Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P378-cd

Challenges and Solutions for Autonomous Systems Safety – Findings from three International Workshops (IWASS)

Ingrid Bouwer Utne

Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. E-mail: ingrid.b.utne@ntnu.no

Christoph Alexander Thieme

Department of Software Engineering, Safety, and Security, SINTEF Digital, Trondheim, Norway. E-mail: christoph.thieme@sintef.no

Marilia Ramos

B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles (UCLA), USA. E-mail: marilia.ramos@ucla.edu

Ali Mosleh

B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles (UCLA), USA. E-mail: mosleh@ucla.edu

Autonomous systems intend to be a stepping-stone towards safer and more efficient operations. Extensive mapping and monitoring of land, space, and the oceans, renewable energy harvesting and production, inspection of physical structures difficult to access, operation of subsea systems, land-based, maritime, and air transportation are emerging areas for high autonomy. Still, the corresponding advancements in software, hardware, and interactions with humans and the environment involve complexities that pose major challenges concerning safety, reliability, and security (SRS). Society is hesitating to allow for widespread use of highly autonomous cars or ships. The industrial use of autonomous systems depends on effective and transparent standards for safety, verification, and certification, for which it is essential to develop credible methodologies for assessing risk, acceptance criteria, testing and verification. Hence, risk management must become an important driver in the early design process and during operation. Autonomous systems must have sufficient integrity, be capable of determining if it can continue operating with degraded performance, and cooperate with human operators. Since 2019, the International Workshop on Autonomous Systems Safety (IWASS) has been bringing together multidisciplinary experts from academia, industry, and authorities to discuss the SRS challenges and potential solutions. This paper aims to provide an overview of the discussions and results from the three editions of IWASS and discuss potential ways to enhance SRS in future developments and implementations of autonomous systems.

Keywords: Autonomous systems, Safety, Reliability, Security, IWASS, Autonomy.

1. Introduction

Autonomous system development is enabled through new and enhanced applications of artificial intelligence (AI), machine learning, cheaper and abundant sensor technology, and more powerful and integrated software and hardware systems and solutions. Autonomous systems often operate in unstructured environments and depend on safe interaction between their different subsystems and components: the hardware, software, and the human operator or supervisor, when applicable. However, reaching higher levels of autonomy is a challenge concerning how to ensure safety, reliability, and security (SRS) in the design, operation, approval, and acceptance of such systems (Utne et al, 2017). To develop safe solutions for autonomous systems it is essential to:

- Identify, understand, analyse, and evaluate risks in autonomous systems operations;
- Integrate safe solutions in the design;

- Monitor, and follow up that the risk level in operation is acceptable;
- Implement regulations and procedures that ensure safe operations;
- Convey safety to society to establish trust in the systems.

These safety challenges are interdisciplinary. They require a joint effort by research and industry in different fields to develop feasible solutions. Hence, the International Workshop on Autonomous Systems Safety (IWASS) was established in 2019 to discuss and find solutions to the SRS challenges of autonomous systems, covering autonomous maritime, marine, land vehicle, railway, and aerospace systems. So far IWASS has been conducted three times, bringing together around 50 international experts from academia, industry, and the authorities to discuss the SRS challenges and potential solutions to the advancements of autonomous systems (IWASS, 2019; 2021; 2022).

This paper aims to provide an overview of the discussions and results from these workshops[a], and state potential ways to enhance SRS in future developments and implementations of autonomous systems. The remainder of the paper is structured as follows: section 2 presents the main SRS challenges. Section 3 discusses potential solutions to these challenges. Section 4 offers concluding remarks.

2. Main safety, reliability, and security challenges

2.1. Overall concepts and topics

Safety is a "state where risk has been reduced to a level that is as low as reasonably practicable and where the remaining risk is generally acceptable". Security is the freedom from or resilience against harm created through voluntary actions targeting directly or indirectly the system (Rausand & Haugen, 2020; ISO/IEC Guide 51, 2014). Reliability, on the other hand, can be defined as "the ability of an item to perform a required function, under given environmental and operational conditions and for a stated period time" (Rausand & Høyland, 2004).

Reliable systems are not necessarily safe and secure. The difference becomes apparent, for example, when the software for an autonomous system is considered. Instead of stopping when being operated outside its design envelope, the control software may try to recover the system. Safety features, for example, may be exploited by hostile agents to gain control or access. Secure systems are not always safe, for example, a secure system may be overly complicated for its users leading to unsafe operation.

In the IWASS workshops, the following topics were identified as most important with respect to impact on autonomous systems SRS:

- The role of humans;
- SRS modelling and methods for understanding complexity and cascading failures;
- Security and cybersecurity;
- Demonstrating safety – verification and validation (V&V);
- Risk acceptance;
- Legal and regulatory aspects;
- Ethical and social aspects;
- AI, data analytics, and decision support in resilient autonomous systems.

These topics are further elaborated in the next sub-sections.

2.2. The role of humans

One of the main motivations for developing autonomous systems is to improve safety by avoiding human failure since they rely less on humans in operation. In many cases, however, humans will still be needed for intervention, remote control, or monitoring, as established by the Level of Autonomy (LoA). Human failures may occur due to different reasons, initiating an incident or contributing to one initiated by a system failure or external factors. For example, humans may use an autonomous system in a different context, exceed the operating envelope, or not act as expected for emergency response. Similarly, a software failure may provide misleading information to operators or not provide the necessary data, causing human failure.

[a] https://www.risksciences.ucla.edu/iwass-2023-home

Despite humans still being involved in operation, many systems are not designed to ensure that the human operator is able to intervene when necessary (e.g., providing insufficient time for situation assessment and action of a safety driver). Autonomous systems modify how humans interact with the system and, thus, may lead to new human failure mechanisms and influencing factors. Some of these factors include automation complacency, boredom, lack of trust in automation, and the human being out of the loop (Ramos & Mosleh, 2021). Addressing these challenges includes holistically analysing the human role in autonomous systems' operation, i.e., considering the interactions between the human and the system in different operational phases and LoAs, and designing the system to prevent or mitigate potential human errors.

2.3. SRS assessment methods and modelling

Most current quantitative assessment methods used in conventional risk and safety assessments rely on the separation between software, hardware and humans. This means that system components are assumed to be independent and often analysed separately (Mosleh, 2014). As a result, interactions between components, emerging properties, and complexity are often neglected or simplified in the use of conventional methods. As a result, complex systems may not be sufficiently represented and analysed.

Some qualitative methods, such as the Systems-theoretic Process Analysis (STPA) (Leveson & Thomas, 2018) and the Functional Resonance Analysis Method (FRAM) (Hollnagel, 2012), consider the different system elements, including emerging properties and system interactions. These methods provide useful qualitative insights but are of limited value in prioritizing risks and mitigating measures.

For hardware subsystems and components, such as engines, valves, or drive trains, there are mature quantitative methods and mathematical approximations of failure probabilities and degradation.

For software safety, SRS assessments are more difficult to perform. Software reliability, which is different from software risk, is approximated by, for example, the remaining number of errors in the software, which does not describe how the software may fail and the corresponding system and operational effects. Commonly used approaches for software SRS assessment in the industry build on checklists or focus on fulfilling formal requirements as proof for SRS compliance (ISO/IEC Guide 51, 2014). The interaction between software components, including from different suppliers, is challenging as thousands of lines of code may need to be analysed and checked. Recently, risk analysis for software has been addressed to an increasing extent (Aldemir et al, 2010; Thieme et al, 2021).

An often-discussed topic in risk assessment for autonomous systems is the need for new and adapted methods. A holistic approach is required for the SRS assessment, considering the potential interactions, and outcomes (Hollnagel, 2012).

2.4. Security and cybersecurity

Cyber security, data and information technology (IT) security, and physical security are major challenges for autonomous systems. Cyber security and software risks differ from traditional security issues and hardware failures, since past behaviour cannot be used to predict future behaviour, particularly in situations where the systems continuously learn.

Attackers may exploit the autonomous system´s behaviour and vulnerabilities, which may put passengers and goods at risk. Vulnerabilities occur through flaws in the design and development process of hardware or software, and by the human users. Hardware hacking, e.g., by introducing micro computers into the system, may allow access for an attacker (Wyglinski et al, 2013). The complexity of the autonomous systems may mask vulnerabilities and hide intrusion or access by attackers.

Communication protocols between components developed many years ago without any security and cybersecurity mechanisms may also lead to vulnerabilities, for example, through poorly integrated system components, wireless communication and/ or entertainment systems, systems for remote monitoring, jamming and spoofing (Meland et al, 2021; Vinnem & Utne, 2018; Haas & Möller, 2017). Whereas a jammed sensor is not able to fulfil its function, a spoofed sensor will produce fake signals. Jamming and spoofing have enabled hijacking and stolen autonomous systems (Yagderlei et al, 2015; Parkinson et al, 2017).

2.5. Legal and regulatory aspects

Legal and regulatory aspects are challenging since many existing regulations do not consider autonomy, and regulators face the challenge of adjusting or developing new regulations for autonomous and semi-autonomous systems. Simultaneously, system developers face the challenge of demonstrating that the systems are safe to regulators. An example is ships, regulated by the International Maritime Organization (IMO). Regulations from IMO move slowly since shipping is an international industry, and regulations are implemented and enforced by each IMO member state. One of the regulatory issues is that several conventions require vessels to be sufficiently crewed to maintain a safe lookout. Hence, current "autonomous ships" are manned with a safety crew to manually take control if a dangerous situation occurs.

Liability is another challenge in regulating autonomous vehicles, as it may be challenging to determine who should be responsible if an accident happens. Should, for example, developers of anti-collision algorithms be responsible for a collision or to what extent is the remote driver or supervisor responsible for a mal-functioning system and/or failure to intervene?

To comply with regulations for autonomous systems we need to test and use them to assess their safety. Nevertheless, we do not want them on the road, in the sky, or at sea before we know they are safe. Another challenge is that different industries and countries have their own regulations regarding requirements to the development, testing and deployment of autonomous systems. Adopting international regulations is thus challenging. The IMO, for example, only gives recommendations for adoption by member states, which means that a vessel may not be accepted in a country that did not adopt this resolution. Involving regulators may be useful for gaining mutual understanding of the challenges, but in some countries or industries, the regulators cannot become involved to ensure that they remain impartial and not favour specific solutions.

2.6. Ethical and social aspects

How ethics and moral are implemented in the decision-making and behaviour of autonomous systems will influence the societal acceptance.

The German Ethics Commission on Automated and Connected Driving (2017) provided the first official guidelines for the ethical choices of autonomous vehicles. One of the rules here states that human life should have priority over other animals' life. Another rule states that distinction based on personal characteristics, such as age, should be prohibited.

Also, some ethical aspects are related to liability. In the U.S., for example, the more a person earns, the greater their liability exposure. A potential effect of this may be that producers of autonomous vehicles protect themselves against major liability claims by adjusting the car´s driving behaviour in accordance with the average wage in a region (Himmelreich, 2018).

Communicating safety to society is a must to gain trust in autonomy and societal acceptance. Regulators may have to define some level of ethical requirements and expectations to autonomous systems. Determining such requirements, however, may be challenging as ethics is not an objective and quantifiable topic. Biases related to equity may find their way into the decision algorithms, and hence, databases used for training of AI decision systems. Software designers and programmers are responsible for the algorithms and behaviour, and for selecting training data.

2.7. *Demonstrating safety*

Demonstrating safety of autonomous systems is not trivial and is related to regulatory, societal, and ethical requirements. An important concern is how to derive the requirements against which to verify the system. V&V may be a particular challenge with autonomous systems because of the system's complexity and lack of consensus on regulation and ethical guidelines. Furthermore, compliance with one set of requirements may compromise other performance aspects; for instance, put a safety driver in a car and it's all good from a liability and system failure perspective, but that creates a challenge to human performance with respect to time to act and sufficient situation awareness.

An additional concern is to determine the best verification processes to use for autonomous systems. Given their complexity, embodiment in the real world, and potential for adaptation or learning, continuous and integrated processes are recommended. Determining if the risk associated

with these systems is acceptable can only be achieved when performance has been verified and validated. Verifying autonomous systems may be more resource demanding than conventional systems since there is increased focus on system behaviour. Foreseeing abnormal scenarios for autonomous systems may be more critical to avoid these becoming less robust and innovative concerning handling the unforeseen. This may cause a "system state explosion".

2.8. Risk acceptance

Risk tolerance may vary among industries depending on the regulations, stakeholders' acceptance and trust, and the risk level of existing human-operated systems. The safety of an autonomous system has been compared to its human-operated counterpart; i.e., an autonomous vehicle should be as safe as a human-driven one. Nevertheless, the perspective of "as safe as existing conventional systems" raises some issues.

Firstly, historical data for autonomous systems is insufficient for comparison with human-driven ones. A suggested approach is a behaviour comparison, meaning that an autonomous system should behave similarly to a human-operated system. However, characterizing a human operator's behaviour is not straightforward. For instance, car drivers' behaviour can differ greatly based on age, experience, country, etc. Hence, to benchmark the autonomous cars' behaviour against a human-driven car becomes highly challenging.

Secondly, autonomous systems' operation's "unknown unknowns" may become ignored. Autonomous systems comprise new elements, such as connectivity issues, emergent failures from unsafe interaction between subsystems, security concerns related to hacking and spoofing, etc. Thus, the risk level of autonomous system operation is challenging to compare with conventional systems.

Finally, one of the reasons for developing autonomous systems is to increase the safety level. Hence, designing systems "as safe as current systems" may limit the benefits that can be achieved; namely, developing systems that are considerably safer than current ones.

Safer systems may also increase users' trust and increase public acceptance of new systems; in particular for those in which the public's tolerance concerning accidents involving an autonomous system is lower than for a human-operated one.

2.9. Artificial intelligence

Autonomous systems depend on AI and data analytics, which can be applied in two ways: (i) as part of the systems' intelligence, i.e., information processing, decision-making, or motion control, and (ii) as part of the safety assurance process of autonomous systems.

A key requirement to AI-based systems is that they must be able to detect if they are outside their operating envelope. This should include the detection of anomalies not included in training data sets and the appropriate reaction to these as this compares to a human driver who adapts to a new situation and identifies untrained situations.

A challenge with safety and AI is that whereas conventional software is claimed to be deterministic, predictable, and explainable, the output may sometimes be unpredictable for software with AI methods. In conventional software, this means that a specific output can be expected for a given input and it is possible to explain why. Hence, if faulty inputs are given (e.g., from a faulty sensor), it is possible to determine the expected output. Since it is not known how the AI's decisions are made, for example, for deep neural networks and other "blackbox" approaches, uncertainty increases and there is more difficulty when testing autonomous systems with AI, since deriving conclusions from limited "edge cases" may not hold true over the whole state space of input combinations.

3. Recommendations on potential solutions

The challenges described in the previous Section show that several recommendations (keywords in bold) for potential solutions are relevant for achieving safe autonomous systems:

It must be possible to adequately analyse and test the constituent **system elements.** This means that the system needs to be modularly structured so that subsystems and components can be investigated regarding impact on system performance and safety. AI and control systems must be developed to be verifiable, regulated, transparent, and explainable. More knowledge is

needed to understand the limitations of the technologies used for situational awareness.

It is necessary to define the system´s **operational envelope** and expected behaviour. At the same time, the possibility of violating or exceeding the operational envelope cannot be overlooked, which means that fail-safe solutions must be implemented. Before the final design and commissioning, a restricted envelope may help to assess system compliance and the expected behaviour.

Operations that include **human operators** and supervisors with shared control capabilities cannot be driven by technology development or legislation requirements only. Human capabilities must be carefully considered with respect to performance factors, such as available response time and the ability to perceive and process information in an event that may require human interaction. These capabilities may change through less involvement in normal operation.

Autonomous systems are often described by their **LoA**, or Degree of Automation, which generally assumes a linear progression of autonomy. Higher LoAs imply less human involvement, which may lead to the conclusion that humans are less relevant to the system safety. However, while the task load may be reduced in higher LoAs, the tasks may demand significantly higher levels of interaction and cognitive effort and be critical for the system's safety. In the design of the systems, it is important to keep in mind that this may lead to a false perception of simplicity (and safety) in the task-switching and shared control between humans and system. Hence, the LoAs must be revisited to clarify the human role in system operation.

The answer to “how safe a system is” concerns **risk levels**, which should be assessed against pre-defined acceptance levels. This means that the “as safe as current systems” approach to autonomous systems safety is challenging with respect to i) comparing systems objectively, and ii) potentially overlooking the opportunity for developing safer systems than existing ones. Furthermore, system **safety** performance must be adequately and accurately assessed, demonstrated, and communicated to stakeholders.

Risk analysis is a valuable tool for assessing system safety, and risk specialists/risk analysts must be engaged early in the design process along with specialists from other disciplines. Conducting useful and high-quality risk assessments of autonomous systems, thus, requires the right interdisciplinary team with a shared understanding and common language of work. A feasible risk assessment approach that identifies and includes the most relevant stakeholders and disciplines in the risk assessment must be developed and applied.

Risk science provides several methods for achieving safer autonomous systems. However, complexity and possible cascading failures pose several challenges concerning risk assessment methods. Many existing methods are inadequate and lack integrated modelling of autonomous systems' hardware, human, and software. Clear criteria for when an existing method is adequate for an analysis and when a new one is needed, are, however, still to be developed. These criteria must reflect the benefits of any new methods to the industry, as the adoption of methods requires resources in terms of training, knowledge, and changes in the safety assessment processes. Authorities, regulators, or third parties must also accept the use of new methods.

In general, there is a need for a "**framework**" consisting of various qualitative and quantitative methods for identifying, analysing, and evaluating different hazards and hazardous events combining both simulation and more traditional "discrete logic" approaches. Experts in other disciplines that focus, for example, on the use of the system, must be involved in the assessment. Domain experts, risk analysts, and regulators could define which disciplines should be involved in the assessment.

The complexity of autonomous systems could be addressed in risk assessment through the compartmentalization of the systems, related to the above-mentioned “system elements”. A challenge lies in defining the subsystems' boundaries and the correct integration of the sub-models with each other. When choosing the assessment method, it is important to consider the objective and context of the analysis, as well as its validation or verification.

Cybersecurity needs to be explicitly incorporated and considered in the risk assessments, even though the reliance on communication technologies to inform humans about the status of the autonomous system may lead to vulnerabilities and there is more reliance on sensors and raw environmental data. Hence,

multiple sensors that measure the same thing may be incorporated so that the system can identify if one or more of these sensors has been maliciously compromised. **Environmental conditions** impact the safety of several systems and operations. Due to climate change, risk assessment cannot rely solely on historical data.

AI-based methods, where the AI learns from data to facilitate assessment, are also seen as promising tool (Hegde & Rokseth, 2020). AI-based methods could help identify scenarios by combining system knowledge and knowledge of earlier assessments. However, the AI-based methods need to be built with domain knowledge and data used to tune the model. In some AI techniques, for example deep learning, little is known about the model parameters. Hence, a challenge is to develop transparent and trustworthy methods and models. Risk models using Bayesian networks (BN) for supervisory risk control, e.g., as proposed in (Utne et al; 2020) could be one approach to explainable AI. Ethical development and deployment of AI is required, meaning that guidelines for these processes need to be provided to the industries.

V&V should be performed early and throughout the whole system design process to provide feedback into the design, development, and deployment process. It is important to ensure that software aspects and testing are covered. It is necessary to demonstrate that the system complies with relevant standards and regulations, including the development processes and V&V approaches used. Safety needs to be continuously evaluated throughout a system's lifetime. In particular, a self-learning (AI based-) system needs continuous and integrated verification processes. Such evaluations should consider learnings and operational, environmental, and organizational changes. The large quantity of data generated by connected systems should be leveraged for updating risk assessments, monitoring and verification to evaluate the system performance and its relevant subsystems.

Ethical issues are often closely linked to legal matters, such as the implications of AI failures, misuse, or liability concerns, which have sometimes been considered a barrier for realizing highly autonomous systems. A solution may be to require the implementation of an operational data recorder for autonomous systems to clarify legal issues, and to provide information for explaining causes, errors, and faults in accident investigations.

4. Conclusions

This paper discusses the SRS challenges and presents potential solutions for autonomous systems, based on the discussions during IWASS 2019-2022. Even though autonomous systems may be a step towards safer and more efficient operations, more software and advanced control systems lead to increased complexity and emerging risks that are challenging to identify, analyse, evaluate, monitor, and mitigate. The third IWASS workshop concluded that there is not one simple "key" to solving the SRS challenges:

- Autonomy creates a new dimension in the human-machine relationship – from the operators directly involved to the people interacting with these systems externally.
- Designing autonomous systems with the ability to explain why they take specific actions and made certain decisions may lead to increased trust by stakeholders.
- Risk specialists must be engaged early in the design process, which is not always the case.
- Risk tolerance / acceptance remains a challenge with several proposed solutions, including different types of safety envelopes and constraints.
- The methodology for analysing and evaluating risks of software-intensive systems is advancing, but a framework is needed, including simulation, for more precise predictions of system/operational performance.
- LoA introduce risks and functional failures that are important to analyse in different operational modes, including the effects of shifts in the LoA and shared control with the human operator.
- V&V remains a challenge, which is closely linked to the methodological risk assessment problems. If risk methodology is improved, V&V may become easier. V&V efforts must be trustworthy and acceptable.

Finally, the gap between academic research, regulators, and industry with respect to disseminating theory, knowledge, experiences, and recommendations related to risk assessment approaches, needs to be closed. Learning across industries and application areas should be enhanced, which means that different people with diverse backgrounds and experience should be involved in

developing the solutions. Different disciplines must communicate and avoid operating in silos, to prevent suboptimal solutions. Interdisciplinary cooperation across different industries involving various stakeholders to find solutions to safer autonomous systems and operations is the continuing goal of the past and future IWASS workshops.

Acknowledgements

The work is partly sponsored by the Research Council of Norway through the Centre of Excellence funding scheme, 223254, AMOS, project UNLOCK 274441, and project ORCAS 280655.

References

Aldemir, T., S. Guarro, D. Mandelli, J. Kirschenbaum L.A. Mangan, P. Bucc, M. Yau, E. Ekici, D.W. Miller, Sun, X, S.A. Arndt (2010). Probabilistic risk assessment modelling of digital instrumentation and control systems using two dynamic methodologies. *Reliability Engineering and System Safety, 95 (10)*, 1010-1039.

Federal Ministry of Transport and Digital Infrastructure of Germany (2017). Ethics Commission on Automated and Connected Driving. doi:10.1126/science.186.4158.38.

Hegde, J., B. Rokseth (2020). Applications of machine learning methods for engineering risk assessment – A review*, Safety Science*, 122, 104492.

Himmelreich, J (2018). Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. *Ethical Theory Moral Pract* ;21:669–84.

Hollnagel E (2012). FRAM – The Functional Resonance Analysis Method. 1st Ed. Farnham. UK: Ashgate.

ISO, IEC. ISO/IEC Guide 51: Safety Aspects - Guidelines for their inclusion in standards. Geneva, Switzerland: 2014.

IWASS (2019). Proceedings, International Workshop on Autonomous Systems Safety, Trondheim, Norway. Ramos, M., C. Thieme, I.B. Utne, A. Mosleh (eds), ISBN: 978-82-691120-2-3.

IWASS (2021). Proceedings, International Workshop on Autonomous Systems Safety (online webinar). Thieme, C, M. Ramos, I.B. Utne, A. Mosleh (eds), DOI: 10.34948/N33019.

IWASS (2022). Proceedings, International Workshop on Autonomous Systems Safety. Dublin, Ireland. Thieme, C, M. Ramos, I.B. Utne, A. Mosleh (eds). ISBN 978-82-691120-4-7.

Leveson, N.G., J.P. Thomas (2018). *STPA Handbook*. 1. Cambridge, MA, USA: 2018.

Meland, P. H., K. Bernsmed, E. Wille, J. Rødseth, D. A. Nesheim (2021). A Retrospective Analysis of Maritime Cyber Security Incidents. *TransNav* 15 (3): 519–30. doi:10.12716/1001.15.03.04.

Mosleh, A (2014). PRA: A Perspective on strengths, current Limitations, and possible improvements. *Nucl Eng Technol* 2014. doi:10.1109/MM.2013.18.

Ramos, M, A. Mosleh (2021). Human Role in Failure of Autonomous Systems: A Human Reliability Perspective. 2021 Annual Reliability and Maintainability Symposium (RAMS), pp. 1-6.

Rausand, M, S. Haugen (2020). *Risk Assessment – Theory, Methods, and Applications*, 2nd ed, Wiley.

Rausand, M., A. Høyland (2004). *System Reliability Theory. Models, Statistical Methods, and Applications*, Wiley.

Thieme, C.A., A. Mosleh, I.B. Utne, J. Hegde (2020). Incorporating software failure in risk analysis—Part 2: Risk modeling process and case study. *Reliability Engineering and System Safety*, 198, 106804

Utne, I.B., A.J. Sørensen, I. Schjølberg (2017). Risk Management of Autonomous Marine Systems and Operations. In Proc 36th Int Conf Ocean Offshore Arct Eng. 10.1115/OMAE2017-61645.

Utne, I.B., B. Rokseth, A.J. Sørensen, J.E. Vinnem (2020). Towards supervisory risk control of autonomous ships. *Reliability Engineering and System Safety, 196*, 106757.

Vinnem, J.E., I.B. Utne (2018). Risk from cyberattacks on autonomous ships. *Saf. Reliab. – Safe Soc. a Chang. World*. doi:10.1201/9781351174664-188.

Wyglinski, A.M., X. Huang, T. Padir, L. La, T.R. Eisenbart, K. Venkatasubramanian (2013). Security of autonomous systems employing embedded computing and sensors. *IEEE Micro*. *33 (1), 80-86.*