SAFETY & RELIABILITY

SAFE SOCIETIES IN A CHANGING WORLD

Editors

Stein Haugen, Anne Barros, Coen van Gulijk,
Trond Kongsvik & Jan Erik Vinnem

CRC Press
Taylor & Francis Group
A BALKEMA BOOK

PROCEEDINGS OF THE 28TH INTERNATIONAL EUROPEAN SAFETY AND RELIABILITY CONFERENCE (ESREL 2018), TRONDHEIM, NORWAY, 17–21 JUNE 2018

Safety and Reliability – Safe Societies in a Changing World

Editors

Stein Haugen & Anne Barros
NTNU, Faculty of Engineering, Trondheim, Norway

Coen van Gulijk
University of Huddersfield, Faculty of Computing and Engineering, Huddersfield, UK

Trond Kongsvik
NTNU, Faculty of Economics and Management, Trondheim, Norway

Jan Erik Vinnem
NTNU, Faculty of Engineering, Trondheim, Norway

CRC Press
Taylor & Francis Group
Boca Raton London New York Leiden

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A BALKEMA BOOK

CRC Press/Balkema is an imprint of the Taylor & Francis Group, an informa business

Typeset by V Publishing Solutions Pvt Ltd., Chennai, India
Except:
'Failure Mode Effects & Criticality Analysis (FMECA) using Bayesian Dirichlet-multinomial conjugate pair' by W. Baun

'Lessons learned from an unexpected uranium accumulation event' by D.G. Harrison & A. Smith

Published by: CRC Press/Balkema
Schipholweg 107C, 2316 XC Leiden, The Netherlands
e-mail: Pub.NL@taylorandfrancis.com
www.crcpress.com – www.taylorandfrancis.com

ISBN: 978-0-8153-8682-7 (Hardback)
ISBN: 978-1-351-17466-4 (eBook)

Safety and Reliability – Safe Societies in a Changing World – Haugen et al. (Eds)
© 2018 Taylor & Francis Group, London, ISBN 978-0-8153-8682-7

Risk-based regulation and certification of autonomous transport systems

S.O. Johnsen, Å. Hoem & T. Stålhane
Faculty of Information Technology, NTNU, Trondheim, Norway

G. Jenssen & T. Moen
SINTEF Technology and Society, Trondheim, Norway

ABSTRACT: Autonomous transport systems in all modes—road (i.e. autonomous cars), aviation (i.e. drones), shipping and rail are coming. Regulation and testing are on-going in Norway. Risks of autonomous systems are uncertain due to missing data, emerging technology and variation in framework conditions. However, accidents of autonomous cars seem to be 1/3 or 1/2 of current levels. Incidents are different, needing outside interventions sometimes. Based on review of experiences across the modes and regulations, we suggest agile and transparent learning in the whole autonomous ecosystem, between all modes. System certification are needed, and system responsibilities must be clarified. Structures for orchestrating transport (i.e. control of many autonomous vehicles with possible common failures) and marking autonomous transport, should be established. In the interfaces between humans and systems there are differences in autonomy as imagined vs. performed, leading to new incidents and accidents. Emerging safety/security issues must be explored.

1 INTRODUCTION

This paper discusses experiences of autonomous transport systems, to establish a framework for risk based governance. Risk and risk governance are based on the process described by Renn (2005), starting with problem framing; risk appraisal (hazards and vulnerabilities); risk judgment; risk communication and risk management. The implementation of autonomy can reduce transport risks but it can also introduce new risks in the interfaces between the autonomous system and the environment (such as humans). As discussed in Lund and Aarø (2004), risk reduction must be based on a broad set of actions such as regulation, technical design, training and awareness.

Based on involvement in the regulatory process in Norway and experiences of autonomous transport systems we have discussed new emerging risks and threats. We see the need for establishing framework such as regulatory actions and clarification of responsibilities as autonomy is being implemented.

In the following we have defined autonomous systems and concepts such as Levels of Automation (LOA) used to specify degree of automation.

1.1 *Definitions and terminology*

Safety is related to accidental harm, while security is related to intentional harm. Safety is defined as: "*the degree to which accidental harm is prevented, reduced and properly reacted to*", Firesmith (2003). Security: "*the degree to which malicious harm is prevented, reduced and properly reacted to*".

In Parasuman and Riley (1997) automation and autonomy is described as "*The execution by a machine agent (usually a computer) of a function that was previously carried out by a human*". Automation can be done by various means i.e. 1: Remote controlled (Surveyed and/or externally controlled); 2: Autonomous (based on own sensors and systems); 3: Cooperative and connected (based on own sensors and other traffic information) or 4: A combination of 1–3. The terms autonomous and automated has been used interchangeably in some papers. We have made a distinction. By *autonomy* we mean a system that is non-deterministic in that it has a freedom to make choices, and by *automated* we mean a system that is more deterministic in that it will do exactly what it is programmed to do. This is based on the taxonomy and discussion of autonomy from Vagia et al. (2016).

When trying to scope risks of autonomous systems we must include the regulation, risk governance, organizational framework, interfaces to humans and the autonomous system (a combination of software components and cyber physical systems). The system is often a collection of systems being developed by different stakeholders. Thus, we have used the concept of autonomous

ecosystem, AEC. This is inspired by the concept Software Ecosystems (SEC). SEC consists of components developed by actors both internally and externally of the company, i.e. outside the traditional borders to a group of private persons and actors. Manikas et al. (2013) defined a software ecosystem as: "*the interaction of a set of actors on top of a common technological platform that results in a number of software solutions or services. Each actor is motivated by a set of interests or business models and connected to the rest of the actors and the ecosystem as a whole with symbiotic relationships, while, the technological platform is structured in a way that allows the involvement and contribution of the different actors...*". Arguments for using such a concept is the realization that development increasingly is taking place outside of organisational silos due to the need for speed of development, need for supporting applications, reduction of development costs, competition. This is creating the need to address governance challenges in an ecosystem framework.

An example of an autonomous ecosystem is Intelligent Transport Systems (ITS) consisting of autonomous vehicles, integrated with traffic control, electronic payments and other systems. Autonomous ecosystems handle information, but also actual critical processes such as transport (via automobiles, boats, drones and trams). These ecosystems must be safe and secure. The systems must be able to handle unanticipated events, breakdowns and be able to go to a safe and secure (end-)state.

To explore the main risks of autonomous systems, we need to clarify responsibilities i.e. LOA in task execution. LOA is described by steps going from no automation where the humans are fully in control to a fully automated system with no human interaction. Sheridan and Verplank (1978) introduced 10 steps of automation, going from LOA1: Fully Manual Control to LOA10: Fully Autonomous Control. The LOA has been adapted to the car industry by the Society of Automotive Engineers (SAE), describing six levels of autonomy in driving, SAE (2016). Going from no autonomy (level 0), through driver assistance, partial automation, conditional automation, high automation, to full automation (level 5). The design of the autonomous transport system must ensure that the system maintains an accepted level of performance despite disturbances, including threats of an unexpected and malicious nature. Our approach is to speed up learning and knowledge sharing between modes, since the autonomous systems have different maturity and experiences in aviation, rail, road and sea.

The concept of resilience engineering is an important strategy to handle unanticipated incidents. Hollnagel, Woods and Leveson (2006) define resilience as "*the intrinsic ability of a system to adjust its functioning prior to or following changes and disturbances, so that it can sustain operations even after a major mishap or in the presence of continuous stress*". Handling of unanticipated incidents and continue to operate safe is a key ability of autonomous transport systems.

Based on the preceding introduction, the research questions (RQ) we want to explore are:

- RQ1: What are the major risks introduced by autonomous transport systems?
- RQ2: What regulatory issues should be prioritized to handle these risks?
- RQ3: What are the way forward, i.e. main approaches and issues needed to mitigate major risks of autonomous transport systems?

2 SCOPE, CHALLENGES AND METHODS

When discussing autonomous ecosystems, we include the organisational framework, regulation, human interactions and understanding in addition to the actual systems in autonomous systems and the infrastructure. This is described in Figure 1.

2.1 *Challenges and problems*

When introducing new technology such as autonomous systems, one of the basic challenges is to understand emerging risks. Safety, security and resilience have often been identified late when vulnerabilities have been exploited and unwanted incidents have been published. There has been a tradition in the software industry that vendors seldom have to pay for these unwanted incidents even if they are due to poor quality, poor focus on safety, security or resilience. The consequences and costs have been given to users, organisations and society. In autonomous transport, the consequences can be loss of lives and/or environmental damage. In addition, when discussing vulnerabilities in autonomous ecosystem, one challenge is that there is not one single supplier, but a set of suppliers involved. It can be difficult to identify responsibilities and manage competencies, if framework conditions (regulation/responsibilities) are missing.

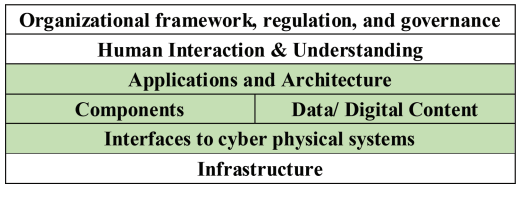

Figure 1. Scope of autonomous ecosystems—AEC.

2.2 *Methodology and approach*

We have based this paper on empirical data from users of autonomous transport systems, a targeted literature review of autonomy and safety in addition to discussion of suggested regulation of autonomous road transport in Norway.

We have explored experiences of autonomous transport systems from St. Olav Hospital in Norway, where autonomous systems have been used from 2006 to 2017. St. Olav has 10,500 employees, and covers an area of 200,000 M^2. We are involved in pilot projects with self—driving shuttle busses in three Norwegian cities. Trials addressing feasibility of Mobility as a service (MAAS) linking up to public transport (first and last mile). We are involved in trials with eco-friendly autonomous ships/vessels for cargo and passenger travel along the Norwegian coastline.

We have performed a literature review based on a keyword search of autonomy, safety, security and resilience using SCOPUS, ACM Digital Library, IEEE Explore, Springer Link and Science Direct.

We have been involved in a hearing of regulation related to testing of autonomous vehicles in Norway from the Ministry of Transport and Communications—MTC (2016). The suggested regulation was distributed in December 2016, comments to be given within March 2017 and the regulation were proposed to be approved as law in December 2017. Our comments were based on the literature review, experiences from St. Olav's and other public comments.

The taxonomy used to register incidents has been based on Blanco et al. (2016). They collected a broad set of naturalistic accident data from autonomous driving, using a taxonomy of crash seriousness going from most serious at C1 to negligent at C4.

C1: Crashes with airbag deployment, injury (needing doctor visit), rollover, more damage than $1,500, require towing, police reportable.
C2: Minimum of $1,500 worth of damage, crashes such as large animal strikes and sign strikes.
C3: Crashes involving physical conflict with another object, but with minimal damage. Includes most road departures, small animal strikes, all curb and tire strikes potentially in conflict with oncoming traffic and with higher risk potential if no curb.
C4: Tire strike only with little or no risk element (e.g., clipping a curb during a tight turn), considered to be of such minimal risk that most drivers would not consider these incidents to be crashes.

3 RESULTS AND DISCUSSIONS

In the following section, we have documented experiences from autonomous systems at St. Olav Hospital; some selected findings from our literature review; and key issues discussed during regulation of testing of autonomous transport systems.

3.1 *Findings from autonomous systems at St. Olav*

St. Olav Hospital has installed an automated transport system called Transcar LTC2 Automated Guided Vehicle System (AGV) from Swisslog. They installed seven AGVs in 2006, and additional 14 AGVs at the end of 2009. From 2010 to 2017 they have had 21 AGVs in operations. Each week the 21 AGVs transport medicine, food, clothes and garbage, in total 70–80 tonnes. (Each AGV can transport a load of 500 kg, and is transporting 3.6 tonnes each week). The speed is slow, moving at approximately 2 km/hour (maximum speed is 5 km/h). The AGVs can send signals, open doors, and reserve elevators to deliver goods. There are different suppliers of door and elevator automation. When there are conflicts that cannot be resolved, a signal is given to the operational centre. The centre is manned by an operator that can intervene through the system, or go to the place where there is a conflict.

The AGVs can communicate (i.e. deliver pre-programmed messages) such as "Please move—you are in my way", or "Elevator is reserved—please move out of elevator". A key issue related to the awareness building between automated transport systems and humans are the above-mentioned communication from the AGVs, supporting the understanding that the automated system need to inform the bystanders about their perceptions and what they are going to do next, that helps staff, patients and visitors to learn to interact with the AGV's and to anticipate their behaviour.

In the Transcar LTC2 Operations and Maintenance manual it is written "*Always maintain a distance of 1.5 meters between the vehicles and people or objects.*" This safety guideline is not possible to implement at St. Olav due to space limitations.

There are traces on the floor indicating that the AGVs are always following the same pathway, thus (new) common failures may happen.

There has been a total of 100–130 minor incidents per year (5–6 per AGV) categorised as C4 by us. Minor repairs are done on the AGVs, changing around 50 components per year. There are around 15 emergency stops each year, categorized as C3, where components must be changed. We do not have data indicating that there has been any incidents of category C2 or C1. Reported incidents are minor crashes due to faulty navigation, for example due to objects placed in the route travelled that is not detected.

When interviewing the users some incidents that can be generalized were reported:

- The AGVs have problems with pallets close to the walls. The AGV uses the wall as reference in steering. A misplaced pallet results in a lateral shift of the AGV position and may sometime end up with a collision. Initially the operators used a great deal of time to clear the transport road area (in the basement) from clutter (i.e. parked bicycles, pallets with supplies); this work has been reduced now—but maintenance and design should take into accord these limits of AVGs.
- The AGV collided several times with the forklifts, since the LiDAR sensor (light detection and ranging) had a limited vertical field of view and was seeing a free zone (space) under the forklift. This was mitigated by placing a black rubber skirt under the forklift. The same kind of collisions happened when using stepladders on the floor in the AGVs pathway, since the LiDAR did not detect the object. Thus, one issue has been the ability to see and identify objects in relation to the AGVs sense of its own size and position. This may be a general challenge with autonomous transport systems. The death accident of Joshua Brown, described by NTSB (2017) and NHTSA (2017), was between a Tesla and a trailer crossing the road—a white trailer giving poor contrast and with substantial height above the ground. Some similarity with the forklift problems at St. Olav. A rubber skirt under the trailer may have increased visibility/visual signal of the trailer.
- The AGVs can open doors, reserve and use elevators. Sometimes there has been conflicts between the AGVs and the users, needing human intervention through a central control.
- Software updates of AVGs, elevators and doors has led to interface problems, thus there is a need to look at the AVGs as a part of an ecosystem.

During the 11 years' operating the AGVs there has been no reporting of human injuries at St. Olav. However, at the AHUS hospital (AHUS, 2009), with the same system—one incident happened in 2009, where a nurse sustained a minor injury when colliding with the AGV (i.e. category C3 or C4).

In summary, the AGV system has had an impressive safety record at St. Olav's Hospital. Key issues of safe operations are related to an ecosystem approach planning the interaction between technology, organisation and humans. Based on preparation trough pilots; low speed; communication between automated systems and humans to inform surrounding people of the AGV's intended behaviour. The unexpected may happen, thus there was a need to establish a manned control centre that can intervene during operations.

3.2 *Key findings from literature review*

In Axelrod (2014) the focus is on software assurance of safety-critical and security-critical systems. The perception is that use of the current methods has not achieved the wished-for level of protection, and that there are missing security principles and standards. There seems a need for incentives or regulations to implement protective and immunizing measures in software. A requirement could be that these measures are included in a certification process. On governance, it is suggested to establish software assurance standards at the United Nation (UN) level; to have a risk based approach; to share best of breed methods; and the need to discuss liabilities for damages occurring because of an attack or security-related errors.

International governance of security of the infrastructure is addressed through several channels such as standard bodies (i.e. ISO, IEC) and international bodies such as OECD, EU, NATO and UN. Autonomous systems are international—involving many actors with different agendas. In GCIG (2016) there is a discussion of governance of emerging technology as it is integrated into critical infrastructure, such as transport systems. It is suggested that manufacturers should follow the principle of privacy and security by design, when developing new products. They must be prepared to accept legal liability for the quality of the technology they produce. Buyers should collectively demand that manufacturers respond effectively to concerns about privacy and security. Governments can play a positive role by incorporating minimum security standards in their procurement. It is suggested that government regulations should require routine, transparent reporting of technological problems to provide the data required for a transparent market-based cyber-insurance industry. It is suggested to establish an agreement (a compact) based on collaboration between government, industry and private society supporting this evidence based decision making.

In Koscher et al. (2010) vulnerabilities in cars are pointed out, such as the possibility to control a wide range of automotive functions and completely ignore driver input from dashboard, including disabling the brakes, selectively braking individual wheels on demand, stopping the engine, and so on. Attacks were easy to perform and the effects were significant. It is possible to bypass rudimentary network security protections within the car, and perform attack that embeds malicious code in the car that will completely erase any evidence of its presence (after a crash). There is a discussion of the challenges in addressing these vulnerabilities in the existing ecosystem.

In Lima, et al. (2016) semi-autonomous and fully autonomous cars are described as coming

from the development stage to operations. The autonomous systems are creating safety and security challenges. These challenges require a holistic analysis, under the perspective of ecosystems of autonomous vehicles. These systems will become important critical information infrastructures, simultaneously featuring connectivity, autonomy and cooperation. Threat analyses and safety cases should include both (random) faults and (purposeful) attacks.

In DHS (2015), there is a discussion of Cyber-Physical infrastructure risks in the future smart cities. Several examples of unwanted incidents are described in transportation systems (i.e. autonomous vehicles; trains) in electricity distribution and management and in water and wastewater systems sector. It is suggested to the regulator to work with standards and regulations in addition to communication and increased engagement by giving direct assistance. Challenges mentioned are the need to establish goal based standards and regulations as new technology is implemented and to focus on dissemination of best practices and systematic education.

In (Cerrudo, 2015) there is an empirical evaluation of "smart cities" looking at a broad set of technologies of traffic control, management of energy/water/waste and security. Known vulnerabilities are in traffic control systems, mobile applications used by citizens, smart grids/smart meters and video cameras. The issues are lack of cyber security testing and approval, lack of encryption, lack of City Computer Emergency Response Teams (CERT), and lack of cyber-attack emergency plans. There are reasons to anticipate that we establish potential for serious incidents, if these issues are not addressed and mitigated.

In Frei (2010) there is a discussion of the security dynamics of general software ecosystem (SEC), applicable to autonomous ecosystems. They examine 27000 vulnerabilities in the decade (1996–2008). The paper explores several policies such as security through obscurity, responsible disclosure of vulnerabilities (a suggested policy) or security through transparency. One key insight is that secrecy prevents people from assessing their own risks, which contributes to a false sense of security. Responsible disclosure means that the researcher discloses full information to the vendor, expecting that mitigation is developed within a reasonable timeframe. An increasing number of organizations has adopted some form of responsible disclosure. A risk based regulatory regime are dependent on such an open discussion of the risks.

In summary, if we want systems that are safe, secure and reliable, both safety, security and reliability must be built together. There has been documented several vulnerabilities and responsible disclosure of vulnerabilities to the vendors, seems to be a beneficial policy. Some sort of communities of practices, and a CERT of autonomous systems should be established. There is missing international regulation or compacts based on private public partnerships to ensure privacy, safety, security and resilience. Vendors must ensure this quality by design, and must be prepared to accept legal liability of the technology they produce. Regulations should require routine, transparent reporting of technological problems to provide data for a transparent market-based cyber-insurance industry, and a risk based regulatory regime.

3.3 *Key issues when discussing regulation*

3.3.1 *Selected issues from all forms of transport*

Risks of autonomous transport are not well known at present. To increase knowledge and learning, experiences, taxonomies, regulations and relevant incidents should be gathered and disseminated from all modes—autonomous road systems (vehicles), air transport (i.e. drones), rail (unmanned metro and rail systems) and shipping. Accident investigators and rule-makers (such as "The Accident Investigation Board in Norway") should develop methods for investigation of accidents of autonomy and report their findings.

Shipping: Completely unmanned ships seem to give large benefits and enables new transport systems, some of these issues are documented in Rødseth (2017). There is a need for onshore control centres to manage autonomous shipping operations. Norway has focused on autonomy in sea transport. A network, Norwegian Forum for Autonomous Ships (NFAS) at nfas.autonomous-ship.org, has been established. A more general research program called Centre for Autonomous Marine Operations and Systems (AMOS) has been initiated at the Norwegian University of Technology and Science, ref www.ntnu.edu/amos. The Trondheimsfjord has been selected as a national testing area in collaboration with The Norwegian Maritime Authority and The Norwegian Coastal Administration. At the end of 2017 three testing areas has been established in Norway (Trondheimsfjord, Storfjord and Horten). Test areas has also been established in Finland and China. Risk levels of autonomous ships are influenced by existing incidents and new incidents (i.e. caused by new automation, and former incidents mitigated by crew now being removed). Work is ongoing to explore safety of autonomous sea transport, and to explore a taxonomy of LOA for shipping, Rødseth et al. (2017). In Trondheimsfjord an autonomous passenger ferry is going to be tested in 2018–2019. The authorities need to set rules and requirements based on acceptable risk levels. There

is an increased need for Human Factors knowledge to improve the quality of interfaces (i.e. "human in the loop" control when needed) between humans and the autonomous systems.

Aviation: To govern the use of Remotely Piloted Aircraft Systems (i.e., drones) in Norway, regulation has been established, Civil Aviation Authority—CAA (2016). The operator must be certified through an exam, CAA (2017). Experiences of remotely piloted aircraft Systems, Waraich et al. (2013), documents that mishaps may happen (i.e. 50 mishaps occur every 100,000 flight hours' vs human-operated aircraft where there is one mishap per 100,000 flight hours). The high mishap rate is related to poor attention to human factors science and design in ground control centres, Waraich et al. (2013). Several pilot projects with drones are planned, transporting goods/persons.

Rail/Metro systems: By automated metros (rail systems) we mean systems where there is no driver in the front cabin, nor accompanying staff, also called Unattended Train Operation (UTO). UTO have been in operations from 1980. In UITP (2013), there are listed 674 km of automated metros consisting of 48 lines in 32 cities. UTO´s are found in Barcelona, Copenhagen, Dubai, Kobe, Lille, Nuremberg, Paris, Singapore, Taipei, Tokyo, Toulouse and Vancouver. Wang et al. (2016), list the arguments for UTO as increased reliability, lower operation costs, increased capacity, energy efficiency and an impressive safety record. There is substantial infrastructure cost to ensure safe on and offloading of passengers and that the track is safe and isolated from other traffic. Four distinct Levels of automation are defined: GoA1: *Non-automated train operation, with a driver in the cabin*. GoA2: *Automatic train operation system controls train movements, but a driver in the cabin observes and stops the train in case of a hazardous situation.* GoA3: *No driver in the cabin but an operation staff on board.* GoA4: *Unattended train operation, with no operation staff on board*. We have at present not found normalized accident data for UTO (incidents based on person km), but no accidents have been reported. It seems that the UTO has exceptionally high safety. However more systematic analysis and normalization of all international UTO transport incidents are needed.

Road Transportation: Google's self-driving cars, where the vehicle systems control all aspects of the driving, have been on public roads in the US since 2009. The safety record has been impressive. However Google engineers are supervising and re-taking vehicle control if necessary. The death accident in 2016 (Joshua Brown) by Tesla in Autonomous driving condition was caused by a tractor-trailer that made a left turn in front of the Tesla, and the car failed to apply the brakes. The Tesla did not "see" the trailer—it was all white and had poor contrast with the surrounding bright white sky. In addition, there was a high gap between the road and the trailer. The National Transportation Safety Board (NTSB, 2017) found that the system's "operational design" was a contributing factor to the crash because it allows drivers to avoid steering/watching the road for periods of time that were "inconsistent" with warnings. Tesla could have taken further steps to prevent the system's misuse. In addition, NTSB faulted the driver for not paying attention and "over-reliance on vehicle automation". It also seems there is a need for better training of drivers related to autonomous systems—a part of driver education and driver license requirements.

There are scarce safety data so far, but data from the period 2009 to end of 2015 has been collected from Googles cars, in Teoh et al. (2017). There were three police reportable accidents (denoted as level C1) in California while driving 2,208,199 km, giving an accident rate of 1,36 police reportable incident pr. million km. This is 1/3 of reportable accidents of human-driven passenger vehicles in the same area. Car accidents involving autonomous cars are different from human driven. Google cars get more rear-ended by other vehicles while stopped or barely moving. There is an element of risk negligence in that the human driver does not fully anticipate the action of the self-driving car. There are also challenges of sustained human attention during lengthy period of autonomous driving, making it difficult for the human operator to intervene i.e. "Human in the loop" challenges. Huffington (2017) documented that Waymo's human drivers had to take control from the automated system (i.e. "disengagement") once for every 5,000 miles in 2016. "Backup" human drivers in Uber's self-driving cars had to take over about once every mile as of March 8, ref Recode (2017). It is a challenge to get situational awareness after having been out of the driving control loop for 5,000 miles. The takeover time of the human driver varies from 2 to 26 seconds, ref Eriksson et al. (2017), challenging the design of autonomous systems to enable human intervention.

Analysing all car accidents, it is suggested that 80–90% of accidents are due to "human errors", thus autonomous cars could reduce the level of accidents substantially. However, autonomy could introduce new types of accidents, due to automation itself or due to human drivers not predicting action from the automation. In Blanco, et al. (2016) it is suggested that accident rates are reduced to ¾ of present, while Teoh et al. (2017) documents accident levels of autonomous systems as 1/3 of human driver systems. The National Highway Traffic Safety Administration (NHTSA, 2017) reported a reduction in vehicle crash-rate

by almost 40% with Autosteer activated in Tesla Model S and Model X, compared to before. In Cummings et al. (2014) it is suggested that the level of accidents could be reduced by 50%. More experiences must be gathered, but significant reduction of accidents is expected.

3.3.2 *Need for systematic open data reporting*

At present there are missing data of incidents (accidents and successful recoveries) related to autonomous systems. Open reporting must be established covering systematic safety records and security stories, being available to researchers and industry actors, such as insurance. The scope must cover actions from the autonomous system but also document perceptions and understanding from the involved human actors. The differences between espoused values (rule based actions/work as programmed in autonomous systems) and actual values (actions/work as being done by humans in interaction with autonomy) can create the basis for errors and accidents. It should be a key area of research to explore accidents because of poor design vs. blaming the human actors. Use of video recording could help, based on regulation protecting personal data; (EU 2016:679). There must be a combination of data gathering in combination with in-depth accident investigation. Accident investigation boards should explore accidents of autonomy, to support rapid learning and changes in addition to improve their methods to analyse autonomy incidents.

3.3.3 *System perspective and human factors*

Safety of autonomous systems are dependent on new designed technology, human factors and organisational issues as discussed by Cummings et al. (2014). The perception should be that most accidents in autonomous systems are a consequence of poor design and poor testing, and that "human errors" are a consequence and not a cause as described by Dekker (2002). Moving trivial functions (that can be programmed) to an autonomous system, means that tough decisions and deviations must be handled by humans. Thus, the science of Human Factors, knowing strengths and weaknesses in cognition and ergonomics, must get a significant position when automation is designed and implemented.

3.3.4 *Responsibilities and certification*

The autonomous system decides based on design approved by the manufacturer. Thus, product responsibilities of accidents and incidents must be placed at the manufacturer (OEM). This is in line with the view of the car OEMs Volvo, Google and Mercedes-Benz (Iozzio, 2016). This is also in line with the supervisory responsibility demanded in the Oil and Gas industry (i.e. where the operator is responsible for the chain of suppliers employed). This supervisor responsibility must be placed on the car OEMs, including the continued updating and adaptation of software in use. Certification is needed, such as the ISA/IEC-62443 scheme of industrial control systems used since 2010. However, certification is still being developed, a survey documenting key issues are found in Martin, et al. (2015).

3.3.5 *Security and risk-based regulation*

Security (for safety) must be included in the development of autonomous systems, and systematic testing (including penetration testing) must be done as a part of certification prior to product release. The precautionary principle must be established as a condition for autonomous transport systems, COMEST (2005).

4 CONCLUSIONS

Related to the research question RQ1 (major risks): The sensors and systems used in autonomous systems, does not have a perfect view of the surroundings, and may also act uncoordinated with their surroundings, thus new type of accidents may happen. There is a need to speed up learning from these incidents and to be aware of communication and information challenges in operations.

Human control and assistance through control centres and via human machine interactions must be designed based on the science of Human Factors in order to avoid higher levels of accidents as documented by Waraich et al. (2013).

We continue to see vulnerabilities and exploitation of software in the public and private sectors. Different perspectives are used in security and safety, due to different adversity models. The security community are addressing threats (directed, deliberate, hostile acts) and the safety community are addressing hazards (undirected events). AEC are so pervasive across all sectors that a silo approach can no longer be acceptable. To ensure that all actors in the value-chain understands this, a silo-based "need to know" principle must be replaced by transparent and open reporting. This can also support a market based cyber-insurance industry.

Related to the research question RQ2 (regulation): There is a need for regulatory action from government to set minimum standards, establish responsibility, and follow up of incidents/accidents. Prescriptive and detailed rulemaking on a national level is wanting, but should be replaced by functional approach demanding the same level of risk in automated systems as in existing systems.

Vendors must have responsibility to ensure safety, security and resilience by design, and must be prepared to accept legal liability for the quality of the technology they produce. Ideally, a formal process of product acceptance and certification (i.e. safety case) should be established before a product can be sold. The manufacturers should establish a proactive focus on (best practice) safety/security standards. There is a need to ensure that there is some sort of a structured learning process (among all relevant actors) when incidents happen.

Related to the research question RQ3 (way forward): Innovative approaches, such as the perspective of Autonomous Ecosystems (AEC) are needed to handle the challenges of autonomous transport systems. The science of Human Factors need to be prioritized to ensure that human intervention can be designed in the system and can be performed in actual operations based on actual human limitations and human strengths to improvise and handle unanticipated events.

Safety has been dependent on publicised accidents and a systematic learning loop between users, the regulator and industry. One component in the learning loop of complex software systems has been reporting and analysis of incidents through computer incident response teams (CERTS). There is a need to establish CERTS of AEC to help coordinate actions.

Rules and mechanisms for updating software in autonomous systems will become more urgent as failures can lead to accidents, thus handling of updates must be addressed in a systematic manner.

Communication between autonomous transport systems and drivers and bystanders must be improved. Autonomous systems are rule based while humans are not, thus there may be misunderstandings and common failures, creating need for interventions through transport centres controlling the flow of transport.

These AEC will be exposed to new strains—thus there must be a focus on how to handle surprises by resilience, to ensure that new demands/ stress/ failures are not impacting transportation in a catastrophic way.

REFERENCES

Axelrod, C.W. (2014, May). Reducing software assurance risks for security-critical and safety-critical systems. In *Systems, Applications and Technology Conference (LISAT), 2014 IEEE Long Island* (pp. 1–6). IEEE.

AHUS (2009) www.rb.no/lokale-nyheter/pakjort-av-robot-pa-jobben/s/1-95-4309194

Blanco, M., Atwood, J., Russell, S., Trimble, T., McClafferty, J., & Perez, M. (2016). Automated vehicle crash rate comparison using naturalistic data. Virginia Tech TI.

Cerrudo, C. (2015). An emerging US (and world) threat: Cities wide open to cyber attacks. *Securing Smart Cities—White paper - IOActive*. www.ioactive.com/pdfs/IOActive_HackingCitiesPaper_cyber-security_CesarCerrudo.pdf

CAA-Civil Aviation Authority (2016) Regulations for Remotely Piloted Aircraft Systems (Forskrift om luftfartøy som ikke har fører om bord mv) retrieved from lovdata.no/forskrift/2015-11-30-1404

CAA-Civil Aviation Authority (2017) Training requirements—retrieved from luftfartstilsynet.no/selvbetjening/allmennfly/Droner/

COMEST (2005) The Precautionary Principle from UNESCOs World Commission on the Ethics of Scientific Knowledge and Technology.

Cummings, M.L., & Ryan, J. (2014). Who is in charge? The promises, pitfalls of driverless cars. TR News, 292, 25–30.

Dekker, S.W.A. (2002). Reconstructing the human contribution to accidents: The new view of human error and performance. *Journal of Safety Research*, 33(3), 371–385.

DHS (2015) Department of Homeland Security, Office of Cyber and Infrastructure Analysis: The Future of Smart Cities: Cyber-Physical Infrastructure Risk

EU (2016:679) On the protection of natural persons with regard to the processing of personal data and on the free movement of such data; Regulation of the European Parliament and of the Council of 27 April 2016

Eriksson, A., & Stanton, N.A. (2017). Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human factors*, *59*(4), 689–705.

Firesmith, D.G. (2003). "Common concepts underlying safety, security, and survivability engineering", *Technical note CMU/SEI-2003-TN-033*, Carnegie Mellon University.

Frei, S., Schatzmann, D., Plattner, B., & Trammell, B. (2010). Modeling the security ecosystem-the dynamics of (in) security. *In Economics of Information Security and Privacy* (pp. 79–106). Springer US.

GCIG (2016) Global Commission on Internet Governance, "One Internet" www.ourinternet.org

Hollnagel, E. Woods D. and Leveson N. (2006). "*Resilience Engineering*", Ashgate.

Iozzio, C. (2016). Who's Responsible When a Car Controls the Wheel?. *Scientific American*, *314*(5), 12–13.

Koscher, K., Czeskis, A., Roesner, F., Patel, S., Kohno, T., Checkoway, S.,... & Savage, S. (2010, May). Experimental security analysis of a modern automobile. In *2010 IEEE Symposium on Security and Privacy* (pp. 447–462). IEEE.

Lima, A., Rocha, F., Völp, M., & Esteves-Veríssimo, P. (2016). Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems. In Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy (pp. 59–70). ACM.

Lund, J., & Aarø, L.E. (2004). Accident prevention. Presentation of a model placing emphasis on human, structural and cultural factors. *Safety Science*, *42*(4), 271–324

Martin, J., Kim, N., Mittal, D., & Chisholm, M. (2015). Certification for autonomous vehicles. *Automative Cyber-physical Systems course paper, University of North Carolina, Chapel Hill, NC, USA.*

Manikas, K., & Hansen, K.M. (2013). Software ecosystems–a systematic literature review. *Journal of Systems and Software*, *86*(5), 1294–1306.

MTC (2016) Ministry of Transport and Communications "Testing of autonomous road transport systems" from www.regjeringen.no/no/dokumenter/horing---forslag-til-ny-lov-om-utproving-av-selvkjorende-kjoretoy-paveg/id2523663/

NTSB(2017) www.ntsb.gov/investigations/AccidentReports/Pages/HWY16FH018-preliminary.aspx

NHTSA (2017) The National Highway Traffic Safety Administration, Office of Defects Investigation resume PE 16-007. Automatic vehicle control systems, Tesla Model S accident in Florida May 7 2016.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), 230–253.

Renn, O. (2005). "Risk Governance—Towards an Integrative Approach" *White paper no.1 – international risk governance council.*

Recode (2017) www.recode.net/2017/3/16/14938116/uber-travis-kalanick-self-driving-internal-metrics-slow-progress

Rødseth, Ø.J. From concept to reality: Unmanned merchant ship research in Norway. I: Proceedings of Underwater Technology (UT), 2017 IEEE. IEEE 2017 ISBN 978-1-5090-5266-0. OCEAN

Rødseth, Ø.J. & Nordahl H. Ed. (2017). Definition for autonomous merchant ships. Version 1.0, October 10. 2017. Norwegian Forum for Autonomous Ships. nfas.autonomous-ship.org/resources-en.html Accessed December 2017.

SAE (2016). SAE International standard "J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems." Revised: 2016-09-30

Sheridan, T.B., & Verplank, W.L. (1978). *Human and computer control of undersea teleoperators*. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Teoh, E.R., & Kidd, D.G. (2017). Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research*, *63*, 57–60

UITP (2013) Observatory of Automated Metros World atlas report. International Association of Public Transport (UITP), Brussels

Vagia, M., Transeth, A.A., & Fjerdingen, S.A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?. *Applied ergonomics*, *53*, 190–202.

Waraich, Q.R., Mazzuchi, T.A., Sarkani, S., & Rico, D.F. (2013). Minimizing human factors mishaps in unmanned aircraft systems. *ergonomics in design*, *21*(1), 25–32

Wang, Y., Zhang, M., Ma, J., & Zhou, X. (2016). Survey on driverless train operation for urban rail transit systems. *Urban Rail Transit*, 1–8.