IWASS

First International Workshop on Autonomous Systems Safety

Whitepaper

11th to 13th March, 2019
Trondheim | Norway

NTNU
Norwegian University of Science and Technology

The B. John Garrick Institute for the Risk Sciences
UCLA ENGINEERING

Welcome to IWASS

Welcome to the First International Workshop on Autonomous Systems Safety! The idea to create IWASS originated from a discussion on the challenges concerning of autonomous ships' safety. Through our knowledge on the similarities shared by autonomous systems, we envisioned how different applications involving autonomy could learn from each other. This insight was the first step towards the creation of IWASS as a platform for an integrated discussion on risks, challenges, and – more importantly – potential solutions concerning safe autonomous systems and operations.

The awareness on autonomous systems' similarities is not a novelty in the field. Yet, to our knowledge, no event to date has assembled experts on different autonomous systems with the purpose to discuss safety, reliability, and security (SRS). In the past, similar events have been organized around a specific type of autonomous system (e.g. cars, ships, aviation) or a particular aspect of safety related to such systems (e.g., the risk of cyber-attack, software reliability). IWASS distinguishes itself from these events – and complements them – by bringing these topics together in a unique attempt to focus on proposing solutions for common SRS challenges.

IWASS is a workshop by invitation only. The invitees have been thoroughly chosen for their potential contribution to the workshop; they are affiliated with numerous organizations (both in academia and in the industry), and have different areas of interest, coming from a variety of disciplines, and having produced significant work.

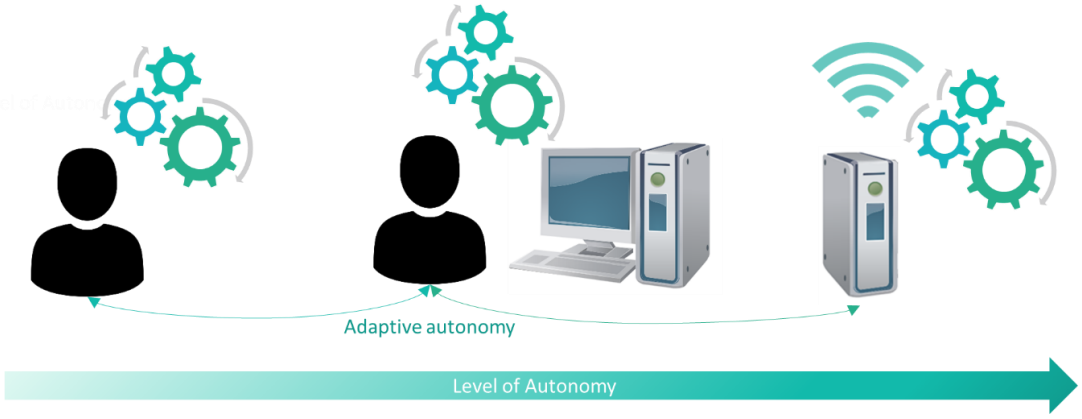
IWASS is arranged in Trondheim, Norway, in March 2019. The first day of the workshop focuses on selected presentations, whereas the second day is a combination of presentations and discussion, with breakout groups focusing on specific topics and preparing a report on their conclusions. In the third day, IWASS will be concluded with a general discussion on the breakout groups' findings and the key results of the workshop.

Following the conclusion of the workshop, proceedings will include a summary of discussions, participants' views, and recommendations on the path towards safe, reliable, and secure autonomous systems. Participants will be invited to read the proceedings and leverage their knowledge, identifying potential collaborations with other participants for future editions of IWASS. This whitepaper provides a starting point concerning some of the topics that will be addressed at IWASS, including the current state of the art on autonomous systems development and challenges it faces. In the following pages, we discuss challenges in respect to risk assessment techniques, human-machine interaction, cyber security, regulatory issues, and ethical aspects.

Autonomy and autonomous systems

The introduction of automation in a wide range of activities has changed how society interacts with machines. For years, automation was applied only to physical activities, rather than cognitive aspects as sense-making and decision-taking. The advent of artificial intelligence, machine-learning, and easier access to powerful software and sophisticated hardware have brought a new revolution into how we interact with automated systems, both as users as well as operators. The outcome of this revolution are highly automated and autonomous systems.

Autonomy can be defined as a system's ability to make independent decisions and to adapt to new circumstances to achieve an overall goal. This is achieved without additional input from human operators or other systems[(1)]. Automation, on the other hand, is often understood as the reproduction of an action, without any choice made by the machine executing the action[(2)].The degree of autonomy of a system may be assessed through Level of Autonomy (LoA). Several authors have proposed different scales for LoA[(3)], either generalizable to autonomous systems or specific to an industry[(1)]. In general, the LoA scale starts at a lower level autonomy in which information reception from the system and surroundings, situation assessment, decision-making, and command giving to the hardware are responsibilities of human operators. The LoA scale progresses to a higher level, when these tasks become responsibilities of a software. Between the lower and higher levels, these tasks are shared between software and human, as illustrated at the Figure below.



Simplified illustration of Levels of Autonomy

A system may be designed with an adaptive autonomy(4), or dynamic autonomy(5), i.e., it may operate as highly autonomous during part of its operation or for performing certain tasks, and then operate in a lower autonomy level for other types of operations. An autonomous system may also be both manned and unmanned.

Many areas of life and business comprehend systems with some level of autonomy. For instance, autonomous chatbots are found on the internet, autonomous manufacturing systems are taking up production, and autonomous transportation systems are being tested on land, in water, and in the air. Although the first industrial sectors to introduce some level of autonomy into transportation were aeronautics and the aero-spatial domains, significant investments have recently fast-tracked the development of autonomous cars, and put those under the spotlight.

The rapid evolution of technology enabling autonomous cars can be illustrated by the Grand Challenge, an event organized by DARPA[1]. The Grand Challenge consisted of a competition of autonomous cars to go through California's Mojave Desert. In 2004, no car finished the race and the most successful one reached a maximum of seven miles of the course. In the following year, five vehicles completed the race. In fact, Google's first project on autonomous cars was launched in 2009 with a team from DARPA veterans. The development of autonomous cars is driven today by giants of the tech and auto industry, such as Google and Tesla, Ford, and General Motors. These are followed by smaller startups as May mobility and Drive.ai.

Autonomy is also applied in other land transportation systems, such as buses and trains. China has launched the world's first self-driving bus in August 2015. The bus drives with guidance from cameras, lidars, and a master controller, along with a human driver behind the wheel, who should take over control in case of any problems. Other examples include the Norwegian city Stavanger, where the mass-transit company is testing autonomous buses, and and Catalonia, Spain, where an autonomous bus called Èrica is being tested to help citizens become familiar with driverless technology. In Finland, three cities are expected to receive autonomous buses by 2020. The technology will be provided by the Japanese company Muji, and it should be the first autonomous bus in the world suited to all types of weather.

[1] DARPA (Defense Advanced Research Projects Agency) is an agency of the United States Department of Defense.

Land transportation on railways has also advanced using automated and autonomous systems. Automatic metros have been used for a long time – being present in over 25 cities. Highly autonomous trains' journeys, on the other hand, started in 2018 in Western Australia, by Rio Tinto Company, and were a breakthrough. The company claims that by the end of the year, the train has completed more than 1 million km autonomously with remote supervision.

The revolution of autonomous transport modes has reached the maritime sector, as well. Yara Birkeland, an autonomous and electric container vessel developed by Yara and Kongsberg, is expected to go through the first operational tests at the start of 2019, and to conduct fully autonomous operations by 2020[(6)]. DNV ReVolt, an unmanned, zero-emission, shortsea concept vessel developed by DNV GL, is being tested in a 1:20 scale, in collaboration with the Norwegian University of Science and Technology (NTNU) [(7)]. In addition, NTNU is currently testing a 1:2 scaled autonomous passenger ferry, which is expected to run on full scale in 2020 [(8)].

In aviation, automation was initially applied in military operations. The Hewitt-Sperry Automatic Airplane first flew in 1917 and was designed as a pilotless aircraft to deliver explosives during World War I. From those early flights, the aviation industry has propelled itself further, with systems such as autopilot and auto-throttle.

Discussion on autonomy in aviation ranges now from autonomous unmanned aerial vehicles (UAV) systems to pilotless commercial aircrafts. Unmanned systems are not only re-shaping transportation systems, but also allowing exploration and research of harsh remote environments with no human life exposure. The Arctic Unmanned Aircraft System Initiative of the Canadian government is testing drones to monitor Canadian Artic for oil spill, ice coverage, marine habitats and activity on the oceans[(9)]. Unmanned aircraft and remotely operated ground vehicles have also been used to monitor Japan's Fukushima nuclear power plant accident in places too dangerous for humans[(10)]. Currently, UAVs use range from policing and surveillance to product deliveries and aerial photography. Civilian UAVs now vastly outnumber military UAVs.

Autonomous Underwater Vehicles (AUVs) are also used for tasks in harsh and unstructured environments, such as for ocean monitoring, in detailed mapping of the seafloor, and for inspection of subsea infrastructure. Similarly, autonomous systems have been used in space exploration. NASA has a team responsible for developing a

suite of intelligent system technologies to extend ground support for deep-space exploration. In addition, to reduce manpower requirements and account for the time delays in communications, the International Space Station (ISS) incorporates advanced autonomous feature. These include smart sensors for failure recognition, diagnostics and prognostics, model-based reasoning for scheduling maintenance, and automation of low-level routine tasks[11].

The rapid development of the technology-enabling systems with some degree of autonomy is driven by the extensive benefits it brings to the wide range of applications above mentioned. Autonomous systems may bring enhanced solutions to city traffic, cargo transport, data collection and knowledge building of harsh environments, and space exploration. The development of autonomous system applications is, however, not without challenges.

Recent accidents have put emphasis on the need to discuss the safety aspect of these systems. The media has particularly featured recent accidents involving autonomous cars, especially the ones causing fatalities. In 2016, two accidents led to drivers' fatalities, in China and in the United States of America[12],[13]. These were followed by two accidents in 2018 in the U.S, which led to a pedestrian fatality and a driver fatality[14],[15]. More recently, in January this year, a self-driving car hit and destroyed a Promobot, an autonomous robot who was attending the Consumer Electronics Show in Las Vegas[16]. The car continued to move for 50 more meters before coming to a halt, leaving the robot non-assisted.

Promobot robot (source: @promobot instagram)

Other incidents involving autonomous systems include, among others, an autonomous bus that collided with a truck in Las Vegas in 2017, an autonomous train that crashed into a wall during a test in India in 2017, a U.S. military drone that was hijacked in 2011, among others. In light of these incidents, development of safe solutions for autonomous systems are, more than ever, crucial for their use. In particular, it is essential to:

- Recognize, understand and assess the risks involved with autonomous systems operations;
- Implement safe solutions in the design phase of these systems
- Monitor, follow up, and ensure that the risk level is acceptable during operation;
- Establish regulations and procedures that assure safe operations;
- Communicate safety to society in order to establish trust in autonomous systems.

Autonomous systems development: what are the challenges?

A common challenge concerning all autonomous systems refers to safety, reliability and security goals being met. Safety can be defined as the state where freedom from unacceptable risks is achieved, or the condition where a system is successfully operating[17]. Reliability, on the other hand, can be defined as the probability of a system or component working as intended under specified conditions for a specified amount of time[18],[19]. It is important to note that reliable systems are not necessarily safe. A reliable autonomous system may execute an action each time perfectly but, in conjunction with external circumstances, such a reliable action can lead to an accident.

The difference between reliability and safety becomes more apparent when the software used in autonomous systems is considered: The software may be executed reliably but may not be safe. For instance, instead of stopping when being operated outside its design envelope, the control software may attempt to recover the system. Similarly, a safe system is not necessarily secure. Security can be defined as the freedom from unacceptable risks being created through voluntarily actions from outside the system.[18],[19] Safety features may be exploited by hostile agents in order to gain control of or access to an autonomous system. Conversely, a secure system may be not safe for users, e.g., due to an over complicated operation.

In the following pages, we will present five key areas that can pose a challenge for SRS of autonomous systems.

Interaction of software, hardware, and human operator

One of the complexities that characterize autonomous systems is the strong interaction among its different components. These are hardware, software, computer hardware and the human operator or supervisor, when applicable. All of these interactions occur in a partially unknown and difficult to predict environment. Human operators are often seen as responsible for accidents, either by initiating them or by not responding properly in the course of events. Indeed, one of the motivations for autonomous systems development is their potential to rely less (or not rely at all) on humans for operation and, consequently, for accidents where human failure would be involved to be avoided. However, depending on the LoA of the autonomous systems, it will still rely on humans for remote control, for onboard operation in part of their task, or for monitoring.

In autonomous systems, operators may use system's functionalities out of the intended context or design envelope, or not behave as expected when their actions are required for emergency response. Their interaction with the system may, thus, voluntarily or involuntarily, jeopardize the SRS of the system. Likewise, a failure of the software may provide misleading information to operators or not provide the necessary data, thus leading to human failure. Similarly, the hardware may produce noise or faulty signals that are interpreted incorrectly by the software, which may lead to unanticipated and often unwanted effects. Software, in turn, may not work as intended and lead to faulty activation of actuators or display imprecise information, due to the discrete nature of the software – both in time and enumeration.

Finally, interactions may create vulnerabilities that can be used by malicious agents to take control of the autonomous system. The challenges regarding the SRS lie in identifying failures that may arise from this complex interaction, as well as from the propagation of those throughout the system's components and subsystems. Solving this challenge will allow for providing valuable contributions to the identification and development of efficient risk-reducing measures and SRS management strategies.

Assessment methods for safety, reliability and security

The software-hardware-human interaction discussed above is one of the main challenges for SRS assessment of autonomous systems. Most current quantitative assessment methods used in conventional risk and safety assessments rely on the separation principle. System components are assumed to be independent of each other and are often analyzed separately[20]. The interaction among components and emerging complexity is thus often neglected or reduced to a minimum. This makes it possible to use proven methods; however, complex systems may be abstracted and not sufficiently represented.

Some qualitative methods incorporate the different system elements, assessing the emerging properties and system interactions. These are, for example, STPA[21] or FRAM[22]. Such methods, while providing useful qualitative analysis, are still very limited in unravelling complex failure modes and mechanisms in addition to being qualitative and of limited value in prioritizing risks and risk reducing measures. The assessment of hardware with respect to SRS is generally well established. Mathematical approximations of failure probabilities of elements, such as engines, valves, or drive trains, are well developed and publicized. However, computer hardware is subject to different failure mechanisms and patterns and the established methods only apply to a limited extend.

For software, SRS assessments are more difficult to establish. Reliability is approximated by such measures as the remaining amount of errors in the software, which does not clarify how the software may fail. In particular, the interaction of different software components, from possibly different suppliers or development teams, is challenging. Several thousand lines of code need to be analyzed and checked for possible interactions. Risk analysis for software has been addressed recently, which is different from reliability methods[23]. Many of the commonly used approaches for software SRS assessment in the industry build on checklists and or focus on fulfilling formal requirements as proof for SRS compliance[19].

An additional challenge concerns security assessment of AS, including, but not limited to cybersecurity. New threats and vulnerabilities may emerge with autonomous systems. The complexity of autonomous systems may mask vulnerabilities, and attackers may use the complexity to hide their intrusion or access.

The assessment of still unencountered threats, malicious intentions and attackers is a key step for addressing security[24].

Autonomous systems are complex, with emerging properties from the interactions of the systems' components. Therefore, a holistic approach is required for the SRS assessment, considering the possible interactions and their potential outcomes and implications[22].

Cyber security

Cyber security and security in general may be one of the major challenges concerning autonomous systems. The autonomous behavior may be exploited, and passengers and goods may be endangered. Security addresses the malicious exploitation of vulnerabilities through threat-agents to cause harm or benefit from it. This is often connected with hacking, where software vulnerabilities are abused and the attacker accesses the target system to control it or extract information[25]. Hardware hacking is another method to access a computerized system. Microchips or micro computers are introduced in the system and allow an attacker to access the computer system[26].

Practices and components that can create vulnerabilities are shared among different types of autonomous systems, for example, communication protocols between components that have been developed many years ago and do not have any security mechanisms. Vulnerabilities may also arise from poorly-integrated system components, wireless communication and/ or entertainment systems, remote monitoring systems, inadequately trained machine learning systems[24],[27]–[31]. A cyber-attack may not always target the autonomous system itself. A ransom ware or a virus may inflict collateral damage to the autonomous system and disable it.

Although autonomous systems may not have an email address or allow downloading of files, the user or operator may connect to the system using his/ her own device. This may open the system for intrusion or give access to malware[32]. Another aspect of cyber security for autonomous systems is jamming and spoofing of sensor systems[24]. A jammed sensor is not able to fulfil its function due to a disturbing signal that disables it. A spoofed sensor, on the other hand, will produce fake signals. Jamming and spoofing may affect, among others, visual sensors, radio wave sensors

and global navigation satellite systems. It has been demonstrated that by jamming and spoofing autonomous systems can be hijacked and stolen[25],[30].

Autonomous Systems should be developed having in mind these vulnerabilities. A sound cyber security management system is required from early development stages on.

Legal and regulatory aspects

Legal and regulatory aspects may be particularly challenging for unmanned autonomous transport systems. Transport systems are regulated to, above all, assure their safety regarding communities, users and drivers. However, these regulations, when developed, did not contemplate autonomy being introduced in these systems. Regulators are thus facing the challenge of developing or adapting existing regulations to accommodate autonomous and semi-autonomous vehicles (AVs); and to keep up with the pace of technology development. Developers, on the other hand, face the challenge of demonstrating and communicating safety of their systems to regulators.

Autonomous ships are a current example of the abovementioned challenges. Ship operations are broadly regulated by the International Maritime Organization (IMO)[2]. Although having a centralized regulation scheme brings uniformity of regulatory approach, IMO regulations also move slowly. One of the legal issues is the safe manning requirements applicable to merchant vessels. Several conventions require that vessels shall be properly manned to maintain a safe lookout, which is a challenge for unmanned autonomous ships.

In general, such requirements may demand major adaptations within current regulations. For instance, the autonomous bus to be adopted in Stavanger, Norway, will have to operate with an employee onboard, in order to comply with Norwegian legislation. This employee must be able to manually override the autonomous controls with a brake button if a dangerous situation occurs.

Road traffic is generally regulated by The Vienna Convention on Road Traffic[33], an international treaty, since 1968. The convention initially stipulated that a human driver must always remain fully in control of and be responsible for the behavior of their vehicle in traffic. The treaty has been signed and ratified by 75 countries, and

[2]IMO develops guidelines, and those are implemented and enforced by each member state.

examples of non-signatory countries include the United States and China. The fact that the U.S. is not a signatory, combined with the possibility of federal states establishing their own legislation, may have influenced that it was one of the pioneers in legislation for autonomous cars. Nevada was the first US state to authorize the operation of autonomous vehicles, in 2011. Since then, 21 other states have passed legislation related to autonomous vehicles. Recently, the US National Highway and Transportation Safety Administration (NHTSA) released new federal guidelines for automated driving systems (ADS). It has a voluntary nature, without compliance requirement or enforcement mechanism.

On December 2016, an act implementing an amendment to the Vienna Convention on Road Traffic entered into force in Germany[34]. The amendment allows the transfer of driving tasks to the vehicle itself, provided that the technologies used are in conformity with the United Nations vehicle regulations or can be overridden or switched off by the driver. Once again, a licensed driver is required to be behind the wheel to take control if necessary.

Liability is another challenge in regulating AV. Who should be responsible when an accident happens? Will anti-collision algorithms developers be responsible when a collision occurs? To what extent is the remote driver or supervisor responsible in case s/he does not act in time to override an action from a mal-functioning system?

In addition to the questions above, some ethical aspects must be assessed in terms of liability. For instance, in the U.S., the income of the victim is related to her/his liability damages – the more someone earns, the greater her/his liability exposure. To protect themselves against major liability claims, AV manufacturers may adjust the car's driving behavior according to the average income in an area[35]. The problem of regulations for autonomous vehicles comes with a catch-22: we need to test and use AVs to assess their safety; yet we do not want them on the road / ocean / sky until we know that they are safe.

Ethical and social aspects

"Never in the history of humanity have we allowed a machine to autonomously decide who should live and who should die, in a fraction of a second, without real-time supervision. We are going to cross that bridge any time now, and it will not happen in a distant theatre of military operations; it will happen in that most mundane aspect of our lives, everyday transportation." (36)

The above quote is retrieved from the report of the developers of the Moral Machine[3]. The experiment, launched in a website, was developed to collect large-scale data on how people would want autonomous vehicles to solve moral dilemmas. The interest in the platform was significant, and they collected almost 40 million decisions from nearly all countries of the world. The experiment presents users with an unavoidable accident scenario, and offers them the choice of the car to swerve or stay in course. The outcome of this choice is to spare one group over the other during a collision; for instance, if the car stays in course it may run over pedestrians, and if swerving it will collide with a fixed object and danger the passengers. They collected decisions data over nine main factors, as sparing men versus women, or humans versus pets.

The type of choice the users confronted in the Moral Machine follows the framework of the trolley cases, and has been addressed by ethics researchers on analyzing autonomous cars. The choice on who to harm in case of unavoidable accidents is a necessary question regarding the development of autonomous vehicles. Should this decision be fixed and embedded in the algorithms during development? Will cars use machine learning and "replicate" human-alike decisions? These questions become more difficult to address given the results of the Moral Machine experiment. Although there were some consensuses regarding some dilemmas, as sparing humans over animals; significant socio-geographical differences arose when dealing with other choices. For instance, a preference to spare younger characters / people is less pronounced in far eastern countries and in some Islamic countries, and higher in Latin America. The same is true for the preference in sparing higher status characters(36).

[3] http://moralmachine.mit.edu/

Imitating human drivers' behavior for establishing moral decisions is, thus, a challenge given the socio-geographical differences. In addition, humans may show unethical biases when driving, such as deciding whether to yield at crosswalks based on pedestrians' race and income[37]. Ethics of autonomous vehicles are not restricted to the trolley problem[38]. Mundane traffic situations, such as approaching a crosswalk with limited visibility, making a turn, navigating through busy intersections, or factors related to how liability is determined raise important ethical question[35].

The first and only attempt so far to provide official guidelines for the ethical choices of autonomous vehicles is the German Ethics Commission on Automated and Connected Driving[39]. One of the rules states that, in a dilemma, protection of human life should have priority over other animals' life. Another rule affirms that distinction based on personal features such as age, should be prohibited. How ethics and moral are implemented on AS will influence its societal acceptance. People's willingness to buy autonomous vehicles and tolerate them on the roads will depend on the palatability of the ethical rules that are adopted. In addition to moral aspects, trust in autonomy is an important factor for societal acceptance. Trust in automation is a highly discussed subject in the human factors and human reliability community.

In short, autonomy creates a new depth in the human-machine relationship from the users side, the operators that supervise it or remotely control it, and the people interacting with the autonomous systems externally. Communicating safety to society is thus a must to gain trust in autonomy and societal acceptance.

Call for discussion

Addressing all the issues mentioned in this whitepaper is not trivial and requires interdisciplinary and international cooperation. We believe that, as a community composed of different types of expertise related to autonomy, we can work together to create safe, reliable and secure autonomous systems for a safer, cleaner, and more efficient future.

Welcome to IWASS and to Trondheim!

Velkommen til IWASS og Trondheim!

Marilia Ramos, PhD

Dr. Marilia Ramos has a PhD in Chemical Engineering from the Federal University of Pernambuco, Brazil. Her expertise is on Risk Analysis and Human Reliability, and she has extensive experience in projects concerning the oil and gas field. Currently she is a postdoctoral research fellow at the Department of Marine Technology, NTNU, and applies risk and reliability analysis to autonomous surface vessels. Her main research interest is on the human-software-hardware interaction in such systems.

Christoph A. Thieme, PhD

Dr. Christoph Thieme obtained his PhD in Marine Technology from NTNU. He has experience with risk analysis and modelling of autonomous marine systems. Currently, he is a postdoctoral research fellow at NTNU in the UNLOCK project, working on risk assessment methods development and applications on autonomous control systems.

Ingrid B. Utne, PhD

Dr. Ingrid Bouwer Utne is a Professor in marine operation and maintenance at Department of Marine Technology, NTNU. Utne is an affiliated Researcher in the Center of Excellence on Autonomous Marine Operations and Systems (NTNU AMOS) where she is managing the research/industry projects UNLOCK and ORCAS. These projects focus on supervisory risk control and bridge the scientific disciplines of risk management and engineering cybernetics aiming to enhance safety and intelligence in autonomous systems.

Ali Mosleh, PhD

Dr. Ali Mosleh is Distinguished University Professor and holder of the Knight Endowed Chair in Engineering at the University of California in Los Angeles (UCLA), where he is also the director of the Institute for the Risk Sciences. He is also honorary professor at several universities in Europe and Asia. He conducts research on methods for probabilistic risk analysis and reliability of complex systems and has made many contributions in diverse fields of theory and application. He was elected to the US National Academy of Engineering in 2010 and is a Fellow of the Society for Risk Analysis, and the American Nuclear Society. Prof. Mosleh is the recipient of many scientific achievement awards, and has been a technical advisor to numerous international organizations.

References

1. Vagia M, Transeth AA, Fjerdingen SA. A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Appl Ergon*. 2016;53:190-202. doi:10.1016/j.apergo.2015.09.013
2. Clough BT, Leader TA, Force A, Afb W. FAO backs biotech crops. *Inf - Int News Fats, Oils Relat Mater*. 2004;15(7):438.
3. Sheridan TB, Verplank W. *Human and Computer Control of Undersea Teleoperators*. Cambridge; 1978.
4. Sheridan TB. Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. *IEEE Trans Syst Man, Cybern Part ASystems Humans*. 2011. doi:10.1109/TSMCA.2010.2093888
5. Laurinen M. Remote and Autonomous Ships: The next steps. *AAWA Adv Auton Waterborne Appl*. 2016:88. http://www.rolls-royce.com/~/media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/aawa-whitepaper-210616.pdf.
6. Kongsberg. Autonomous ship project, key facts about YARA Birkeland. https://www.km.kongsberg.com/ks/web/nokbg0240.nsf/AllWeb/4B8113B707A50A4FC125811D00407045?OpenDocument. Published 2017.
7. Alfheim H, Muggerud K. Development of a Dynamic Positioning System for the ReVolt Model Ship Henrik Alfheim. 2017.
8. NTNU. Autoferry – Autonomous all-electric passenger ferries for urban water transport. https://www.ntnu.edu/autoferry. Published 2018. Accessed August 13, 2018.
9. Transport Canada. Drones in the Canadian Arctic. https://www.tc.gc.ca/en/programs-policies/programs/national-aerial-surveillance-program/drones-canadian-arctic.html. Accessed February 1, 2019.
10. Lillian B. Engineers Developing Drones to Inspect Fukushima Daiichi Nuclear Disaster. https://unmanned-aerial.com/engineers-developing-drones-to-inspect-fukushima-daiichi-nuclear-disaster. Accessed February 1, 2019.
11. Committee for NASA Technology Roadmaps Aeronautics and Space Engineering Board Division on Engineering and Physical Sciences. *Autonomy Research for Civil Aviation Torward a New Era of Flight*. Washin; 2014.
12. Boudette NE. Autopilot Cited in Death of Chinese Tesla Driver. *The New York Times*. https://www.nytimes.com/2016/09/15/business/fatal-tesla-crash-in-china-involved-autopilot-government-tv-says.html. Published September 14, 2016.
13. Singhvi A, Russell K. Inside the Self-Driving Tesla Fatal Accident. *The New York Times*. https://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html. Published 2016.
14. T.S. Why Uber's self-driving car killed a pedestrian. *The Economist*. https://www.economist.com/the-economist-explains/2018/05/29/why-ubers-self-driving-car-killed-a-pedestrian. Published 2018.

15. Ohnsman A. Fatal Tesla Crash Exposes Gap In Automaker's Use Of Car Data. *Forbes*. https://www.forbes.com/sites/alanohnsman/2018/04/16/tesla-autopilot-fatal-crash-waze-hazard-alerts/. Published April 16, 2018.
16. Cozzens T. Autonomous car hits autonomous robot in bizarre collision. https://www.gpsworld.com/autonomous-car-hits-autonomous-robot-in-bizarre-collision/. Accessed February 1, 2019.
17. Woods DD. Essential characteristics of Resilience. In: Hollnagel E, Woods DD, Leveseon NG, eds. *Resilience Engineering -Concepts and Precepts*. 1st ed. Surrey, UK; Burlington, USA: Ashgate; 2006:21-34.
18. Rausand M. *Risk Assessment - Theory, Methods, and Applications*. 1st Ed. (Barnett WJ, ed.). Hoboken, New Jersey, USA: John Wiley & Sons; 2011.
19. ISO, IEC. *ISO/IEC Guide 51: Safety Aspects - Guidelines for Their Inclusion in Standards*. Geneva, Switzerland; 2014. www.iso.org.
20. Mosleh A. PRA: A Perspective on strengths, current Limitations, and possible improvements. *Nucl Eng Technol*. 2014. doi:10.5516/NET.03.2014.700
21. Leveson NG, Thomas JP. *STPA Handbook*. 1. Cambridge, MA, USA; 2018.
22. Hollnagel E. *FRAM – The Functional Resonance Analysis Method*. 1st Ed. Farnham. UK: Ashgate; 2012.
23. Aldemir T, Guarro S, Mandelli D, et al. Probabilistic risk assessment modeling of digital instrumentation and control systems using two dynamic methodologies. *Reliab Eng Syst Saf*. 2010. doi:10.1016/j.ress.2010.04.011
24. Petit J, Shladover SE. Potential Cyberattacks on Automated Vehicles. *IEEE Trans Intell Transp Syst*. 2015. doi:10.1109/TITS.2014.2342271
25. Yağdereli E, Gemci C, Aktaş AZ. A study on cyber-security of autonomous and unmanned vehicles. *J Def Model Simul*. 2015. doi:10.1177/1548512915575803
26. Wyglinski AM, Huang X, Padir T, Lai L, Eisenbarth TR, Venkatasubramanian K. Security of autonomous systems employing embedded computing and sensors. *IEEE Micro*. 2013. doi:10.1109/MM.2013.18
27. Bothur D, Zheng G, Valli C. *A Critical Analysis of Security Vulnerabilities and Countermeasures in a Smart Ship System*.; 2017. http://ro.ecu.edu.au/ism/209.
28. Haas RE, Möller DPF. Automotive connectivity, cyber attack scenarios and automotive cyber security. In: *IEEE International Conference on Electro Information Technology*.; 2017. doi:10.1109/EIT.2017.8053441
29. Hassani NTNU V, Ocean S, Trondheim ntnuno, AntónioAnt P, Pascoal AM. *CYBER SECURITY ISSUES IN NAVIGATION SYSTEMS OF MARINE VESSELS FROM A CONTROL PERSPECTIVE*.; 2017. http://proceedings.asmedigitalcollection.asme.org.
30. Parkinson S, Ward P, Wilson K, Miller J. Cyber Threats Facing Autonomous and Connected Vehicles: Future Challenges. *IEEE Trans Intell Transp Syst*. 2017. doi:10.1109/TITS.2017.2665968
31. Vinnem JE, Utne IB. Risk from cyberattacks on autonomous ships. In: *Safety and Reliability – Safe Societies in a Changing World*. ; 2018. doi:10.1201/9781351174664-188
32. Cárdenas AA, Amin S, Sinopoli B, Giani A, Perrig A, Sastry S. *Challenges for Securing Cyber Physical Systems*.

33. UNITED NATIONS CONFERENCE ON ROAD TRAFFIC. *Convention on Road Traffic*. Vienna; 1968.
34. Amendments to Article 8 and Article 39 of 1968 Convention on Road Traffic. 2016;2016(34):1306-1308.
35. Himmelreich J. Never Mind the Trolley: The Ethics of Autonomous Vehicles in Mundane Situations. *Ethical Theory Moral Pract*. 2018;21(3):669-684. doi:10.1007/s10677-018-9896-4
36. Awad E, Dsouza S, Kim R, et al. The Moral Machine experiment. *Nature*. 2018;563(7729):59-64. doi:10.1038/s41586-018-0637-6
37. Coughenour C, Clark S, Singh A, Claw E, Abelar J, Huebner J. Examining racial bias as a potential factor in pedestrian crashes. *Accid Anal Prev*. 2017;98:96-100. doi:10.1016/j.aap.2016.09.031
38. Roff H. The folly of trolleys: Ethical challenges and autonomous vehicles. *Brookings*. https://www.brookings.edu/research/the-folly-of-trolleys-ethical-challenges-and-autonomous-vehicles/. Published 2018.
39. Federal Ministry of Transport and Digital Infrastructure of Germany. *Ethics Commission on Automated and Connected Driving*.; 2017. doi:10.1126/science.186.4158.38

Organizers and sponsors

NTNU Norges teknisk-naturvitenskapelige universitet

Department of Marine Technology, Norwegian University of Science and Technology, Trondheim, Norway

The Norwegian University of Science and Technology in Trondheim (NTNU) is the largest university focusing on Science and Technology in Norway. NTNU has an international focus and has world-leading research in several science disciplines. The Department of Marine Technology at NTNU is a world leader in education, research, and innovation for engineering systems in the marine environment. It is specialized in methods and techniques that facilitate the assessment, development, and sustainable operation of oil and gas extraction at sea, ship technology, fisheries and aquaculture technology, offshore renewable energy, and marine robotics for mapping and monitoring the ocean and polar environment.

The B. John Garrick Institute for the Risk Sciences
UCLA ENGINEERING

The B. John Garrick Institute for the Risk Sciences, University of California, Los Angeles, USA

The B. John Garrick Institute for the Risk Sciences has declared its mission to be the advancement and application of the risk sciences to save lives, protect the environment and improve system performance. The purpose of the Garrick Institute is for the research, development, and application of technology for (1) quantifying the risk of the most serious threats to society to better enable their prevention, reduce their likelihood of occurrence or limit their consequences and (2) improving system performance with respect to reliability and safety. The institute is hosted at the Department of Engineering at the University of California in Los Angeles.

DNV GL

DNV GL is a global quality assurance and risk management company. DNV GL provides classification, technical assurance, software and independent expert advisory services to several industries. Combining technical, digital and operational expertise, risk methodology and in-depth industry knowledge, DNV GL assists its customers in decisions and actions with trust and confidence. With origins stretching back to 1864 and operations in more than 100 countries. DNV GL are dedicated to helping customers make the world safer, smarter and greener.

Rolls Royce Marine

Rolls Royce Marine is a leading supplier of offshore and marine energy solutions, deck machinery and automation systems. In addition, RRM provides services related to complex system integration, and vessel design. RRM is a leader in marine ship intelligence, automation and autonomy, testing successfully a fully autonomous ferry transit in Finland in late 2018. RRM is now a part of the Kongsberg Group.

Research Council of Norway

The Research Council of Norway serves as the chief advisory body for the government authorities on research policy issues, and distributes roughly nine billion Norwegian kroners to research and innovation activities each year. The Research Council of Norway co-financed the IWASS workshop through the MAROFF knowledge-building project for industry ORCAS (Project number 280655) and the FRINATEK project UNLOCK (Project number 274441).