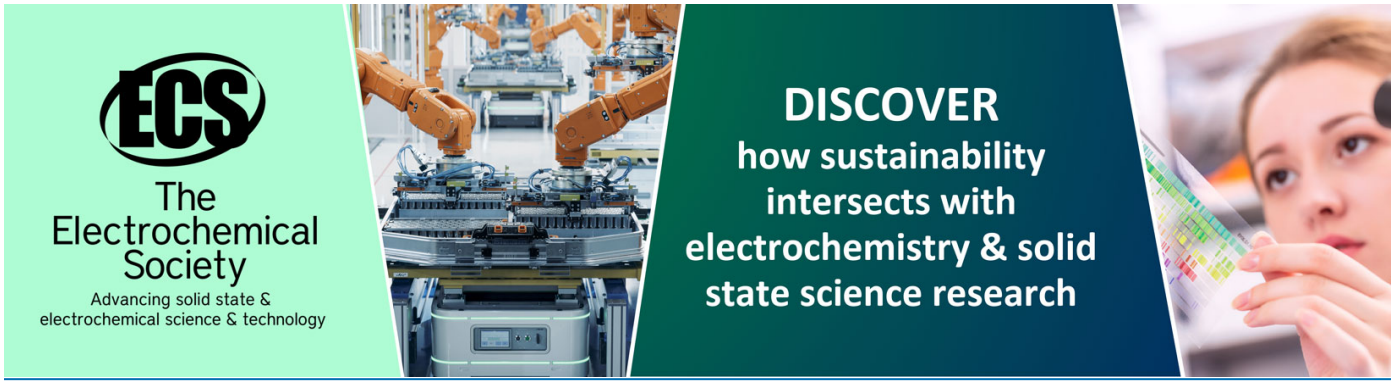
Experiences of main risks and mitigation in autonomous transport systems



IOP Conf. Series: Journal of Physics: Conf. Series **1357** (2019) 012012 doi:10.1088/1742-6596/1357/1/012012

Experiences of main risks and mitigation in autonomous transport systems

S O Johnsen[1], Å Hoem[1], G Jenssen[1] and T Moen[1]

[1]SINTEF, Trondheim, Norway

Stig.O.Johnsen@Sintef.no

Abstract. This paper discusses experiences, risks and mitigation in autonomous transport systems, to improve learning and risk-based governance. This is based on literature reviews, experiences of autonomous transport systems and involvement in the regulatory process. The maturity varies between rail, road, aviation and shipping. Issues that may be transferred and adapted between modes are selected. The three research questions are: to describe major risks introduced by autonomous transport systems; how to mitigate the main risks through design and regulation; and suggest a way forward. The risk is dependent on operational domain, surrounding infrastructure, meaningful human control (including control centrals with human based interfaces), sensor ability, reliable propulsion and technology that are secure. Meaningful human control and system certification is a challenge. There is a need to gather data from incidents. In Norway autonomous shipping and road transportation are prioritized. Autonomy in rail and aviation, has not been prioritized in the same manner.

1. Introduction

This paper discusses experiences of autonomous transport systems, to establish a framework for risk-based design and governance. Autonomous systems have different maturity in aviation, rail, road and sea. As an example, aviation has implemented a high level of automation in combination with advanced infrastructure and are ultra-safe, the other modes have varying experience with automation and varying safety levels. Our approach has been to explore the different modes in order to speed up learning and knowledge sharing. Issues that may be transferred and adapted between modes are selected. Analysis of accidents has shown that when cumulative experience is doubled, accidents are reduced by a fixed percentage, [1] and [2], thus sharing of experiences should mitigate accidents.

Collaboration between automated systems and humans may involve the need for human actions or control. The design of the collaboration should be based on meaningful human control. By meaningful human control we mean that control actions are based on involving humans in the loop in such a manner that the human actor can handle the situation within human limitations based on appropriate human based design. I.e. the competent human actor is able and motivated to influence the behaviour of the system in the right way at the right time, as discussed in [3].

Risk and risk governance are based on [4], starting with problem framing; risk appraisal (hazards and vulnerabilities); risk judgment; risk communication and risk management. The implementation of autonomy can reduce some 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 [5], risk reduction must be based on a broad set of actions such as regulation, technical design, training and awareness. Based on experiences of autonomous transport systems, and involvement in the regulatory process in Norway, we have explored emerging risks and threats. In the following we have defined autonomous systems and concepts such as levels of automation (LOA) to specify degree of automation.

1.1. Definitions and terminology

By automated we mean a system that will do exactly what it is programmed to do. By autonomy we mean a step further, a system that can make a choice free from outside influences and change its initial way of programmed actions, having some notion of "free will", where actions are not determined completely by previously existing causes i.e. non-deterministic as in [6]; based on [7]. In [8] automation and autonomy are 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 1: Remote control (Surveyed and/or externally controlled); 2: Local control i.e. autonomous (based on own sensors and systems; includes cooperation based on traffic information) or 3: A combination of remote and local control based on the situation.

When trying to scope risks of autonomous systems we must include regulation, risk governance, organizational framework, interfaces to humans and the autonomous system (a combination of software components and cyber physical systems). The systems in use is often a collection of systems being developed by different stakeholders thus the concept of Autonomous Ecosystem (AEC) is used, inspired by Software Ecosystems (SEC). SEC consists of components developed by actors both internally and externally of the company [9]. Arguments for AEC is that development is taking place outside organisational silos due to the need for speed of development, need for supporting applications, reduction of development costs and competition. This creates the need to address governance challenges of an ecosystem framework. An example of an autonomous ecosystem is intelligent transport systems (ITS) consisting of autonomous vehicles, integrated with control facilities. These autonomous ecosystems handle information, but also actual safety critical processes such as transport (via automobiles, boats, drones and trams). The systems must be able to handle unanticipated events, breakdowns and be able to go to a safe and secure (end-)state. Regulation of autonomous systems must be established. Issues related to product liability becomes more important as the automation system are in control [10]. A risk-based approach is beneficial in order to reduce the main risks. To develop risk understanding, regulation and best practices, there is a need to gather a full set of data of incidents and accidents. In order to do this, we need to clarify responsibilities of the levels of automation (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. In [11], ten steps of automation were introduced, 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 LOA in driving, [12]. Going from no autonomy (level 0), through driver assistance, partial automation, conditional automation, high automation (level 4), to full automation (level 5). Our focus has been the step from high automation (level 4), to full automation (level 5). The design must ensure that the system maintains an accepted level of performance despite disturbances, including threats of an unexpected and malicious nature.

A key definition in [12] is scope, i.e. Operational Design Domain (ODD) for the autonomous vehicle describing the area of operation (i.e. selected specific road or area), type of road/infrastructure needed, environmental constraints (weather), other allowed traffic in the same area, speed limits, allowed time of operations. This is a key definition of interest in all modes: sea/ air/ road/ rail.

Handling of unanticipated incidents and continue to operate safe must be key ability of autonomous transport systems. The concept of resilience engineering is an important strategy to handle unanticipated incidents. In [13] they 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*".

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 key design and regulatory issues should be prioritized to handle these risks?
- RQ3: What is the way forward, i.e. main approaches and issues needed to mitigate major risks of autonomous transport systems?

2. Scope, challenges and methods

2.1. Challenges and problems

Needs for safety, security and resilience have often been identified late when vulnerabilities have been exploited and unwanted incidents have been published. When discussing vulnerabilities in autonomous eco-system, one challenge is that there is not one single supplier, but a set of suppliers involved. It can be difficult to identify responsibilities if framework conditions (regulation/ responsibilities) are missing. 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.

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 systems in Norway (focused on road transport). We have performed the 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 also performed a review of drone research from database of the Norwegian research council in the period 2008 to 2021 and a literature review based on the search string "Safety and Security of Unmanned Aircraft Systems/ or Drones /or UAS", limited to the period 2013 to 2019.

We have explored experiences of autonomous transport systems from St. Olav Hospital in Norway, where autonomous systems have been used from 2006. St. Olav has 10,400 employees and covers an area of 200,000 Square Meters. We have explored industrial autonomous road transportation. We are involved in projects with self -driving busses in three Norwegian cities. We have been involved in a hearing of regulation of testing of autonomous vehicles in Norway from the Ministry of Transport and Communications, [14], distributed 2016, approved as law in 2017. Our comments were based on literature review and gathered experiences.

We base our analysis on the Human Factors Analysis and Classification Systems for accidents HFAC [15]. HFAC consists of the following causes (with more descriptions in parenthesis):

- 5: External factors (Poor regulation, Poor Design or Wrong LOA, Poor regulatory oversight)
- 4: Organisational issues (Poor planning, Poor safety management, Poor procedures; Lacking meaningful human control; Poor training)
- 3: Unsafe supervision (Poor risk governance, Unsafe operations, Poor infrastructure/ODD)
- 2: Operational preconditions (Human Factors of operators, Poor practice of procedures, Poor quality of infrastructure)
- 1: Unsafe acts (Knowledge based mistakes; Routine based violations; Sensor or program failures of autonomous system)

The taxonomy used to register the seriousness of incidents has been based on [16] 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. (Applicable to all modes):

- 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 strike potentially in conflict with oncoming traffic and with higher risk potential if no curb.

- C4: Tire strike only with little 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, we have documented experiences from eco-systems/software systems in autonomy; and a summary of key issues of autonomous systems used in road; air; sea and railroad.

3.1. Findings from literature review of autonomy and safety/security/resilience in software systems

In [17] the focus is on software assurance of safety-critical and security-critical systems. The perception is that current methods has not achieved the wished-for level of protection, and that there are missing security principles and standards. A requirement is that 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 because of security-related errors. All this related to HFAC#5: External factors.

International governance of security of the infrastructure is addressed through standard bodies (i.e. ISO, IEC) and international bodies such as OECD, EU, NATO and UN. In [18] 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. Governments can play a role by incorporating minimum security standards in their procurement. 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 an evidence-based decision making.

In [19] 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. 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).

In [20], 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). 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 [21] there is an empirical evaluation of “smart cities” looking at a broad set of technologies of traffic control. Known vulnerabilities are in traffic control systems, mobile applications, smart grids/smart meters and video cameras. The issues are lack of cyber security testing, approval, lack of encryption, lack of City Computer Emergency Response Teams (CERT), and lack of cyber-attack emergency plans. We may establish potential for serious incidents, if these issues are not addressed.

In [22] 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. A risk-based regulation is dependent on an open discussion of the risks.

In summary, key issues impacts External factors (Hfacs#5). If we want systems that are safe, secure and reliable, both safety, security and reliability must be built together. Vulnerabilities exist. Responsible disclosure of vulnerabilities to the vendors are needed. Communities of practices, and a CERT of autonomous systems should be established. There is missing international regulation 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. Industry and regulators need to establish goal-based standards, certification and to focus on dissemination of best practices and systematic education.

3.2. Findings from autonomy in road transportation

3.2.1. 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. From 2010 to 2018 they have had 21 AGVs in operations. Each week the 21 AGVs transport medicine, food, clothes and garbage, in total 70-80 tonnes. (An AGV transport 500 kg and transport 3.6 tonnes/week). The speed is approximately 2 km/hour (max speed is 5 km/h). The AGVs can communicate, open doors, and reserve elevators. 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, manned by an operator that can intervene or go to the place where there is a conflict. The AGVs has created the need for separated road areas and the staff in control centre.

A key issue related to the communication between automated transport systems and humans are the messages from the AGVs, supporting that the automated system needs to inform about their perceptions and what they are going to do next. The AGVs deliver pre-programmed messages such as “Please move – you are in my way”, or “Elevator is reserved – please move out of elevator”.

Operational statistics is poor, it does not easily document the travel length of the AVG's or systematize or document the level of incidents or accidents. We have had to perform interviews and analysis to gather data. We estimated that one AVG is utilized 4 hours each day and travels 8043 meters per day, for 21 AVG's a total of 42 226 km in a year. There has been a total of 100-130 minor incidents per year (5-6 per AGV) categorised as C4, i.e. estimated to 2,4 minor C4 incidents per 1000 km travel. Minor repairs are done on the AGVs, changing around 50 components per year. There are around 15 emergency stops each year, categorized as C3 (0,36 per 1000 km), where components must be changed. We do not have data of any incidents of category C2 or C1. Reported incidents are minor crashes due to faulty navigation, due to objects placed in the route. Interviews suggested:

- *Adapt surrounding infrastructure (clear the track of objects); Improve quality of sensors:* The AGVs have problems with some object close to the walls. 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); design should mitigate the limits of AVGs, i.e. poor ability of sensors to identify objects and that the infrastructure must be adapted, i.e. HFACS#3 Need for adapted infrastructure.
- *Adapt surrounding infrastructure (make objects visible to the automated vehicle); Improve quality of sensors:* The AGV collided with the forklifts, since the LiDAR sensor (light detection and ranging) had a limited vertical field of view and was seeing a free zone under the forklift. This was mitigated by placing a black rubber skirt under the forklift. The fatal accident of Joshua Brown, [23] and [24], between a Tesla and a trailer crossing the road – a white trailer giving poor contrast and with substantial height above the ground., i.e. some similarity with the forklift problems at St. Olav. A rubber skirt under the trailer may have increased visibility of the trailer i.e. need for mitigating poor visual ability from the automated system, i.e. HFACS#3.
- *Adapt surrounding infrastructure (need for central control room with operator);* The AGVs can open doors and reserve elevators. Sometimes there has been conflicts between the AGVs and the users, needing human intervention. There is need for a risk analysis and an alarm strategy to ensure that safety critical issues are avoided, i.e. HFACS#3 Need for adapted infrastructure.
- *Maintain interfaces to cyber physical systems:* Software updates of AVGs, elevators and doors has led to problems, i.e. doors and elevators has not responded to communication. Sometimes the AVG has locked the elevator from other use (critical patient transport). This has been mitigated by

duplicated elevators, i.e. planned resilience in critical infrastructure. There is a need to look at the AVGs as a part of an ecosystem to ensure that all updates are tested in the ecosystem.

In summary, the AGV system has had an impressive safety record at St. Olav's Hospital. Risks and situations of hazards have been: 1) Collisions due to poor status of self and surrounding infrastructure and 2)"Lock" situations where the AVG reserves an object that may be critical (i.e. elevator). Incidents from the use of the AVG robots have been related to external factors (ODD-Design of paths and infrastructure; design of meaningful human interactions) and organisational issues (Poor and missing procedures). The learning points relevant to all modes are:

- Conduct pilots with a few autonomous systems to learn and prepare of the infrastructure.
- Design and adapt infrastructure and surroundings to the automated vehicles and build resilience through duplication of key components (example: elevators that may be reserved).
- Design paths to reduce risks avoiding congested areas, adapted to the use of automated systems.
- Design to reduce risks by establishing low energy operations through low speed of the robots.
- Design for remote human interaction - A manned control centre was established to intervene during unexpected issues. (Still manned by one to two persons).
- Focus on building trust through communication between automated systems and humans.
- Gather operational data to speed up learning and adaptation – document incidents and accidents especially at interfaces to humans. Use established taxonomies as [16].

3.2.2. Road Transportation in general: Google's self-driving cars have been on public roads in the US since 2009. The safety record has been impressive (Google engineers are supervising and re-taking vehicle control if necessary). The fatal 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 – and there was a high gap between the road and the trailer. NTSB, The National Transportation Safety Board [23] 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, i.e. HFAC#5 – poor design. 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”. There is a need for better training of operating autonomous systems, i.e. a part of driver license requirements. There are scarce safety data so far, but data from the period 2009 to 2015 has been collected [25]. 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 per million km. This is 1/3 of reportable accidents of human-driven passenger vehicles in the same area, i.e. 66% reduction. 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 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. 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 [26]. It is a challenge to get situational awareness after having been out of the control loop for 5,000 miles. The human takeover time varies from 2 to 26 sec [27] challenging the design of autonomous systems to enable rapid (in-time) human intervention.

It is suggested by some that 80-90% of accidents are due to “human errors”, thus autonomous cars could reduce the level of accidents substantially. This is debatable – in Norway the risk of fatal road accident is 3 per billion vehicle km, while in a similar human environment, in the US it is 7,3 [28]. The risk in Norway is 40% of the risk in the US, indicating that the causes are more complex. Autonomy could introduce new types of accidents, due to automation or due to human drivers not predicting action from the automation. In [16] it is suggested that accident rates are reduced to ¾ of present, reduction of

25%. In [29] it is suggested that the level of accidents could be reduced by 50%. More experiences must be gathered, but experiences so fare indicates a reduction of 25-66%.

Several pilot projects have been conducted in Norway both with personnel transport (autonomous buses in two Cities Oslo and Stavanger) and transport of goods at a Caulk mine [30]. The focus on developing regulation, infrastructure support and innovation has made Norway among the top three in autonomous vehicle readiness as described by KPMG [31]. The four main criteria used by KPMG are: Policy and Legislation; Technology and Innovation; Infrastructure; and Consumer Acceptance.

Comments related to suggested regulation [32], has been to learn across different modes (air, road, rail, sea); ensure systematic reporting of all incidents; focus on both safety and security; need for supporting infrastructure and need for remote control. In summary – main risks of autonomous systems are related to poor image of self and poor support from surrounding infrastructure; risk reduction activities should be based on learning from pilot projects; gathering of incident data; risk based mitigation of unwanted incidents; some sort of certification based on "best practices" and clarification of remote operation support. Accidents due to automation will happen, changing from a human driver to an automated system may decrease the level of accidents with 25-66%.

3.3. Findings from autonomy in aviation

Manned aviation is piloted (with a high degree of automation), unmanned aviation systems (UAS i.e. drones) are usually remotely controlled (i.e. Remotely Piloted Aircraft Systems), but they can also be autonomous. Manned aviation has ultra-high safety, with an accident rate of 1.08 accidents per 1 million flights and have experienced two years with no hull-losses among IATA members (in 2015 and 2018). Key issues for this high level of safety has been a risk based focus on accidents (and the dissemination of learning and improving from accidents based on "no-blame" culture), international strict regulation, large industry actors with a high quality of airplanes and infrastructure (control systems, airports), a system perspective on accidents with a focus on the science of Human Factors (from 1945 and onward), and risk based training and selection (i.e. Crew Resource Management – CRM- training from early 1990s). These issues should be in focus during design of autonomy in the other transport modes, to improve safety.

Drones have often been used in tasks being dangerous, dirty, dull or where speed of delivery is critical. Use of drones is expanding, a compound annual growth rate of 100%. High-speed delivery, inspections of equipment, aerial surveys and photography are increasing. As an example, drone delivery of blood has been in operation since 2016 in Rwanda. To govern the use of drones in Norway, regulation has been established, Civil Aviation Authority [33]. The operator must be certified through an exam [34]. In [35] the Mean Time Between Failures (MTBF) was estimated to 1000 hours for drones. Approximately 100 times higher than MTBF in manned flights. Based on 1000 - failures, the division of failures were in: Power plant (411); Ground Control system (273); Navigation system (146); Electronic system (67); Mainframe (54) and Payload (53). The risk of these failures is dependent on the consequences of the failures such as weight (lightweight drone observation, or heavy industrial transport) and the area of use (in an urban area/ city or in remote and sparsely populated areas). Failures can have significant impact in an industrial (heavy) drone flying in a city and crashing by high speed (i.e. high energy impact). Suggestions for risk assessments are documented in [36]. Security issues related to the use of drones [37] are:

- Fly away attacks (stealing the drone) or Take-down a flying drone - through GPS spoofing
- Injecting falsified sensor data or attacks to destabilize the drone making it to crash.
- Malicious hardware or software, such as backdoors to take control
- Lock-out owner of the drone from connecting; or manipulating data used to navigate
- Steal user data (pictures, streaming video) or related privacy issues such as unsafe data-storage

These issues can be mitigated by securing drone access by strong passwords in user authentication; limiting devices allowed to connect (i.e. enforce authentication); disabling services with poor security; encrypting communication, certify software dependent on risk levels and continuously upgrade software in use. Experiences of UAS from the US government [38], documents that mishaps may happen (i.e.

50-100 mishaps occur every 100,000 flight hours' vs human-operated aircraft where there is one mishap per 100,000 flight hours). The mishap rate of UAS is 50-100 times higher than manned operations. Main causes are related to poor attention to human factors science, such as poor design of ground control centres, [38]. Checklist to improve the design of control centres are suggested by [39].

Use of industrial drones out of line of sight is developing slowly, with many barriers. The use of drones is increasing but regulation, best practices and testing in key areas has been lagging in Norway.

In general there is a need to: establish a registry of owners/ drones; establish systems for drone detection in critical areas (and forced landing if needed); ensure that liability can be directed to owners and responsible developers; ensure that drones are insured; ensure that specific drones have the ability to go to a safe/secure state (i.e. safe landing); establish systematic documentation and reporting of all drone incidents; specify certification scheme and quality assurance of drones; support pilot projects to test the use of drones (i.e. beyond line of sight and in personnel transport); support improved development of practices and regulation to ensure more rapid deployment of drones.

3.4. Issues from autonomy in rail and metro systems in operation from 1980

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 [40] 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. The arguments for UTO as increased reliability, lower operation costs, increased capacity, energy efficiency and an impressive safety record [41]. 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 LOA's 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 operations.

We have at present not found normalized accident data for UTO (incidents based on person km), but no significant accidents have been reported with loss of life or significant harm. It seems that the UTO has exceptionally high safety. However, a systematic analysis and normalization of all international UTO transport incidents are needed. There has not been any significant prioritization of autonomous rail and metro system in Norway.

3.5. Issues from autonomous shipping

Norway has prioritized autonomy in sea transport through national strategies, regulation and pilot projects, documented through white papers [42], [43] and [44]. The expressed goal is to see Norway in the forefront of exploiting autonomous shipping, facilitate for sustainable innovation, protect environment, vessel and the seaman based on safety and security.

Completely unmanned ships seem to give benefits and enable new transport systems, some issues are documented in [45]. The implementation of autonomous shipping creates the need for onshore control centres to manage autonomous shipping operations. A network, Norwegian Forum for Autonomous Ships (NFAS) at nfas.autonomous-ship.org, has been established. A research program, Centre for autonomous marine operations and systems (AMOS), has been initiated at the Norwegian University of Technology and Science, ref www.ntnu.edu/amos. At the end of 2018 three national testing areas has been established in Norway (Trondheimsfjord, Storfjord and Horten). At least six test areas have been established internationally in Finland, China, Germany; the Netherlands and US.

Actual risk levels of autonomous ships are not known; however, some analyses have been performed. As an example,[46], based their assessment on a review of 100 maritime accidents; they suggested that probability of incidents may be reduced but that the consequences may increase if no mitigating actions were established. Examples were decrease of the occurrence of navigational accidents (e.g. collision, and groundings) but an increase of consequences due to missing intervention from humans (during groundings, fire or collisions). However, this may be mitigated through design and practices, and Human

Factors must be used in design and operations [46]. A review of risk assessment of autonomous shipping highlighted the challenge of missing empirical data, the need for including human intervention and other systems in the assessments and need for improved casual model [47]. Work is ongoing to explore safety and a taxonomy of LOA for shipping [48]. Several pilot projects are underway, an example is an autonomous passenger ferry will be tested in 2019-2020 in Trondheimsfjord. The authorities are requiring appropriate design, documentation and third-party verification of these pilot projects.

Based on the safety experiences from aviation as described, we see a need for an open sharing of incidents, best practices and a "no-blame" culture (using a system perspective, being more reluctant to prosecute mariners as a result of accidents), rapid goal based international regulation, fewer and larger industry actors in ship building (of components), more of a system perspective on accidents with a focus on the science of Human Factors (reducing the use of the phrase "Human error" when the system is to blame) and continuous focus on risk based team selection and training (such as CRM). There is a need for Human Factors knowledge to improve the quality of interfaces between humans and the autonomous systems to ensure meaningful human control. Human Factors was seen as a "new" science in the maritime area in 2006 [49] and should be strengthened in shipping. There is a need to reduce the complexity of operations (ODD), (maybe define strict sea-lanes for autonomous shipping); build supporting infrastructure to ensure that the ship is aware of position and surroundings; perform a risk based design to mitigate main challenges (as suggested in [46]); ensure third party verification; establish taxonomies for incident reporting and gather data of incidents and accidents to be shared in an open manner; design control centres based on the science of Human Factors in order to ensure meaningful human control (even during stress). The handling of the unexpected will be a key issue in autonomous shipping; thus, resilience engineering is an important design strategy.

4. Summary and conclusion based on the research questions

The research questions (RQ) were: RQ1: What are the major risks introduced?; RQ2: What key design and regulatory issues should be prioritized to handle these risks? and RQ3: What is the way forward?

Related to the research question RQ1 (major risks): Autonomous systems are dependent on advanced internal logic, signals from sensors (creating understanding of self and surroundings) and environmental signals and controls (from infrastructure and control facilities). The risks introduced are related to probabilities of failures; in control structure, components, failures/weaknesses of infrastructure and failures of control facilities (control centres). The sensors do not have a perfect view of the surroundings and may act uncoordinated with their surroundings/guiding infrastructure, thus new type of accidents may happen. The choice of operational area, ODD, is a key issue; thus, the road/sea/rail/air ODD should be controlled as much as possible – as an example specified and reserved routes for autonomous vehicles should be evaluated. Objects interacting should be adapted to limitations of autonomous systems, (increasing visibility, as an example rubber skirts underneath trailers should be evaluated, in order to mitigate accidents as happened with Joshua Brown.)

Consequences may be dependent on the ability to handle the unanticipated and go to a safe/secure state. Since accidents are unplanned or not foreseen, mitigating actions may not be in place and consequences may increase. As an example, an unchecked/undiscovered fire in an autonomous ship may have greater consequences than fire in a manned ship where humans are present. The need for meaningful human control must be explored and ability to go to a safe/secure state must be designed.

There is a need to speed up learning from incidents and to be aware of communication challenges. Human control and assistance through control centres and via HMI (Human Machine Interface) must be designed based on the science of Human Factors in order to avoid increased accidents [38].

Vulnerabilities of software continues. 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 (mitigation through design and regulation): A key issue is to establish risk-based design and regulation based on experiences and accidents. There must be a combination of data gathering with accident investigation. Scope must cover actions from the autonomous system and document understanding from the involved human actors. Video recording could help, limited by regulation; EU [50].The differences between espoused values (rule-based actions 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. A learning process among all actors must take place when incidents happen.

Mitigation must be based on design of the total system, including the surrounding infrastructure and control centres based on meaningful human control. There is a need for regulatory action 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.

The autonomous system decides based on design and maintenance approved by the manufacturer. Product responsibilities of cars must be placed at the manufacturer, in line with the view from industry such as Volvo, Google and Mercedes-Benz, [51]. In general as the supervisory responsibility demanded from other industries as outsourcing increases, such as the Oil and Gas industry. A formal process of product acceptance and certification (i.e. safety case) should be established before a product can be sold/used. The manufacturers should establish a proactive focus on (best practice) safety/security standards. Certification is needed (i.e. ISA/IEC-62443 scheme of industrial control systems used since 2010) but is still being developed, a survey are found in [52].

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 pre-cautionary principle must be a condition for autonomous systems, [53].

Related to the research question RQ3 (way forward): Innovative approaches, such as using the scope of autonomous ecosystems (AEC), are needed to handle the challenges. The science of Human Factors needs to be prioritized to ensure that meaningful human intervention can be designed and performed in actual operations. Design must be based on actual human limitations and human strengths to improvise and handle unanticipated events. Autonomous systems are rule based while humans are not, thus there may be misunderstandings and common failures, also creating need for interventions through communication, or transport centres controlling the flow of transport. Rules and mechanisms for updating software in autonomous systems will become more urgent as failures can lead to common accidents, thus handling of updates must be addressed in a systematic manner. Transport will be exposed to new strains, there must be a focus on how to handle surprises by resilience, to ensure that new demands/ stress/ failures are not impacting transport in a catastrophic way. 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.

Risks of autonomous transport are not well known. To speed up learning, experiences, regulations and relevant incidents should be gathered and disseminated from all modes -autonomous road systems, air transport, rail and shipping. Safety of autonomous systems are dependent on new designed technology, human factors and organisational issues, [29]. The perception should be that most accidents in autonomous systems are a consequence of design and poor testing, and that “human errors” are a consequence and not a cause as described in [54]. Moving trivial functions 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 of humans, must get a significant position when automation is designed and when accidents are analyzed. Accident investigation boards should explore accidents across all modes, to support rapid learning and changes. The scope of the accident investigation should be based on an ecosystem perspective. Accident investigators should improve methods for investigation

using methods that highlights human factors such as HFACS, in collaboration with international actors to share best practices.

References

[1] Johnsen S & Håbrekke S 2009 Can organisational learning improve safety and resilience during changes Safety, *Reliability and Risk Analysis: Theory, Methods and Applications,* 805-815.
[2] Duffey R and Saull J 2002, Know the Risk: Learning from Errors and Accidents: Safety and Risk in Today's Technology: Butterworth-Heinemann ISBN-13: 978-0750675963
[3] Heikoop D D Hagenzieker M Mecacci G Calvert S Santoni De Sio F & van Arem B 2019 Human behaviour with automated driving systems: a quantitative framework for meaningful human control. *Theoretical Issues in Ergonomics Science,* **1**-21.
[4] Renn O 2005 Risk Governance – Towards an Integrative Approach *White paper no.1, IRGC.*
[5] 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
[6] Hoefer C 2003 Causal Determinism.
[7] Vagia M Transeth A A & Fjerdingen S A 2016 A literature review on the levels of automation during the years *Applied ergonomics,* **53**, 190-202.
[8] Parasuraman R & Riley V 1997 Humans and automation: Use, misuse, disuse, abuse. *Human factors,* **39(2),** 230-253.
[9] Manikas K & Hansen K M 2013 Software ecosystems–a systematic literature review *Journal of Systems and Software,* **86(5),** 1294-1306.
[10] Smith B W 2017 Automated driving and product liability *Mich. St. L. Rev.,* **1.**
[11] Sheridan T B & Verplank W L 1978 Human and computer control of undersea teleoperators Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.
[12] SAE 2016 International standard J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Revised: 2016-09-30
[13] Hollnagel E Woods D and Leveson N 2006 Resilience Engineering *Ashgate.*
[14] MTC 2016 Ministry of Transport and Communications: Testing of autonomous road transport systems from www.regjeringen.no
[15] Shappell & Wiegmann 2000 The human factors analysis and classification system-HFACS.
[16] Blanco M Atwood J Russell S Trimble T McClafferty J & Perez M 2016 Automated vehicle crash rate comparison using naturalistic data *Virginia Tech TI.*
[17] Axelrod C W 2014 Reducing software assurance risks for security-critical and safety-critical systems *Systems, Applications and Technology Conference (LISAT)* IEEE (pp. 1-6).
[18] GCIG 2016 Global Commission on Internet Governance One Internet www.ourinternet.org
[19] Koscher K Czeskis A Roesner F Patel S Kohno T Checkoway S., ... & Savage,S 2010 Experimental security analysis of a modern automobile. *IEEE Security* (pp. 447-462).
[20] DHS 2015 Department of Homeland Security, Office of Cyber and Infrastructure Analysis *The Future of Smart Cities: Cyber-Physical Infrastructure Risk*
[21] Cerrudo C 2015 An emerging US (and world) threat: Cities wide open to cyber attacks Securing Smart Cities – *White paper - IOActive.*
[22] Frei S Schatzmann D Plattner B & Trammell B 2010 Modeling the security ecosystem-the dynamics of (in) security *Economics of Information Security and Privacy* (pp. 79-106).
[23] NTSB 2017 https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1702.pdf
[24] The National Highway Traffic Safety Administration 2017 Office of Defects *Investigation PE 16-007. Automatic vehicle control systems, Tesla S accident in Florida May 7 2016.*
[25] 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
[26] Recode 2017 www.recode.net/2017/3/16/14938116/uber-travis-kalanick-self-driving-internal-metrics-slow-progress
[27] 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
[28] WHO 2018 World Health Organization *Global Status Report on Road Safety*
[29] Cummings M L & Ryan J 2014 Who is in charge? The promises, pitfalls of driverless cars *TR News* **292**, 25-30.
[30] BronnoyKalk 2019 www.prnewswire.com/news-releases/volvo-trucks-provides-autonomous-transport-solution-to-bronnoy-kalk-as-300753545.html
[31] KPMG 2019 Autonomous Vehicles Readiness Index, retrieved at https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/02/2019-autonomous-vehicles-readiness-index.pdf
[32] SINTEF 2017 Public Consultation on Norwegian regulation of autonomous (self driving) Cars of 1st March 2017 : www.sintef.no/projectweb/hfc/sarepta/publikasjonerreferanser/
[33] CAA-Civil Aviation Authority 2016 Regulations for Remotely Piloted Aircraft Systems (Forskrift om luftfartøy som ikke har fører…) from lovdata.no/forskrift/2015-11-30-1404
[34] CAA-Civil Aviation Authority 2017 Training requirements – retrieved from luftfartstilsynet.no/selvbetjening/allmennfly/Droner/
[35] Petritoli E Leccese F & Ciani L 2017 Reliability assessment of UAV systems. *IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace)* (pp. 266-270).
[36] EASA 2016, Prototype Commission Regulation Unmanned Aircraft Ope. *Note,* **22, Annex 1**.
[37] Altawy R & Youssef A M 2017 Security, privacy, and safety aspects of civilian drones: A survey *ACM Transactions on Cyber-Physical Systems,* **1(2),** 7.
[38] 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
[39] Kaber et al 2019 www.ise.ufl.edu/kaber/publications/supplemental-information-for-publications/ ref *Enhancement and application of an unmanned aerial vehicle supervisory control interface evaluation technique: Modified GEDIS-UAV.*
[40] UITP 2013 Observatory of Automated Metros *World atlas report* International Association of Public Transport (UITP), Brussels
[41] Wang Y Zhang M Ma J & Zhou X 2016 Survey on driverless train operation for urban rail transit systems *Urban Rail Transit,* **1**-8.
[42] Maritime Strategy 2015 Maritime muligheter – blå vekst https://www.regjeringen.no/
[43] National Transport Plan 2017 Nasjonal transportplan 2018–2029 Stortingsmelding#33 www.regjeringen.no/no/dokumenter/meld.-st.-33-20162017/id2546287/
[44] Ocean Strategy 2017 Ny vekst, stolt historie https://www.regjeringen.no/
[45] Rødseth Ø J 2017 From concept to reality: Unmanned merchant ship research in Norway *Proceedings of Underwater Technology (UT)* IEEE ISBN 978-1-5090-5266-0. OCEAN
[46] Wróbel K Montewka J & Kujala P 2017 Towards the assessment of potential impact of unmanned vessels on maritime transportation safety *Reliability Eng. & Systems* 155-169.
[47] Hoem Å S 2019 The present and future of risk assessment of MASS: literature review *ESREL*
[48] Rødseth Ø J 2019 Defining Ship Autonomy by Characteristic Factors *Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships*. SINTEF Academic Press.
[49] Hetherington C Flin R Mearns K 2006 Safety in shipping: human elem. *J Saf Res* **37**:401–411
[50] EU 2016 On the protection of natural persons with regard to the processing of personal data and on the free movement of data; *Regulation of the EU and of the Council of 27 April 2016*
[51] Iozzio C 2016 Who's Responsible When a Car Controls the Wheel? *Scientific American* **314**(5)
[52] Martin J Kim N Mittal D & Chisholm M 2015 Certification for autonomous vehicles. *Automative Cyber-physical Systems course paper,* University of North Carolina, NC, USA.
[53] COMEST 2005 The Precautionary Principle *UNESCOs World Commission on the Ethics of Scientific Knowledge and Technology*
[54] 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