supervisor intervention. In Figure 1A, the supervisor is present in the cabin of a highly automated machine which moves along a pre-determined path in an off-road setting. The highly automated off-road mobile machine is designed to plan a new trajectory avoiding static obstacles most of the time. The supervisor is aware of this capability and therefore trusts that it will either stop or adapt trajectory adequately reducing the risk of the aforementioned hazard. Unfortunately, for both the supervisor and the machine there was no timely response to the static obstacle resulting in a severe collision. Conversely, Figure 1B, shows a similar scenario where the highly automated off-road mobile machine is designed to plan a new trajectory avoiding static obstacles most of the time. The supervisor is aware of this and therefore gets increasingly frustrated with the machine for coming to a stop when it could easily continue.

Considering our assertion and acknowledging the impracticalities associated with human supervision, we recognize that the design process for safer highly automated off-road machinery can benefit from extending the safety risk assessment process. This extension should focus on hazard awareness and the capacity of the highly automated off-road machine to avoid hazards through its timely reactions, as well as the supervisor's ability for event-based intervention. In this article, we propose that, alongside the conventional risk assessment approach outlined in most off-road machine safety standards, there remains a need to enhance our comprehension for assessing the risks of human-machine interaction with a focus on hazard awareness and capacity for event-based intervention. In section 2, we propose a concept for future risk assessment for highly automated off-road mobile machinery. We conclude our work in Section 3.



Figures 1A and 1B *illustrate the consequences of over-reliance on event-based supervisor intervention. In Figure 1A, a supervisor in the cabin of a highly automated machine in an off-road setting trusts the machine's capability to avoid static obstacles, but a severe collision occurs due to a lack of timely response. Conversely Figure 1B depicts a frustrated supervisor as the machine stops unnecessarily despite its capability to recognize obstacles.*

2 FIRST STEPS IN FURTHER EXPLORATION

In highly automated off-road machinery, a supervisor's presence remains vital, especially for high-risk machinery compliance. While such machinery should be capable of executing its own safety functions [3], human presence is still necessary for unforeseen scenarios, albeit not always reliable. Acknowledging this, we've begun our exploration by refining and extending a risk assessment method, inspired by EN ISO 19014:2018 [23].

First, we re-defined *Controllability* as "the capacity of the highly automated off-road machine to avoid harm to the agent(s) at risk of a hazard, through its timely reactions and the supervisor's capacity for event-based intervention". Here, we define *agent*(s) as either human(s), machine(s), or other environmental factors considered to be at risk of harm during a risk assessment. *Harm,* being the physical injury or damage to the health of people or damage to property or the environment, *risks* the probability of its occurrence and *hazard* as the source of harm discussed in terms of its severity ranging from unremarkable to a fatality.

In our risk assessment concept, we initially assert that the Level of Controllability (LoC) hinges on a combination of Machine Awareness of the Hazard (MAH) and Degree of Decision Making (DoDM). MAH refers to the degree of a machine's cognitive understanding of potential hazards, whether known all of the time, most of the time, sometimes or are unknown, before, during or after the time of action. This concept is an extension from EN ISO 19014:2018, which focuses solely on the operator's hazard awareness. The level of MAH aims to describe the machine's awareness of a hazard. This level of MAH when combined with DoDM then comes to a LoC which in turn could provide insights into describing minimum required Human Engagement Levels (HEL). The intent is that this can then enhance risk assessments by better describing the safety case for timely human event-based intervention. Machine Awareness of the Hazard (MAH) can be spread across 5 levels, ranging from MAH[0…4]:

- MAH 0 – Hazard is **not known** at the time of action.
- MAH 1 – Hazard is **known at** the time of action **sometimes.**

– MAH 2 – Hazard is **known at** the time of action **all of the time.**
– MAH 3 – Hazard is **known before** action, **most of the time.**
– MAH 4 – Hazard is **known before** action, **all of the time.**

This is complemented by the Degree of Decision Making (DoDM), which classifies the extent to which the logic-driven component of a safety-related control system could ensure a safe state when a hazard arises. This logic-driven element is the safety reasoning part of the safety-related control system. Decision-making refers to the system's ability and capacity to react to hazards and execute the necessary actions to achieve a safe state. A safe state is attained when appropriate risk reduction measures for relevant hazards are implemented. In our conceptual risk assessment, we align DoDM with human response times when fully engaged with the machine. For instance, in a driving scenario, a typical response time is 0.5-0.55 seconds [22], which we consider as our baseline. Based on this, the DoDM safety reasoning should be similar to a human, categorizing it into 5 levels, ranging from DoDM[0…4] :

– DoDM 0 – **significantly slower** than human response time to a particular hazard, 2-2.5 s
– DoDM 1 – **slightly slower** than human response time, to a particular hazard 0.8-2.0 s.
– DoDM 2 – **similar degree** of response time to a particular hazard as human, 0.5-0.8 s.
– DoDM 3 – **slightly faster** than human response time, to a particular hazard 0.3-0.5 s.
– DoDM 4 – **significantly faster** than human response time, to a particular hazard 0.2-0.3 s.

We propose that combining MAH and DoDM can provide us with a LoC as shown in Table 1.

Table 1 – Proposes levels of Controllability (LoC) as an interaction between the Degree of Decision Making (DoDM) and Machine Awareness of the Hazard (MAH).

	MAH 0	**MAH 1**	**MAH 2**	**MAH 3**	**MAH 4**
DoDM 0	0	0	0	0	0
DoDM 1	0	1	2	3	4
DoDM 2	0	2	2	3	4
DoDM 3	0	3	3	3	4
DoDM4	0	4	4	4	4

Human response time decreases as trust in machine automation rises as discussed in Section 2. Our proposed LoC requires higher HEL (Human Engagement Level) levels at lower LoC degrees, organized into 5 levels, ranging from LoC[0…4]:

– LoC 0 – **Significantly worse** than a human in a similar hazardous situation, requires **active "hands-on" human engagement,** the human is present inside the cabin, monitors behaviour through a visual feedback system and has hands on the wheel**.**
– LoC 1 – **Worse** than a human in a similar hazardous situation, requires **some "hands-on" human engagement,** the human present inside the cabin could be supervising the machine through a visual feedback system**.**
– LoC 2 – **Equal** to humans in a similar hazardous situation, requires **human presence, but not necessarily "hands-on"** as the machine is expected to be adequately capable of mitigating a hazard**.**
– LoC 3 – **Better than** humans in a similar hazardous situation, requires **human presence**, or relatively active engagement from a supervisor at a distance. Where in event-based scenarios observed through a visual feedback system an emergency safe state can be recovered i.e., through an emergency stop**.**
– LoC 4 – **Significantly better than** humans in a similar hazardous situation, requires **human presence,** at a distance in event-based scenarios observed through a visual feedback system an emergency safe state can be recovered i.e., through an emergency stop**.**

For instance, a machine capable of avoiding obstacles along a predetermined path could typically detect static obstacles reliably. Yet, there are instances where obstacles go unnoticed, hindering the machine's hazard control and limiting its ability to respond promptly and safely. Consider if the machine fails to identify a dynamic obstacle and initiates braking only after a collision occurs. In this case, both the Machine Awareness of the Hazard (MAH) and Degree of Decision Making (DoDM) are at level 1. According to our preliminary LoC Table 1, this implies the need for some level of active human engagement, despite the machine's general capability to avoid hazards at runtime.

We acknowledge the need for a human-in-the-loop safety option for event-based intervention, especially with increasing machine autonomy. However, simply assigning a supervisor to this role may be irresponsible. Our proposal extends the risk assessment process outlined in EN ISO 19014:2018, but it's limited as the response times we advocate for are based on automotive research data. We cannot reliably align them with the response times of highly skilled operators. Nonetheless, we believe that further development of this work, though highly conceptual, has the potential to bridge the gap between highly automated machinery and human engagement in supervision.

3 CONCLUSION

The EU mandates CE declaration for OEMs in the machinery sector, aiming to ensure worker and citizen safety. However, evolving regulations like Machinery Regulation 2023/1230 and the proposed AI Act outpace harmonized standards. This discrepancy poses compliance challenges for both OEMs and regulators. While conventional standards assume constant human oversight for safety, highly automated off-road machinery requires operators to act as supervisors, presenting cognitive limitations and safety risks.

Recognizing the supervisor's vital role for unforeseen scenarios, though its reliability is uncertain, we've embarked on exploring and understanding this further. Inspired by EN ISO 19014:2018, we are refining and extending the risk assessment approach. Our proposed conceptual framework integrates Machine Awareness of the Hazard (MAH) and Degree of Decision Making (DoDM) to determine Levels of Controllability (LoC). It also underscores considering human engagement levels (HEL) alongside machine capabilities. This framework aims to bridge the gap between highly automated machinery and effective supervision, laying the groundwork for future safety compliance within the EU.

Adopting a machine-supervisor awareness concept can foster a user-friendly design approach to achieve and maintain a level of security, safety, and technology readiness that is acceptable for runtime hazard mitigation in highly automated off-road machinery. This advancement is crucial to the engineering design process for tracking and monitoring the performance measures of system environments during standard operations and procedures for risk mitigation [24].

Through our collaboration with stakeholders and academic partners, the supervisor-machine shared decision-making process aimed at improving technology design development highlights the critical need for an adaptive environment using automation to manage runtime hazard mitigation for an integrative and transformative engaged system. This investigative approach, with supervisor awareness at its core, establishes a design process for requirements based on the impact of event-based intervention given the capabilities, resources, interests, and connections related to runtime hazard mitigation in highly automated off-road machinery [25].

Future work intend to enhance our understanding of the associated risks and further refine our conceptual risk assessment proposal. Intending to assist designers of highly automated machinery in defining responsibilities and setting minimum safety requirements for situational awareness and monitoring human engagement. Moreover, future iterations of this effort could pave the way for creating a compelling safety argument in line with Regulation 2023/1230. This, in turn, could streamline the future market entry of compliant highly automated machinery within the EEA.

4 REFERENCES

1 Gospodinova S., "Commission welcomes political agreement on new rules to ensure the safety of machinery and robots," EU Commission, 2022.

2 European Commission, "*On machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC*," European parlement and the council, 2023.

3 European Council, "Regulation *(EU) 2023/1230 of the European Parliament and of the Council*," Official Journal of the European Union, 2023.

4 European Commission: AI regulation proposal, "*Harmonised rules Artificial intelligence (Artificial intelligence act) and amending certain union legislative acts*," European Union, 2021.

5 Koning de, A.M.R., Ghabcheloo R, "*Machine safety conformance limitations for highly automated and autonomous heavy-duty mobile machinery*," in Safety-Critical-Systems Conference, York, UK, 2023.

6 Wickens C. D., "*Engineering Psychology and Human Performance,"* New York: Taylor & Francis Group, 2021.

7 Volvo, “Volvo LX03,” [Online]. Available: https://www.volvoce.com/global/en/about-us/what-we-believe-in/innovation-at-our-core/concept-lab/lx03/. [Accessed 1 4 2024].
8 BOMAG, “BOMAG,” [Online]. Available: https://www.bomag.com/ww-en/press/news-videos/future-study-fully-autonomous-tandem-roller/. [Accessed 1 4 2024].
9 Bishop J. M., “*Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It*,” Journal Frontiers in Psychology - Cognitive Science, 2021.
10 Rauffet P, Botzer A, Chauvin C, Saïd F, Tordet C. “*The relationship between level of engagement in a non-driving task and driver response time when taking control of an automated vehicle.”* Cognition, Technology & Work. 2020.
11 Mackworth N. H., “*The breakdown of vigilance during prolonged visual search*,” Journal of experimental psychology, 1948.
12 Clare E. M., “*Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction*,” Engaging Science, Technology, and Society, p. 21, 2019.
13 website, "*Three Mile Island Emergency*," Dickinson College’s, [Online]. Available: https://tmi.dickinson.edu/. [Accessed 4 2024]
14 BEA, “Final Report Airbus A330-203 Air France,” Bureau of Enquiry and Analysis for Civil Aviation Safety, France, 2012
15 Clercq G. D., "*Paris crash Tesla driver says car accelerated on its own - lawyer,*" Reuters, [Online]. Available: https://www.reuters.com/world/europe/paris-crash-tesla-driver-says-car-accelerated-its-own-lawyer-2021-12-16/. [Accessed 4 2024].
16 Shepardson D., "*Backup driver in 2018 Uber self-driving crash pleads guilty*," Reuters, 29 07 2023. [Online]. Available: https://www.reuters.com/business/autos-transportation/backup-driver-2018-uber-self-driving-crash-pleads-guilty-2023-07-28/. [Accessed 15 04 2024].
17 Burton S, Habli I, Lawton T, McDermid J, Morgan P, Porter Z. Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. Artificial Intelligence. 2020
18 Joshi A., Miller S.P., Heimdahl M.P., “*Mode confusion analysis of a flight guidance system using formal methods,*” InDigital Avionics Systems Conference, 2003. DASC'03. The 22nd Oct. 2003, IEEE.
19 Soares S, Lobo A, Ferreira S, Cunha L, Couto A. “*Takeover performance evaluation using driving simulation: a systematic review and meta-analysis*.” European transport research review. 2021
20 Yadav AK, Velaga NR. “*Modelling brake transition time of young alcohol-impaired drivers using hazard-based duration models.”* Accident Analysis & Prevention. 2021
21 Čulík K, Kalašová A, Štefancová V. “*Evaluation of driver’s reaction time measured in driving simulator.*” MDPI Sensors. 2022
22 Dickerson AE, Reistetter TA, Burhans S, Apple K. “*Typical brake reaction times across the life span.”* Occupational Therapy in Health Care. 2016
23 ISO, “*19014-1:2018 - Earth-moving machinery*,” 2018.
24 Sloan, V., & Amado, F. T., & Smith, C., & Campbell, B. S., & Brown, W. L., & Mei, L., “*Learners’ Peer-to-Peer Interactions of Aerospace and Aviation Education with Unmanned Aerial Systems Designs Using Data Methods Integration*,” Paper presented at 2023 ASEE Annual Conference & Exposition, Baltimore, Maryland. 10.18260/1-2—43379, 2023
25 Brown, W. L., Johnson, E., Mei, L., Cornelius, J., Sharma, D., Zhu-Stone, W., Koning de, A.M.R., Cornelius, T., Dabipi, I., Zhang, L., Wiggins, U., Jackson, E., "*A Peer-to-Peer Guide to Academic Transformation using Research-to-Practice of STEM Learners to Promote a Lifecycle-Oriented Project for Accessibility within the User Community and Environments*," 2023 IEEE Frontiers in Education Conference (FIE), College Station, TX, USA, doi: 10.1109/FIE58773.2023.10343008, 2023