The MBED (Maturity Based Evidence Dependent) Regulatory Framework for Highly Autonomous Systems

Joseph Morelos, Lloyd's Register EMEA, Southampton

Abstract:

This paper articulates the opportunities but also unique challenges that autonomous systems present the marine industry. It describes and explores the experience of aerospace and automotive industries in developing increasingly autonomous assets. Resulting from the unique characteristics of systems that learn and adapt, it elaborates the different certification and regulatory challenges. In response it proposes a unique regulatory framework for autonomy based on the maturation of capabilities – a licencing approach similar to how aspiring captains, airline pilots, doctors and similar professions are evaluated and monitored.

Introduction

Autonomous systems have the potential to revolutionise the transportation industry including marine in the same way that airplanes, cars, ships and trains have radically changed business and society during their widespread commercialisation in the last century.

Technology companies have led the advancement of autonomous capabilities to date. Progress in artificial intelligence, specifically machine learning techniques underpin the sophisticated products and services offered by Amazon, Apple, Google, IBM, Microsoft, etc. across industries as diverse as advertising, self-driving cars, finance, medicine and social media. [1]

As the financial gains from the use of autonomous systems become clear, the disruption of traditional business models and displacement of incumbents became ubiquitous.

In the logistics industry, the rapid development, integration and evolution of autonomous systems began with the "last mile" creating the efficient eco-system that online shopping depends on today. [2] In the continuous quest for customer satisfaction at the lowest possible cost it becomes a question of when not if autonomous systems will be adopted in the "first mile". Lloyd's Register is developing the regulatory framework, tools and technical expertise to assist marine clients in this journey, enabling the safe creation, integration and maturation of autonomous systems across all facets of the maritime industry.

The Size of the Prize

There are two fundamental technological changes in the marine industry today, digitalisation which includes applications such as autonomy and decarbonisation mandating the adoption of environmentally friendly propulsion for ships. These two themes have profound consequences beyond the maritime industry given that ships transport over 90% of the world's trade and is critical to the world economy. The insurance company Allianz reported 2,611 casualties in 2016 including 85 vessels within the total loss category and

hundreds of minor losses and claims.[3] While the severity and number of casualties have seen a steadying decrease over the last few years, the aggregated value of compensations remains in the range of a billion dollars. [4] Investigation of the different claims identify human error as the root cause in 75-90% of the accidents. Increasing autonomy in different ship systems (e.g. navigation, object detection, collision avoidance, health management systems, etc.) can help prevent and mitigate these failures and accidents.

However the impact of autonomy goes beyond reducing maritime accidents and saving human lives. There are obvious potential benefits such as increasing maritime supply chain efficiency with better management of vessel traffic and congestion particularly around crowded ports and waterways. Drawing from the experience and advances in last mile logistics, autonomy and the resulting "on demand" network effects can create efficiency that increases capacity of the installed base. There are numerous industries where autonomy directly increased capacity, which in turn has driven more demand. The exemplar being Amazon who currently provides a best in class 2-hour delivery window from order initiation of selected products. The superior customer service offered by Amazon has revolutionised the retail industry, shifting and creating new demand for their retail and technology services. [5]

From Not Working to Neural Networking

In the past autonomous systems have been severely limited in their ability to characterise their environment especially dynamic phenomena inherent in most environments. The more complex, dynamic and unstructured the environment, the more difficult the process of creating situational awareness and correct operational decisions become. This was evident during the series of DARPA (Defence Advanced Research Projects Agency) competitions beginning with the grand challenge of 2004, where 15 self-driving cars competed in a desert race for a $1 million cash prize. None of the autonomous cars actually came close to finishing the 140- mile journey, the best from Carnegie Mellon university drove a distance of only 7 miles before being immobilised by a rock. The following year amazing progress was achieved with the 2005 grand challenge where five teams finished the gruelling 132-mile race. Even more significant was the 2007 race, set in a dynamic urban environment where self-driving cars had to run alongside other vehicles and obey driving rules – requiring the context and experience of city driving. [6] The capabilities self-driving cars went from failing to understand a desert setting to managing safe interactions within a complex city environment over a period of four years.

This remarkable achievement required advancements across several domains from sensor technology to computing hardware. However central to the leap in capabilities and performance of self-driving cars are the deep neural network (DNN) facilitating computer vision, object recognition, environmental awareness, behaviour recognition and navigational decisions among others. DNNs are a type of machine learning algorithms that are capable of absorbing multiple levels of representation and abstraction loosely based on the functionality of the human brain. Most of the recent progress in autonomous cars are centred on improved DNNs, from more efficient neural network architecture designs to the sophisticated graphics processing units (GPU) hardware– enabling increasingly larger neural networks to process and analyse even bigger data sets. [7]

The Need for Legibility and Transparency

A novel element of machine learning algorithms (MLAs) including DNNs is their capability to learn (to be trained), resulting in performance improvements by adapting to different situations, phenomena and nuances of the real world. The algorithms deployed for deep space exploration will be continuously modified by the data collected from the combination of environment, stressors, the spacecraft's health condition and operational missions to which they are exposed. [8] As the term "machine learning" indicates, the performance of the resulting algorithms is heavily influenced by the quality and quantity of data used in training and the learning methods employed.

Complexity is inherent in most machine learning algorithms (MLAs) making them challenging to certify. Despite more than a year of extensive testing, NASA's remote agent in Deep Space One had errors resulting from concurrent threads within the architecture that nearly resulted in a failed mission. These fault conditions never materialised despite extensive, multiple test runs based on pre-determined failure modes and had to be rectified during the live demonstration. [9] Complexity results in most MLAs being characterised as non-deterministic, creating speculation that they are impossible to certify. Given that most software functional safety/verification standards such as IEC 61508, DO178, ISO 26262 are founded on the principle of determinism, critics argue that these cannot be applied to evaluate MLAs. Misconceptions about non-determinism in highly autonomous systems have been thoroughly investigated by NASA and Rockwell Collins including their impact on safety and certification. The report concluded that while there can be non-deterministic aspects to machine learning algorithms, what is key is to be precise about the source and the mechanisms that generate the resulting unpredictability to understand their safety implications and develop the means to test the consequences. Furthermore it is the dynamic nature of the environment, stressors, operational conditions that then triggers adaption of the MLAs resulting in their optimisation that creates the appearance of unpredictability. [10] Given the permanence of randomness in most environments and operational settings, absolute determinism cannot be achieved. Other investigated sources of non-determinism include concurrency, the reliance on probabilistic MLAs and the uncertainty in the existence of solutions given a time deadline.

LR believes in building trust and confidence in highly autonomous systems by rationalising their behaviour and capabilities. To accomplish this, autonomous systems must exhibit legibility, transparency including the justification of analytical processes involving situation awareness, situation analysis, judgement and action taking events. Robust justification should be in place including the use of objective evidence in proving the correctness of autonomous systems to facilitate a responsible, phased substitution of the functions and responsibilities of competent and experienced crew. The consistency and repeatability in generating correct situation awareness, situation analysis, decision making and action taking should be tracked and documented, including errors, faults and failure events evaluated as part of the maturation and acceptance criteria of autonomous systems.

Autonomy Challenges the Existing Regulatory Framework

Rules and regulations mitigate complexity that result in fault and failure conditions mostly with the concept of redundancy. When an electro-mechanical equipment like a pump, compressor, auxiliary engine, etc. fails, a fully redundant unit starts and maintains the essential function. There are multiple electronic fire and smoke detectors installed within a space to provide sufficient coverage and redundancy to account for a single detector failure. Even static components such as pipes, carrying pressurised and liquefied gas are required to be of double wall design to mitigate the effects of leakage from the primary pipe.

The principle of redundancy does not address software reliability and the resulting failure modes. Electrical, mechanical and structural components exhibit failure characteristics based on wear and tear, operational exceedance and other widely understood phenomena. In contrast software including MLAs will mostly fail as a result of dormant design errors. This means that the failure modes of the deployed software, MLAs will likely be present in and exhibited by the back-up software given their highly correlated design and architecture. [11]

Another inherent feature of the Rules is the use of prescriptive requirements in describing the safety criteria for the design, construction and continued operation of the ship. Most of these prescriptive requirements target the materials, components, equipment and structures of the vessels rather than holistic systems. Software and MLAs cannot exist in isolation, requiring reliable integration with multiple sensors, computing hardware and connectivity in operating essential systems. Generating robust prescriptive requirements that cover different use cases becomes complicated when considering integrated systems. [12]

A further consequence of how the Rules were developed is that assurance and certification are reliant on submitting the exact information on materials, size, capacities, scantling, construction, testing, etc. for technical specialists and surveyors to verify. Given the complexity of autonomous systems and the unique characteristics of MLAs such as the ability to generalise, new verification techniques based on a systems approach should be developed to evaluate their safety including desired behaviours and performance. [13]

The LR Maturity-Based, Evidence Dependent (MBED) Regulatory Framework for Autonomous Systems

Autonomous systems will have to consistently generate the correct situation awareness, situation analysis, judgement and actions to be considered safe throughout their lifecycle. Lloyd's Register believes that autonomous capabilities without appropriate safety justification will find limited use in the marine industry. To this end the LR foundation through the University of York initiated the Safety of Robotics and Autonomous Systems programme to investigate and develop robust regulatory requirements including assurance techniques appropriate for autonomous systems. [14] This strengthens and supports the further development of existing LR Rules and Regulations relevant to autonomous systems namely, the Unmanned Marine Systems (UMS) code and Digital Ships ShipRight Procedure.

Inspired by the existing licensure environment evaluating master mariners, airline pilots and similar professions a maturity based

framework is well placed to evaluate the design, development, testing, and operation of an autonomous system. Throughout its lifetime an autonomous system will experience many changes such as degradation, failures), the environment, random events and situations that are yet to be encountered (the unknown unknowns) that will require some form of learning then adaptation. It would be unreasonable to expect an autonomous system to be robust anticipating all possibilities, let alone exceptional situations.[15] Initially the capabilities of autonomous systems should be evaluated against well-established functions of the system, recognised operational concepts, range of operating environments, known system faults and failure modes, relevant accident investigations and other applicable sources of information. While a maturity based framework can never guarantee perfect autonomous systems exhibiting exemplar behaviours, safety and performance - it can create sufficient evidence by consistently monitoring the track record of the autonomous systems. The implication being should an autonomous system fail to meet the behavioural, safety, performance criteria at any point during the design, development, testing and operation stage, these events are captured and evaluated as evidence for objectively imposing restrictions, limitations or in the worst case of repetitive failures - the withdrawal of approvals (licence).

Design and Development

During the design and development phase autonomy should be evaluated in terms of measurable benefits, such as improvement in safety and performance compared to a conventional system. If an autonomous system cannot provide meaningful advantages or the resulting complexity actually increases the risk compared to the conventional system, these become grounds for declining the use of an autonomous system until it can be modified and made safer. Following the establishment of measurable benefits, the assurance framework then evaluates the holistic evidence that the autonomous system can safely do what it claims to do in terms of behaviour, performance and safety. This begins with the assessment of the designer including their development methods, quality assurance procedures, testing, among other software development quality metrics. After successful qualification of the developer, the assurance framework advances to reviewing the development of a well bounded and defined autonomous system. This includes evaluating the training/learning techniques employed in developing the MLAs, the quality and quantity of training data, the capability to generalise, the topology among other technical requirements. [16] In addition the cyber-physical system is evaluated against the requirements of the LR Digital Ships ShipRight procedure. This risk based review verifies the reliability of the integrated systems, components forming part of the autonomous system covering diverse domains such as cyber security, data integrity, architecture, human factors, hardware, software, networks/connectivity among others. Furthermore the LR Digital Ships Procedure calls for software verification of AL4 and AL5 autonomous systems, requiring an integrated evaluation of the MLAs, the digital system and the physical system. As part of the verification, the relevant hazard requirements are specified and the level of reliance placed on MLAs and software are defined.

Testing

The early control software developed for industrial systems were of relatively simple

design, resulting in their safety and performance proven by applying well-known test runs. This resulted in software verification being mostly a scenario-based testing exercise. The software to be evaluated is embedded into a test rig that replicates the physical system with the specific inputs and outputs then the test rig initiates a series of test runs. Software errors are identified when the actual outputs do not match the range of the pre-determined outputs.[17] As systems grew in both hardware and software complexity, the methods of verification and testing also evolved. While traditional software testing is sufficient for conventional control software, it cannot accommodate the complexity of MLAs and autonomous systems. Given the range of scenarios, operating conditions, environment, etc. the test conditions become incomparably larger. Furthermore given the ability of some autonomous systems to adapt to changes in operational profiles, physical condition of the system, environment, etc. - a series of adaptation may effectively invalidate the previous testing and consequently the basis of the existing approvals. LR through partnerships are investigating these challenges to define and develop appropriate testing environments for autonomous systems.

Operation - Monitoring the Performance of Autonomous Systems

A key component of the maturity based framework is the through-life monitoring of the behaviour, performance and safety of autonomous systems analogous to employing condition based maintenance on equipment, systems, structures of the ship. Having the ability to detect and record anomalous and failure events as they occur, including the evaluation of the same offline will contribute to the improvement of safety and reliability of autonomous systems. This is a key differentiator of the maturity-based framework – the through-life accumulation of evidence and evaluation of autonomous systems based on discrete events rather than periodic surveys, complimenting the design review, risk based work and testing during the approval phase. This means an approved autonomous system making incorrect situation awareness, situation analysis, judgement, actions can be flagged without delay. The subsequent investigation can then impose certain restrictions, limitations or worst withdrawal of the approval contingent on the safety implications and the ability to modify the behaviour of the autonomous systems. Furthermore the practice of evidence collection, investigation of failures and imposition of restrictions should be triggered by the actual events - divorced from the periodic survey and audit cycles. This provides a safety feature analogous to licenced drivers being flagged by the police for driving irresponsibly. Autonomous systems consistently exhibiting bad behaviours with potentially serious safety consequences should be restrained and taken offline at the earliest opportunity.

Conclusion

Widespread adoption of autonomy in shipping can radically transform the maritime industry. Increasing the autonomous capabilities of different systems provide a wide range of benefits such as reducing collisions and grounding (navigation systems), improving energy consumption and minimising pollution (propulsion systems) and minimising congestion of busy ports to name a few. Having said this there are significant safety

challenges remain. Chief among these is the very nature of an autonomous system – its ability to learn and adopt to highly complex and dynamic environments, internal failures, mission profiles and operations. Testing by simulations and sea trials alone may prove inadequate to ensure the dependability of these autonomous systems. For this purpose Lloyd's Register is developing the appropriate regulatory framework, including tools and technical expertise to assist marine clients in this journey. Enabling the safe development, testing and usage of autonomous systems across the maritime industry.

REFERENCES

1. Bloomberg, Fortune
2. Amazon, DHL, Fedex, CB insights future of logistics
3. 2017 Allianz Safety and Shipping Review 2017
4. 2017 Allianz Safety and Shipping Review 2017
5. https://www.nytimes.com/2016/08/11/technology/think-amazons-drone-delivery-idea-is-a-gimmick-think-again.html
6. Then and now, the DARPA grand challenge https://www.darpa.mil/news-events/2014-03-13
7. https://www.nvidia.com/en-us/deep-learning-ai/
8. Source: NASA RAX Deep Space One Mission
9. NASA: Formal Analysis of the Remote Agent Before and After Flight
10. NASA/CR 2015-218702 Certification Considerations for Adaptive Systems
11. Fundamental LR Rules: Software Conformity Assurance to understand the development, quality and testing of the software rather than requiring software redundancy
12. LR DS (old CES) ShipRight requires consideration of 9 domains using a risk based approach. Requiring goals and functions rather granular prescriptive requirements for different use cases
13. LR DS (old CES) ShipRight. Assurance of RBD findings rather than prescriptive rules. System engineering verification approach where the test plan is derived from the RBD processes rather than reviewing/testing each component in isolation
14. http://www.lrfoundation.org.uk/news/2017/lrf-and-the-university-of-york-announce-12m-partnership.aspx
15. NASA/CR 2015-218702 Certification Considerations for Adaptive Systems, Chapter 7.6 Paradigm Shift: Licensure
16. 2002 NASA Verification and Validation of Neural Networks for aerospace applications
17. NASA and RIACS Verification and Validation of Artificial Intelligence