DIGITAL TRAIN CONTROL
FUNCTIONAL SAFETY FOR AI BASED SYSTEMS

Mike Erskine, BEng (Hons), Chem, MBA, FSE (TÜV), FIEAust, RPEQ
David Milburn, BEng(Hons), CEng, MIET, MIEAust CPEng, RPEQ

GHD

Summary

The rail industry has introduced automatic systems to mitigate human error such as the widespread application in many countries of Automatic Train Protection (ATP) and the more limited application or Automatic Train Operation (ATO). In recent years the car industry has been playing “catch-up” with automated driving applications such as Lane Keeping Assist, Auto Emergency Braking, and Adaptive Cruise Control. It is now poised to make a game-changing leap forward to Autonomous Vehicles (AVs). AV and Connected AV technology is rapidly advancing and the race is on to achieve SAE International Level 5 classification (full automation) for driverless cars for public roads. Transfer of these significant advances in technology have the potential to impact the rail sector in dramatic ways, prompting a myriad of challenges. The rail industry has the opportunity to capitalise on this technology to bring about widespread automation of open rail networks far beyond the current limitations of segregated metros and to augment the safety features of automated metros by embracing AV technologies, in particularly Artificial Intelligence (AI) and Machine Learning (ML). Safety and minimising risk remain at the heart of these challenges.

INTRODUCTION

Most transport systems require a human driver and/or human network operator to interpret real time information and respond to this information. The human driver controls the vehicle and network operator manages the network, following the defined rules. Adherence to the rules is often supervised with automated interventions to mitigate the risk of human error, for example, if the driver exceeds their authority (distance/speed) or an operator attempts to set a conflicting route. The driver/attendant and controller can also react to unexpected events (degraded, abnormal and emergency situations).

The notable exceptions are the increasing number of automated metros (GoA4) that successfully provide high performance services within segregated environments with minimal variance (pre-programmed). The next evolution is to provide similar levels of automation and performance on open networks (mainline and light rail) where the environment cannot be easily controlled.

The systems that are developed will need to consider the sensors available on-board and the information available from the connected network. Ultimately, as we move away from human driven vehicles and network management, outside the relative simplicity of the metro environment, autonomous systems will be required to handle all the unexpected events on the rail network, similar to autonomous cars, using Narrow AI. Autonomous rail has less variables, but potentially more catastrophic consequences. When managed correctly AI has the potential to significantly improve safety. However, to achieve societal acceptance, research from the autonomous car industry suggests that this may need to be about two orders of magnitude better than the average driver.

System safety is demonstrated against functional safety standards. This is the foundation for best practice risk management and assurance. Functional safety for railway systems is currently addressed by standards such as EN50126 [2], EN50128 [3], EN50129 [4], originally predicated on predominantly deterministic systems (based on the inputs the output can be predicted).

In the event that the driving functions and/or train controller (signaller) functions and/or network manager functions and/or maintenance scheduling are devolved to autonomous systems based on Artificial Intelligence (AI), then the approach to functional safety standards and methods, will need to adapt. The guidance from the generic functional safety standard (IEC61508) is that AI is ‘positively not recommended’ for a Safety Integrity Level (SIL) greater than 1, and at SIL 1 has ‘no recommendation for or against it being used’. It is clear functional safety standards need to catch-up with technology advances.

The IEC, recognising this situation, is actively developing a new International Standard for a risk management framework for AI [5]. The IEEE are also developing AI related draft standards, in particular a standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems [42] and a Guide for Architectural Framework and Application of Federated Machine Learning [43].

In current systems, Safety Critical Application Conditions (SCAC) that cannot be addressed by the technology are exported to the human domain and usually result in human tasks within the Rule Book. In autonomous applications all functions must be addressed by the system.

A suitable approach is required with society with respect to the use of AI. A period of increasing AI support for drivers (semi-autonomous), with a transition over to AI driving (autonomous), will be required. The expected approach is longitudinal testing (repeated observation), to develop an inductive long-term dynamic safety case (‘proven in use’). This would be delivered through aggressive training environments that use training sets and test conditions that place the system in ‘extreme’ operating environments. In-service experience would be supported by a driver presence to ensure AI algorithms are able to operate safety with the eventual operational demands. There’s no doubt that as we continue to move away from human driven vehicles, automated or autonomous systems will increasingly be required to handle unexpected events on the network.

Realising the full benefits of railway digitisation will depend upon the rail industries ability to successfully implement and manage AI. AI has the potential to enable substantial improvements in the operation and maintenance of rail transport networks for both rail operators and infrastructure managers to improve management decisions, lower costs and enhance competitiveness.

The current and future developments, opportunities and challenges of AI in railway transportation is summarised by the EU in a briefing paper [37]. This considers three key areas where AI can be applied:

- Intelligent train automation (train control);
- Operational intelligence (predict and prevent); and
- Asset intelligence (long-term performance of rail assets).

This paper focusses on intelligent train automation. However, operational intelligence, asset intelligence are important contributors to train automation because success in these areas limits the variability and frequency that the autonomous systems encounter and have to react to unexpected events (degraded, abnormal and emergency situations).

Rail operators are actively pursuing automation (GoA3/GoA4) for open rail networks, including two major European operators, SNCF (Société nationale des chemins de fer) and Deutsche Bundesbahn (DB).

SNCF the French national rail operator has established two consortia to develop two driverless train prototypes for an autonomous freight train, and autonomous express regional passenger train. AI elements for the autonomous passenger train include an automated driving module, capable of reacting to possible hazards on the guideway and for automatic train door closure on open platforms. SNCF’s aspiration is to deploy semi-autonomous trains by 2023 and fully autonomous trains by 2025 [39]. DB the German national operator, is also pursuing driverless operations, and has aspirations to achieve this by 2023 [40]. Other National operators such as Network Rail in the UK are focusing on the application of ATO at GoA2 (with a driver present) [38].

Perhaps the most significant development and precursor to higher levels of automation on open networks for both freight and ultimately passenger operations is the success of Rio Tinto’s AutoHaul project. This project has not resolved or mitigated all the issues associated with ‘Supervising the Guideway’ on open networks, but has been able to demonstrate a safety case to the satisfaction of Office of the National Rail Safety Regulator (ONRSR) for driverless GoA4 operation of heavy haul iron ore trains in an unsupervised open and remote environment of Western Australia. This application is driven by significant commercial and operational benefits. For example, “in a manual system, every time one driver ends their shift and another comes on board, the train needs to stop. On a typical journey a train will stop three times, adding more than an hour to the journey” [41].

It is clear that in terms of automation there is a significant difference between automatic systems applicable to segregated environments such a metros, and the autonomous systems required for open networks. It is important to establish a common language to distinguish between the characteristics of 'Automatic' and 'Autonomous' systems. To promote discussion, the following definitions are offered:

Automatic System: a system that performs task sequences based on pre-defined rules. The information required to understand the environment is provided to enable the system to undertake rehearsed actions (characterised as deterministic).

Autonomous System: a system capable of making independent decisions to respond to all cases in real-time, and in some situations without reference to pre-defined instructions. It must therefore manage the functions of perception, environmental awareness, and spontaneous decision making (characterised as non-deterministic and potentially stochastic).

Autonomous train operations on open networks will need to be equipped with sensors and algorithms that are very similar to those used in self-driving cars based on AI. The rail industry must also avoid technology fragmentation by maintaining standardisation / interoperability throughout the digitization and automation journey. This will be essential to achieve longer-term cost effective solutions on a scale and refresh cycle that can compete with road transport and that builds upon existing or planned investments in train control technology and mandatory requirements.

1 CURRENT SITUATION – THE RISE OF AUTOMATIC SYSTEMS

Human drivers and guards have been responsible for driving and managing the safety of trains. This responsibility has not always been straightforward but as societal expectations have changed and safety has improved the public have remained comfortable with this method of working. Throughout the history of railways there have been safety critical incidents that have resulted in the loss of life and occasionally significant loss of life. These occurrences have spurred changes in the rail environment, often with a focus on removing human error.

Many metros have embraced automation and have ceded the driving function to automated systems with a driver retained for degraded operations. Others have removed the need for a driver altogether. The two fundamental automatic systems to support or undertake the driving task are:

- **Automatic Train Protection (ATP):** fail-safe subsystem that supervises train driving by ensuring that speed and movement limits are observed and intervenes if these are exceeded; and
- **Automatic Train Operation (ATO):** subsystem that automatically drives trains, through control of acceleration and braking, including but not limited to accurate stopping at specified stopping positions and regulation, using operational data provided by a Traffic Management System (TMS) and under the supervision of the ATP.

The experience of automation on metros beyond the provision of ATP is now crossing over to mainline railways. Firstly, with the introduction of ATO with a driver present. Ultimately it is expected that driverless/unattended solutions will be developed and deployed on mainline networks.

1.1 Grades of Automation

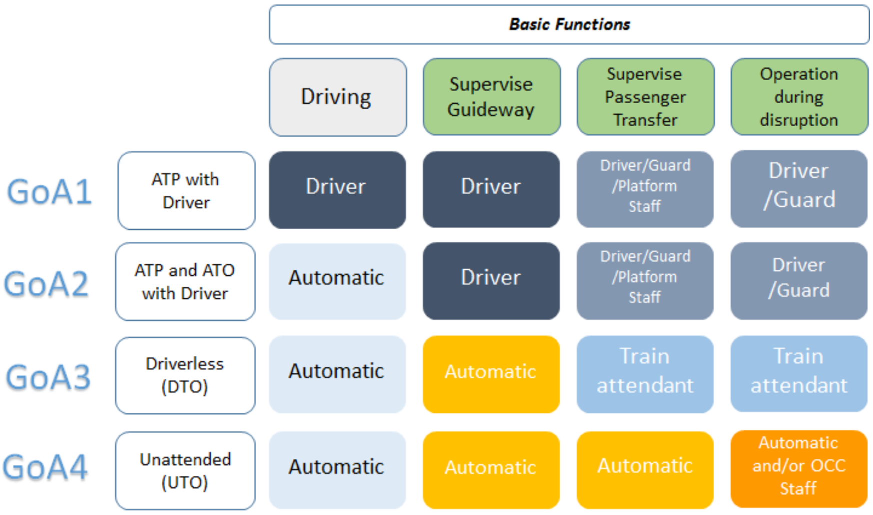
The IEC 62290 series of standards specifies the functional, system and interface requirements for Urban Guided Transport Management and Command/Control Systems (UGTMS). UGTMS covers a wide range of operational needs from non-automated to unattended. IEC 62290-1 [6] defines five Grades of Automation (GoA):

- GoA0: On-sight train operation;
- GoA1: Non-automated train operation;
- GoA2: Semi-automated train operation;
- GoA3: Driverless train operation; and
- GoA4 Unattended train operation.

The definition of GoA arises from apportioning responsibility for given 'basic functions' for train operations between operations staff and the system.

1.2 Basic Functions

The Grades of Automation defined in IEC 62290 [6] are summarised in Figure 1. Increasing automation is based on how the 'Basic Functions' are achieved.



		Basic Functions			
		Driving	Supervise Guideway	Supervise Passenger Transfer	Operation during disruption
GoA1	ATP with Driver	Driver	Driver	Driver/Guard /Platform Staff	Driver /Guard
GoA2	ATP and ATO with Driver	Automatic	Driver	Driver/Guard /Platform Staff	Driver /Guard
GoA3	Driverless (DTO)	Automatic	Automatic	Train attendant	Train attendant
GoA4	Unattended (UTO)	Automatic	Automatic	Automatic	Automatic and/or OCC Staff

Figure 1 – Grades of Automation and Basic Functions

As described in IEC 62290 [6], and illustrated in Figure 1, there are four 'Basic Function' required for train operations, these can be summarised as:

- driving;
- supervise the Guideway;
- supervise Passenger Transfer; and
- managing operation during disruption.

The functionality to support GoA1 and GoA2 concerned with ATP and ATO is within the traditional scope of the signalling and train control discipline. However, the functionality required to support GoA3 and GoA4 requires a broader perspective on what constitutes train control to safeguard the train in response to external events.

Furthermore, the complexity of achieving the 'Basic Functions' through automation depends significantly on the complexity of railway environment (number of tracks, junctions, timetable etc.) and the level of control over the environment that can be asserted. The level of control can be divided into two broad categorisations:

- segregated environments; and
- open environments.

1.3 Automation in Segregated Environments

Segregated environments are where the railway corridor is largely isolated from external influences. This is typical of metros in tunnels or elevated sections with robust security measures and complete separation from road users can be achieved and where passengers/public are unable or cannot easily gain access to the track (guideway) and other potential obstacles can be contained.

Automation in segregated environments is long established. The systems and concepts are very much 'proven in use'. The vast majority of existing automated metro lines are GoA2, there are relatively few GoA3 applications, the trend for metros is now firmly in the direction of GoA4.

The primary motivations for automated metros have been capacity, safety and energy efficiency. These can be achieved through the application of ATO at GoA2 and above. The pursuit of GoA4 is more focused on staff efficiencies (headcount, training, competencies, availability and scheduling) and service flexibility (introducing or removing trains to meet demand).

1.3.1 Benefits of Automation

The benefits of driverless metro operations are well documented in many papers and articles [8] and more convincingly by the growing commercial appetite for their deployment [7]. The benefits of automated operations are particularly well documented by Wang Y, Zhang M, Ma J, Zhou X [8] including:

- lower Operational Expenditure (in some cases 30% lower)
- increased capacity;
- improved reliability;
- increased flexibility;
- increased energy efficiency; and
- higher levels of safety and security.

1.3.2 Managing the basic functions in segregated environments

In segregated environments the ‘Basic Functions’ for metro operations can be relatively easily achieved, for example.

- driving – provided by ATO (for GoA2, GoA3, GoA4);
- supervising the Guideway – provided by Intruder Detection, CCTV with Video Analytics and where fitted PSD/PEB (for GoA3 and GoA4);
- supervising Passenger Transfer – provided by PSD/PEB (for GoA4); and
- operation during disruption - provided by staff in the Operation Control Centre (OCC) and/or staff with quick access to the train and infrastructure (for GoA4).

1.4 Automation in Open Environments

Rail networks that are not segregated are Open Environments and as such the environment cannot be easily controlled. For example, suburban lines, freight lines, high speed lines and mixed traffic mainlines. There has been very little in the way of automation on open networks other than the widespread introduction of train protection systems (GoA1) on mainline networks. This is because it is difficult with existing technologies (and those used for segregated networks) to automate all the ‘basic functions’.

However, the driving function can be automated and the use of ATO protected by ATP with a driver present (GoA2) is now emerging as a credible option beyond the metro environment. A good example is the Thameslink programme in London. The first project to implement GoA2 for high frequency passenger services (through the central section) using ATO over ETCS (AoE), this delivers a service frequency of 24 trains per hour (tph) [38]. Other projects are now following, for example, Digital Systems Program (Sydney) and Cross River Rail (Brisbane).

1.4.1 Drivers responsibilities

When considering AI for train control, it is useful to understand what train drivers currently do. The train driver has numerous responsibilities other than the obvious tasks of applying the brake and traction to comply with signals (or in cab authorities), adhering to the timetable, including stopping and departing at stations and opening and closing doors. The ‘hidden’ roles of the train driver are considered in detail by Karvonen H, Aaltonen I, Mikael Wahlström M, Salo L, Savioja P, Norros L [9].

In fact, the inspection of most operational railway Rule Books and driver policies and instructions will reveal a considerable number of tasks expected of a train driver which go beyond the basic driving task. Although, many of these task are relevant to both metros and mainlines, there are a far greater number of scenarios that need to be covered for open mainline networks. As an example, the Rail Safety and Standards Board (RSSB) Train Driver Manual (GERT800) [10] that compiles the content of the national Rule Book (GE/RT8000) for Great Britain, that is relevant to the role of the train driver is 948 pages long.

Some of key responsibilities of a driver, relevant to the higher grades of automation (GoA3 and GoA4) for in service functions on open networks, are the responsibilities related to dealing with:

- trespass;
- animals on the line;
- lineside fires;
- obstructions;
- floods and snow;
- road vehicles;
- level crossings;
- track workers;
- infrastructure faults/defects;
- train health/faults/defects; and
- train dispatch.

In terms of hazardous situations early detection increases the opportunity to avoid or mitigate the consequences. All too often, by the time the driver is aware of a hazard and reacts, it is too late to avoid an incident.

1.4.2 Driver Limitations

Drivers, like all people, have limitations and factors that influence and effect performance. Human performance is variable across any sample group. Factors that raise concerns and challengers for drivers and the management of drivers include:

- sleep apnea;
- fatigue;
- drugs and alcohol;
- mental health;
- training deficiencies;
- driver concentration;
- driver distractions; and
- driver attitude towards safety.

Rail statistics for accidents related to human performance are difficult to obtain. However, one example can be seen in the United States where the National Transportation Safety Board (NTSB) formally attributed six rail accidents to sleep apnea between 2000 and 20013 with nine fatalities [11]. More recently, the NTSB attributed two commuter railroad accidents within thirteen weeks of one another, to Obstructive Sleep Apnea (OSA) [12], [13].

The following provides further context. According to the National Transport Commission in Australia (based on data from the United States (National Motor Vehicle Crash Causation Survey)) human error and dangerous human choices cause up to 94% of serious road crashes, with causes including speeding, drink-driving, fatigue and distracted driving [1].

The consequences of human error and performance issues for both rail and road can be mitigated or avoided by introducing automation. In railway operations the basic application of ATP (GoA1) mitigates the fundamental safety risks associated with human error for the driving task (exceeding permitted limits). The further application of ATO (GoA2) eliminates variability in human performance and when well-designed creates a consistent, efficient high performance driving style.

1.4.3 Managing the basic functions in open environments

The transfer of the current methods for managing the Basic Functions from the segregated metro environment to the open environment is insufficient to provide automation beyond GoA2. Although, PSD and/or PEB can be introduced to manage passenger transfer achieving what could be described as

GoA2+. Additional functionality is required to deliver higher levels (GoA3, GoA4) of automation related in particular to the highly complex functions required for supervising an open guideway and managing abnormal, degraded and emergency situations.

1.4.4 Network complexity

In addition to usually being in an open environment, suburban lines, freight lines, high speed lines and mixed traffic mainlines are often part of a complex railway network and operational service model. This can often include features such as:

- multiple tracks;
- complicated junctions (nodes) with numerous routes;
- differential speeds;
- numerous rolling stock configurations;
- highly variable dwell times;
- highly variable turn-around times;
- numerous braking and acceleration characteristics;
- complex heterogeneous timetables with various stopping patterns and many different routes/services;
- train on train delays; and
- bottlenecks that create pinch points.

This adds significant additional complexity that automated metro systems are not required to manage. Automatic metros are usually characterised by their service simplicity, operating a headway service (no timetable or alternate routes), with a uniform stopping pattern on a simple route with identical rolling stock (homogenous).

The characteristics of the services provided by the earlier adopters of ATO for mainline operations are limited to services with simple characteristics similar to metro services. If mainline ATO is to extend beyond simple applications and become widespread then, in addition to managing the risks associated with open environments, network complexity will also need to be addressed. Managing network complexity will require a more capable ATO solution, supported by a TMS capable of providing both individual train and network wide optimisation.

2 FUTURE AUTOMATION – ENABLED BY SEMI-AUTONOMOUS & AUTONOMOUS SYSTEMS

Technology is advancing faster than at any other time in history and we stand on the brink of a fundamental technology revolution. This is generally referred to as the ‘Fourth Industrial Revolution’. Observing and decision making are key functions of the driving task. Until recently, only human drivers had the cognitive abilities, perception and complex reasoning required to safely navigate a train through varying shared, unshared and open environments.

In these environments, the driver is required to drive the train, following a prescribed set of rail/road rules, as well as being prepared to react to unexpected events, intrusions and obstructions. Automatic support for the ‘basic functions’ can be divided into two broad categories:

- **Driving Assistance** (Quality of service; punctuality, capacity, energy saving); and
- **Safety Assistance** (human factor containment; supervision, increased vigilance, early warning, evasive action).

2.1 Driving Assistance

On open networks, there are a number of existing products under the driver’s supervision that provide various levels of Driving Assistance. The current state of the art for Driving Assistance is ATO. As discussed, ATO solutions are well proven in providing Driving Assistance, testimony to this is the ever growing number of STO (GoA2), DTO (GoA3) and UTO (GoA4) metros and a few mainline GoA2 references. The future automation of open networks, therefore begins with the widespread adoption of ATO at GoA2 (with a driver present).

2.2 Safety Assistance

Many mainline routes across the world are fitted with ATP systems with various functionalities. The key variants are intermittent or continuous although many train operators and their regulators allow trains to be manually driven without ATP. The precursor for future automation on open networks must be the provision of a continuous ATP system as the minimum Safety Assistance provision. This provision will ensure that the ATO is always supervised by an independent ATP (SIL4 protection system).

Further safety assistance measures are required to manage the additional risks associated with open networks if Grades of Automation higher than GoA2 are to be achieved for open networks. Even where a driver is retained, additional Safety Assistance measures will improve the overall safety and performance of the train service. It is anticipated that these additional Safety assistance measures will be achieved using a combination of automatic and autonomous systems.

2.3 Interoperability and Standardisation

Across Europe there has been a twenty-five year-long (and ongoing) standardisation project, largely funded by the European Union (EU) to achieve 'Interoperability'. Although, this has been a long and at times slow journey, an open standard for a flexible radio based train control technology now exists in the form of ETCS Level 2 [14], [15].

ETCS is mandated for EU member states as the train control system to satisfy compliance with the Control Command and Signalling (CoCoSig) Technical Specification for Interoperability (TSI) [23] under the Interoperability Directive [24]. Automation on routes subject to the Interoperability Directives must therefore be compliant with CoCoSig TSI and hence be interoperable in accordance with the ETCS specifications. This legal necessity for EU states, combined with the commercial imperative to protect existing investments in ETCS technology, provides a compelling case that mainline automation must be ETCS compliant and integral to the ongoing evolution of ETCS.

2.3.1 ETCS deployment

Although, ETCS was developed to support interoperability across Europe, it is being rapidly adopted by operators worldwide whether they have borders to cross or not. ETCS has become the *de-facto* technology for new and upgraded mainline passenger, high-speed, freight and mixed-traffic railways. ETCS has been deployed on more than 80,000km of railway. Worldwide deployment statistics can be found on the Unife ERTMS website.

2.3.2 Evolution of ETCS

ETCS technology has started to encompass CBTC functionality, with individual suppliers tending to use the same, similar or slightly modified products in both forms of digital train control architectures. In particular, this can be seen in the recent introductions of ATO over ETCS.

The EU are continuing to invest in Advanced Traffic Management & Control Systems under the Innovations Programme 2 (IP2) work-stream within Shift2Rail [22], building on the ETCS specifications and experience form urban CBTC.

The enhancements to ETCS that are in various stages of specification, demonstration and development including, Automated Train Operation. This goal is likely to be achieved with the development of the standards for ATO over ETCS (Subset-125 [20], Subset-126 [21], Subset-130 [22], Subset-131 [23], Subset-132 [20] and Subset-139 [25] these are currently in draft) and ETCS Level 3 (and successors) and ultimately compliant products. The aim is to allow driverless trains to run on ETCS, initially at GoA2 and ultimately reaching a level of full automation (GoA4).

The ongoing enhancements to ETCS are illustrated in Figure 2.

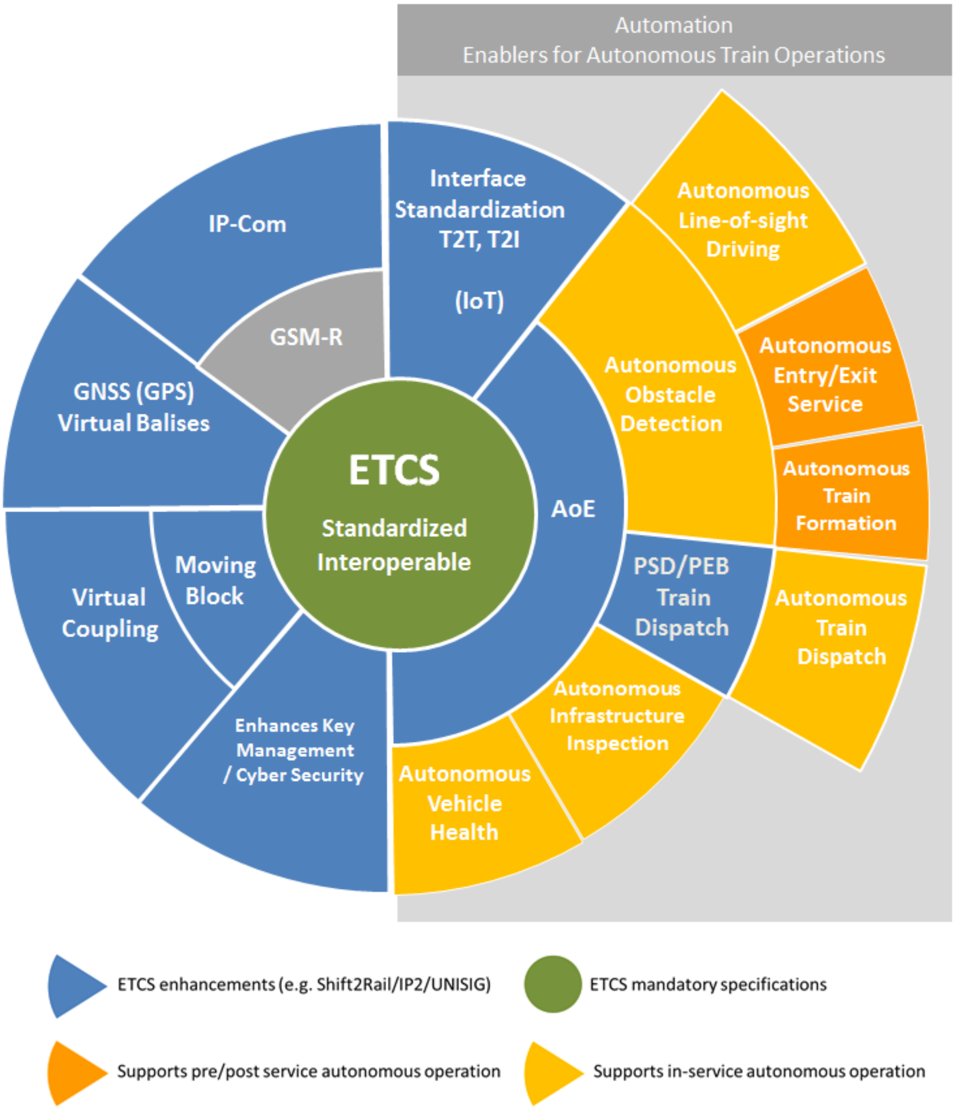

Automation
Enablers for Autonomous Train Operations
IP-Com
GSM-R
Interface Standardization T2T, T2I
(IoT)
GNSS (GPS) Virtual Balises
ETCS
Standardized Interoperable
AoE
Virtual Coupling
Moving Block
Enhances Key Management / Cyber Security
PSD/PEB Train Dispatch
Autonomous Line-of-sight Driving
Autonomous Obstacle Detection
Autonomous Entry/Exit Service
Autonomous Train Formation
Autonomous Train Dispatch
Autonomous Infrastructure Inspection
Autonomous Vehicle Health
ETCS enhancements (e.g. Shift2Rail/IP2/UNISIG)
ETCS mandatory specifications
Supports pre/post service autonomous operation
Supports in-service autonomous operation

Figure 2 – ETCS as the core to building open network automation

2.4 Enablers for open network autonomy

In addition to the enhancement being pursued by the Shift2Rail programme with respect to ATO. In order to achieve true autonomy on open networks a number of additional systems will be necessary to manage the additional risks presented by open networks (external events) and to allow operation when Full Supervision (FS) mode is unavailable or inappropriate.

These additional system will be required to have situational awareness so that they can provided varying degrees of what could be described as Reactive Train Safeguarding (RTS) to respond to events on and around the train and make safety related decisions – replacing (fully autonomous) or augmenting (semi-autonomous) the decision making undertaken by the driver. This autonomous system functionality is not part of an ATO system.

2.4.1 Autonomous system model

All autonomous systems are based upon a similar process, this can be presented as a simple repetitive three stage model:

- information;
- processing; and
- actuation.

The autonomous system continuously cycles through the stages. The model is illustrated in Figure 3.

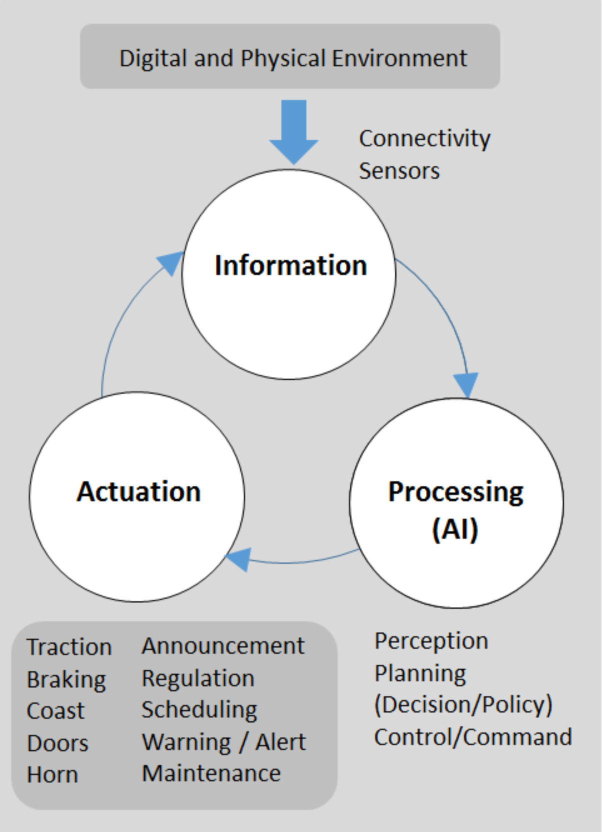


Figure 3 – Autonomous systems model

2.4.1.1 Information

The autonomous train system has to collect raw information about the status/condition of the digital and physical environment. This is achieved through two basic methods:

- Received from other connected systems (intruder detection, PSD/PEB, SCWS etc.); and/or
- Directly acquired by sensors (e.g. GPS, RADAR, LiDAR, IR camera, stereo cameras etc.).

This provides the raw information used by the autonomous system to develop an understanding of the dynamic physical environment.

2.4.1.2 Processing

The processing stage is the algorithms that provide understanding and decision making. In simple systems this can deterministic programming. However, in complex systems this will need to be provided by AI. The processing stage has three steps:

- **Perception:** refers to how and how well the autonomous system understands the raw information and can process the captured information (acquire, select, organize) and interpret this to establish a sufficiently accurate view of the real world appropriate to the function of the autonomous system (discern the difference between an animal and a person, discern the difference between a track worker and a trespasser etc.);
- **Planning:** refers to the ability of the autonomous system to apply policies and make decisions to achieve higher order goals in response to the current circumstances. This is achieved by combining the processed information about the environment (perceived view of the real would) with established policies, domain knowledge and learning regarding how to respond to the presented environment. This leads to the autonomous decision; and

- **Control:** is the transfer of decisions (intentions and goals) into actions commanded for Actuation.

2.4.1.3 Actuation

The actuation stage physically implements the decisions in the real world. Such as braking or accelerating. In the proposed ETCS based autonomous digital train control system the actuation for all supporting autonomous systems would be provided by ETCS and ATO over ETCS (AoE). This arrangement for Autonomous AoE is illustrated in Figure 4.

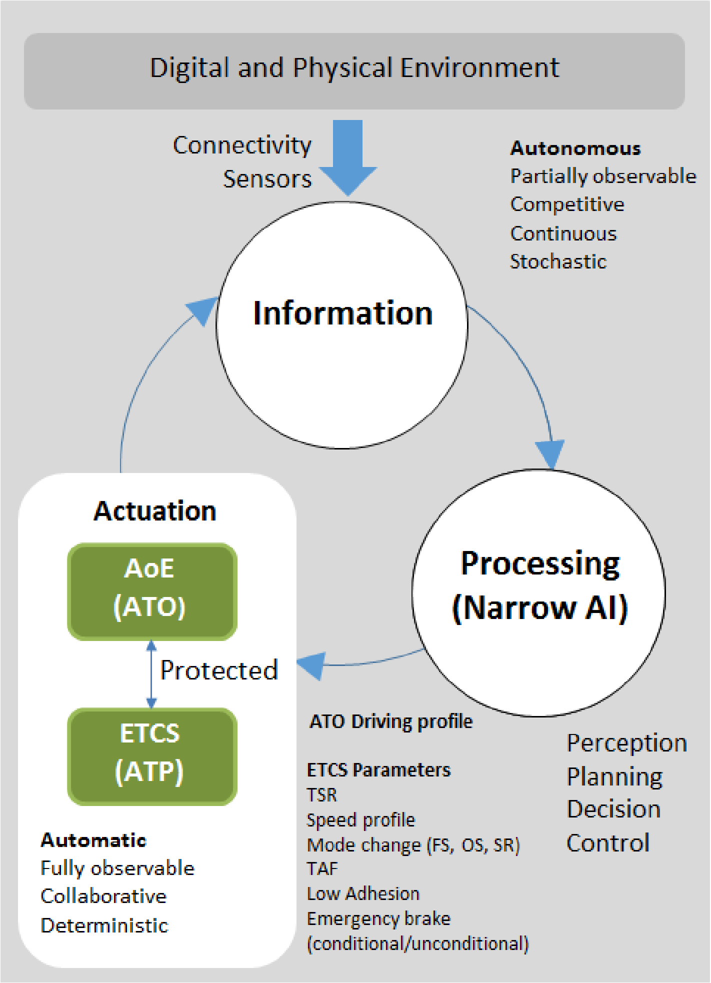


Figure 4 – Autonomous AoE model

This approach preserves all the benefits of ETCS and AoE and allows incremental enhancements to provide semi-autonomous and autonomous functions.

More importantly, the responsibilities of the ‘intelligent agents’ within the autonomous support systems are limited to the supplementary functions that ETCS or AoE cannot provide. All commands from the AI are supervised by ATP within the permitted Movement Authority (MA) or other ETCS values/parameters (e.g. On-sight (OS) ceiling speed). The AI will only be able to command more restrictive parameters or command the revocation of parameters applied by the AI.

2.5 Autonomous system for open network operations

The additional functionality required for open network autonomy can be divided into two groups of autonomous systems, those to support:

- in-service train operation; and
- pre and post service train operation.

The possible evolution of ETCS to support automation in open networks for both In-service train operation and Pre and Post service train operations are illustrated in Figure 2.

2.5.1 In-service train operation

The train can be driven by AoE, however this system cannot provide Reactive Train Safeguarding (RTS) in response to external events. Additional autonomous systems are required, to manage the open network risks. The primary autonomous systems necessary to provide RTS could include:

- Autonomous Obstacle Detection;
- Autonomous Line-of-Sight Driving;
- Autonomous Train Dispatch.
- Autonomous Infrastructure Inspection;
- Autonomous Vehicle Health.

The supporting autonomous systems from an ETCS perspective for 'in-service train operation' are considered in more detail by Milburn D [25].

2.5.2 Pre and Post service train operation

If the whole of the train operation (day in the life of a train) is to be autonomous, then further autonomous systems are required in order to provide Pre and Post service autonomy. Most notably, systems to provide:

- Autonomous Entry/Exit Service; and
- Autonomous Train Formation.

The supporting autonomous systems from an ETCS perspective for 'pre and post service train operation' are considered in more detail by Milburn D [25].

3 ARTIFICIAL INTELLIGENCE

The introduction and acceptance of Artificial Intelligence (AI) and Machine Learning (ML) will be central to the success or otherwise of autonomous train control systems on open networks.

3.1 Overview of AI

Before considering the introduction and management of AI to support train control automation it is important to understand the basic concepts of intelligence, AI and AV maturity. This section is based on a paper by Erskine M [27] that cconsiders AI generically from a Risk Management perspective.

3.1.1 Intelligence and AI

Artificial intelligence can be defined as:

Both the intelligence of machines and the branch of computer science which aims to create it, through "the study and design of intelligent agents" or "rational agents", where an intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success [26]

3.1.2 Coherent Extrapolated Volition

Coherent Extrapolated Volition (CEV) theory is *"meant as an argument that it would not be sufficient to explicitly program our desires and motivations into an AI. Instead, we should find a way to program it in a way that it would act in our best interests – what we want it to do and not what we tell it to"* [28].

There is a potential for AI to go further in what we would want, or in a new direction we didn't anticipate nor desire. Sometimes this can be good, other times, less so.

3.1.3 When humans secede control (trust), the societal permission point

This point isn't just an equivalent of replacing an existing human at current average levels of performance. It goes much deeper than that. For a train, it is the point where a highly trained person within a high integrity system operates. These people are in the very low error level because of high skills to get to that point. They also have double check systems i.e. train control, effectively taking them to a level about a factor of 10 or 100 better than an average person.

It is also hardwired into us from many millions of years, such that for us to let go of control, the system we are to trust has to be significantly better. This is partly because of ego, and partly because of the unknown. When travelling by train, we trust the driver and systems, because there is a benefit in using these services. It is a fragile balance. As the nuclear industry that routinely use the upper levels of Safety Integrity has found out, this societal trust is essential for continuity of business.

It is this region of about two orders of magnitude or better than society on average can do, is where society intuitively will accept highly automated systems. Whilst the safety of achieving this point may be well and truly much better than where we operate now for a certain mode of transport, the cost to achieve it is much higher. If targeting anything less than this point, but even well above current standards, societal acceptance likely won't be achieved, and financial viability is also unlikely.

3.1.4 Mountain of caution

AI for road vehicles is currently in what is called the Narrow AI category. It is akin to a sophisticated machine, but with less verbal interaction with the occupants.

With AVs, a different feature is emerging. The initial response in 2014-2016 to the awareness of AVs being a reality was for a "blue sky" of amazing capability. When it is a mature technology, the indications are that this could be quite possible. That initial view seems to be giving way to a "mountain of caution", based on initial trial results and further research. This "mountain" comes from a few sources:

- The first is from the natural fear or concern of the unknown. Regulators and politicians are normally conservative as many panaceas are put forward to them by a range of people and companies. Quite often, purported benefits don't turn out as predicted; and
- The second is the media fascination with new and different events affecting public opinion and perceptions. The recent Uber and Tesla fatality events were of concern. However, when AI driving is examined in context of overall risk level compared to human drivers, AI is significantly safer, i.e. lower in frequency of accidents per unit of distance travelled.

Perceptually, that AI trust region would appear to be in the order of a hundred times better than our current national average accident and fatality rate, or a 99% reduction in accident rate. This is difficult, because there is a delicate balance between cost to achieve this, and benefit gained, and societal ability to afford such technological performance. There are ways to possibly achieve this very difficult region of individual and societal acceptance. From an emotional perspective, we need to become familiar with the various forms of AI. This is perception, trust, integrity and the other features of the organisations bringing us this technology.

The "Mountain of Caution" needs to have definite attention and a strategy to deal with it. Formalised methods are required for processes of HF analysis and of stakeholder engagement in this respect. There is a path forward in this for companies contemplating significant AI content in its products and services. There are two main components. These can be broadly defined as technical and social. Both are needed, but also need to be complementary, and not independent of each other.

3.1.5 Complementary logic to current narrow AI

Whilst there is significant benefit in AI and other computational engines to determine what shapes are to take certain actions, there may be other risk-based approaches to augment reliability of correct identification. Some of these are subtler. The recent Uber accident has further highlighted the need. Our human intelligence looks at many features along with shape and size. Aspects like trajectory, velocity, background weather, and environmental influence versus purpose, all play a part. More of this can be programmed with the vehicles to perhaps make much higher accuracy identification of objects with potential to interact with the vehicle. Human identification whilst good, is let down by much poorer surveillance of the immediate environment, whereas current scanning AV technology is good and is improving at this.

Multiple deductive and inductive logic processes could be used, along with reductive processes for really high-quality decision making of professional drivers. The dynamic abductive reasoning process is also required for dynamically adjusting and for increased learning. These likewise need to be programmed into AI AV technology and verified as effective through a formal process.

3.1.6 Independence of Assessment of Human Factors and Stakeholder Issues

Internal HF and stakeholder assessment do have benefits, but they can also have constraints at times. The capitalist imperative and internally generated schedules can limit the quality, diversity and magnitude of input of these vital disciplines.

Properly regulated independent assessment can provide a much wider range of assessment benefit. This is partly because of the typical expertise residing in specialist firms, but also because of a much more powerful paradigm. These are outside people with the requisite specialist skills, looking inwards at the project with set regulatory requirements for the first time with the ability to articulate the key concerns of the general population. This is especially beneficial if there is a charter of values that can be referenced.

This has become the trend in recent large government projects in Australia where a range of societal issues and sensitivities exist and need to be managed. The emerging range of large corporate project such as AVs with AI would have a high level of legitimate HF issues and stakeholder concerns that come together as societal permission requirements before acceptance can occur. An important part of societal acceptance for large projects can be an open and independent assessment. Regulatory bodies have realised this and have developed refined processes for rail, aviation, and other key facets of society such as infrastructure design, industry and utilities.

3.1.7 AI development Causes leading to Hazards and Risks

Whilst it is difficult to predict risks for AI applications, it is possible to list some of the more likely causes giving rise to some scenarios. Likely discipline or application level issues associated with AI development and deployment are considered to be:

1. Setting timeline targets or goals that may be too ambitious for the budget or particular capabilities of the AI;
2. Not setting goals or targets that are necessary for the facility or organisation utilizing the AI, and having undue exposure to safety, environmental, social or reputation risk;
3. Quality of information available and how constant it may be for the life of the AI to utilize (possible allowance for retraining needs as information changes);
4. Programming AI without all of the necessary variables;
5. Programming the AI with too many trivial variables – i.e. less relevant and generates spurious results;
6. Programming the AI using older data, older statistical paradigms, perhaps not providing the fullest extent of newer capability;
7. Not using top tier specialists in the learning and checking processes leading to insufficient knowledge, or incorrect knowledge for application;
8. Lack of independent review phase of AI output testing leading to inadequate validation of required learning for the expected spectrum of issues;
9. Lack of suitable interaction between programmers, technical, sociological, environmental and financial specialists leading to deficiencies in the learning process for the AI;

10. Lack of suitable complexity of model to adequately reflect the current and projected situation leading to inability to deal adequately with a reasonable spectrum of issues;
11. Insufficient ethical training leading to safety and social issues;
12. Deliberate training or poisoning of the learning process with unacceptable material, leading dangerous safety, environmental and financial actions;
13. Negative perceptions by potential end users of the artificial intelligence output; and
14. Criticality of what the AI is doing, scenarios related potential harm to life, environment, financial, social.

From these and other risks, and existing processes that we are familiar with, we can look at a range of customised processes that will likely manage issues and have potential for the best output.

3.1.8 Risk Management of these New Standards

Applications of technology in society have become increasingly more complex. We have been broadly successful in developing more and more complex control systems to achieve this desired level of performance and control of these operations. This is driven by our societal expectations, and also the profit that comes with achievement. Along the way, this has generated new challenges to test these systems for reliability. New professional certifications and courses for risk management, including functional safety have been developed in recent years to help achieve and maintain these standards.

Previously, for technical applications, we have been able to verify the outputs of many calculations, except where complex programs are used. In those cases, other complex programs can be used as verification. This can get into the region of what is known as Segal's Law *[29]. A man with one watch always knows the time. A man with two watches is never really sure*. What this means is multiple outputs may yield different results. In risk management, this may be beneficial. If the outputs of a complex system analysed two or more different ways are marginally different but in the same range, then that is usually acceptable, depending on consequences of error. This has been used in the rail industry with complex systems as contained in EN50126, which has been primarily deterministic in its assessment methodologies. By contrast, AI systems are stochastic. Environments in which AI can be used can be defined as follows [34]:

1. complete/incomplete;
2. fully/partially observable;
3. competitive/collaborative;
4. static/dynamic;
5. discrete/continuous; and
6. deterministic/stochastic.

It is these environments that determine what sort of data and how much is needed.

AI applications require different approaches to achieve the really high levels of reliability and performance that we would now routinely achieve. How does one get an AI system operating at better levels than a person can typically achieve? The answer lies in the careful risk management of programming, data to feed it, and suitable programming that will likely achieve reasonable CEV for the AI, i.e. not too constraining, and not too loose. Almost, like a good parent, suitably stretching their children as they develop.

Current AI approaches include statistical methods, computational intelligence, and traditional symbolic AI. Many tools are used in AI, including versions of search and mathematical optimization, neural networks and methods based on statistics, probability and economics. The AI field draws upon computer science, mathematics, psychology, linguistics, philosophy, neuroscience, artificial psychology and other disciplines.

Like people, the AI typically needs the collective benefit of numerical and graphical and textual knowledge and calculations that we can muster for a wide range of scenarios. It needs to be taught "life" skills for the key areas of its operation.

As with people, AI learning will need to be a continuous process as the environment in which it operates continually evolves. Risk management will likewise need to be active through all steps of the learning and operational phases. Some of these are outlined below:

1. Context and Interaction – proper definition and critical mass of context, including CEV consideration. Proper definition of interaction, safety, societal, business and other.
2. Initial Learning – basic training for functionality;
3. Advanced learning – higher performance requirements;
4. Learning Assurance – testing for a wider range of scenarios, or as systems become more complex; and
5. Ongoing Assurance – continued verification of performance, (with software or hardware upgrades, environment changes, or general time-based verification).

3.1.9 Context and Interaction

This contains all the elements of basic risk management as outlined in ISO 31000. As the context becomes exponentially more complex as more social components and output interactions are included, the propensity for problems likewise increases. Key features required are a systematic identification of the environment and the hazards to all stakeholders within and associated with that environment. The level of consequence and frequency will determine how much effort is needed here. This can be expressed in several ways:

1. Insufficient detail and scenarios of the key identified areas required;
2. Insufficient key areas identified;
3. Insufficient HF, societal factors and environmental factors and related scenarios considered;
4. Insufficient scoping of tolerability and acceptability of hazards, events and associated probabilities; and
5. Unbounded limitations definition (boundary) setting of system capabilities with reference to the situation at hand.

Whilst it is important to frame the context properly, there are opportunities to capture issues later. However, not doing enough context analysis knowing that later steps may capture shortcomings is neither recommended nor encouraged. Context setting for broader AI applications perhaps requires more effort than previous technologically predominant/safety challenges. The rail industry typically does put significant effort into their context and risk assessment processes because of the potential consequences of safety critical failures in terms of injuries and fatalities.

AI can benefit from a large learning population, and over a much larger time domain than humans, an emerging example of this is Tesla vehicles [35].

3.1.10 Fallacious decision-making and dynamic environment testing

This is equally important to assure as the validation of good reasoning. There are two aspects to this that need to be considered:

1. What is the potential for a wrong decision or output given good validated training and information (system construct, information and human corruption)? Are there weaknesses in the development of the system?
2. What is the potential for fallacious decision making (error potential) when in operational mode and conflicting inputs or change of environment occur. Will the AI continue to perform with the

Coherent Extrapolated Volition of our originally desired goals in a dynamically evolving environment?

As has been seen, the applications of AI will be much greater than the constrained environments we have often made or designed our equipment or processes for. With technology to date, some of the risks that have occurred are ones where the environment has altered. Our typical response has been to learn from mistakes, or to anticipate and engineer the equipment for an expected but unlikely scenario.

Time dynamic environments will require a level of risk management that isn't commonly utilised. Increasingly, this will occur in social settings, where the societal response may change with time, and even in response to the use of AI in that application.

3.1.11 AI and current Safety Critical Systems

Some recent thinking has emerged in regard to using AI for safety critical systems. Typically, high integrity safety systems are deterministic in nature. AI goes beyond deterministic thinking, i.e. if pressure exceeds "X", then open valve "Y" until pressure "Z" is achieved, and then close valve "Y" again. AI can go well beyond numerical stochastics as well, which is inductive testing by nature. This is where certain patterns occur, then a certain situation is likely (to some extent) to happen, so a response is required. Contextual testing is also limited, i.e. how well does the testing approximate all that could occur in the field? Have all scenarios been contemplated, and does the system recognise a new situation as a deviation to current experience, or as a new situation without a defined response? Longitudinal testing is continued testing over a longer period of time to gain greater exposure to a wider range of events and to test responses. Longitudinal testing has limitations, but has been used so far.

Therefore, testing and assurance regimes need to reflect this inherent situation.

"Although AI will be an engine for progress in many areas, creating real-world systems that realize these innovations will in fact require significant advances in virtually all areas of computing, including areas that are not traditionally recognized as being important to AI research and development… future AI systems will not only draw from methods, tools, and themes in other areas of computer science research, but will also provide new directions for research in areas such as efficiency, trustworthiness, transparency, reliability, and security" (Hager, Bryant, Horvitz, Matarić, and Honavar 2017). These authors identify the development of formal methods as a key enabler for the deployment of AI techniques in dependable applications [30].

In essence, many people are describing the key components of competency based training and all that goes with it. Essentially, AI, like humans, needs to reach an acceptable level, and every now and then, have some retesting to ensure bad habits or wrong learning don't creep in. Transparency will likely be a problem with AI systems for a long time to come [31]. Academics have been grappling with this issue but perhaps ignore some broader features of the limits of human centred logic and resources. Some approaches like Generative Adversarial Networks (GAN's) could be what is required to train AI in certain cases, as the human limits are reached.

It is possible that combinations of deterministic and inductive systems may provide the best overall response. This is akin to our deductive logic in combination with our inductive reasoning for a given situation, and its natural variation, to arrive at a logical action that is most likely to be correct.

The IEC has been tasked with the exercise of developing a new international standard for a risk management framework for AI [5]. It is a substantial issue, and is evidenced by recent conferences starting to emerge on this topic [32]. Currently, the functional safety standard only allows AI up to SIL 1 [36] for fault correction. However, it is neither recommended nor not recommended.

It is in the upper reaches of the functional safety standard that we find some useful principles that relate to development and use of AI in safety critical areas. A high hardware fault tolerance (HFT) along with a high integrity (SIL 4) is needed for a very safe and reliable system. This requires both elements of software and hardware control to achieve this outcome. Critical applications like rail crossings and other key rail interaction areas form part of this regime.

For safety critical and other system, it is not enough to just have AI technology within a technical product. Organisations need to have suitable resources for developing and training the AI, including appropriately developed ethics input and a functional management structure [33].

Regulatory authorities need to appropriately update the current So Far As Is Reasonably Practicable (SFAIRP) philosophy for safety critical systems in the light of what is reasonable for AI training and assurance. They will also need to better articulate the means by which assurance of SFAIRP will then need to be made. If developed suitably, the proposed IEC risk framework for AI and safety critical systems will be very important in these aspects.

3.2 AI for Train Control

AI is already a reality and proven in numerous applications. However, it is likely to be a significant challenge to introduce this into train control applications (Line-of-Sight Driving) and other related railway systems that make safety related/involved decisions (for example, condition led maintenance interventions).

Many existing systems outside the rail industry successful use Narrow AI. This can perform better than humans in a restricted field of information and outputs or actions. For example, simple autonomous vehicles (AVs) in a closed transport environment, such as a single corridor route.

AV and Connected AV technology is rapidly advancing and the race is on to achieve SAE International Level 5 classification (full automation) for driverless cars for public roads.

The technology required to support autonomous train operations on open rail networks as described in this paper requires far less complexity, as the application and safety responsibilities of the AI agents have been minimised and the environment is already partially controlled.

The train can already drive its self under normal operating conditions, with high integrity protection, using proven technology. The AI has to bridge the gap between this capability and the additional functionality required to manage unexpected events and line of sight movements.

3.2.1 AI/ML challenges for train control

There are two major challenges to overcome:

- The first major challenge will be the transfer of technology to the rail domain, not the absence of appropriate base technology; and
- The second and more significant major challenge, will be the emergence of fundamental changes to how functional safety is perceived and the risk assessment techniques employed.

There is clear agreement amongst academics and developers that the existing probabilistic risk assessment methods are insufficient to support the development and implementation of safety critical embedded AI and Machine Learning (ML) algorithms. The deficiencies are clearly articulated by CW Johnson [30] with respect to existing practices for the risk assessment of Human Factors, Software and the emerging risk assessment of AI/ML.

This is argued on the basis of:

- limits of Determinism (predicting the outcome);
- limits of Induction (bottom-up reasoning);
- limits of Deduction (top-down reasoning); and
- limits of Context (influences).

In simple terms, in complex systems, there are far too many unknowns to accurately predict all outcomes (likelihood and consequence of failures). As a proxy, rigorous processes are applied for the development of pre-programmed (deterministic) software algorithms to achieve a desired Safety Integrity Level (SIL). This is essentially the reliability of the safety functions (and only the safety functions).

AI/ML algorithms are not pre-programmed (other than with a learning mechanism), they are taught and they learn. As such AI/ML are non-deterministic (stochastic). This is one of the key strengths/benefits. The programmer does not need to envisage and code all scenarios.

AI/ML algorithms require minimum programming and minimum domain knowledge or expertise to be highly successful at a given ‘narrow’ task. However, this strength is also the greatest weakness when applying established risk assessment techniques to develop a robust safety argument. Many applications, in particular driverless car applications need to overcome this limitation, otherwise the much converted, go anywhere, ‘robo-taxi’ will not become a reality.

The approach to this conundrum has almost universally been to rely upon longitudinal field trials, rather than predictive safety analysis and formal methods. These trials are used to establish inductive safety arguments and are a convenient extension of the ‘proven in use’ concept (refer to EN 50126 [2]). This, and any other approach, raises two obvious and philosophical questions of confidence:

- When do we know that we have done sufficient trial running to know that the potential software failures are at an acceptable level?
- How do we know when have we have encountered sufficient contextual variance to adequately cover all potential real world scenarios?

The answer to these questions, or at least the basis on which to build the argument for sufficiency, are not covered by any current standards for rail or road (or any known standards for any relevant industry). Furthermore, as yet there are no formal techniques available for predicting the behaviour of system that include AI/ML algorithms.

3.2.2 Progressive system assurance for train control AI

The acceptance of the embedded AI/ML algorithms that support the semi-autonomous and autonomous train applications as described in this paper, could be enabled by a series of progressive assurance phases.

- Phase 1: Static training;
- Phase 2: Dynamic training;
- Phase 3: In-service longitudinal testing and training.

3.2.2.1 Phase 1: Static training

This starts with the definition of the initial training set for the specific AI/ML application (autonomous system). The initial training set is essentially the equivalent to the specification for a deterministic pre-programmed algorithm. However, unlike a traditional specification the training set will not be a complete set of requirements as all aspects of the operating environment cannot be fully defined.

During the initial training process the AI/ML will be subject to aggressive training within simulated environments, using training sets and extreme operating conditions and rapid training cycles.

3.2.2.2 Phase 2: Dynamic training

Depending on the autonomous application, the trained model (AI/ML) will then progress to dynamic training/testing in a non-operational (low risk) environment and subject to further rigorous operating extremes.

3.2.2.3 Phase 3: In-service longitudinal testing and training

When sufficient confidence has been established, the AI/ML can be introduced, under strict controls, into operation on in-service trains under the supervision of a responsible driver. This is where the longitudinal testing will be undertaken on numerous vehicles over an extended period. This approach will build-up millions of hours of augmented experience (the collective experience of all the fitted vehicles). The rate of progress will be highly dependent on the size of the population of vehicles using the particular system, the complexity of the associated environments and the breadth of events encountered.

The goal is to reach a point where a compelling argument can be made that the confidence level in the capability of the autonomous system establishes that the risks have been reduced So Far As Is Reasonably Practicable (SFAIRP). The autonomous (AI/ML) system can then be declared to be *‘proven in use’*.

3.2.3 Ensemble learning and federated learning

One very distinct advantage of AI deployed in multiple similar/identical applications is that of *ensemble learning*, where the overall AI learns from the sum of all of the deployed units. This is currently being utilized by Tesla for their AI training to achieve level 5 automation. The same could be done for trains. As an example, this would have benefit in determining the nature of trespassing on the rail corridor, and learning appropriate actions to take to minimize harm in each class of trespass. It is also possible for application of feed forward (predictive) mode, which could be a unique game changer for rail operations for trespass and various derailment scenarios.

To maximise the overall benefit of ensemble learning, it would be useful to develop arrangements for the federated ownership of data for rail systems. This could be very important for the trespass learning example, as well as other learning, which needs to be open, and ethical for train operations that utilizes AI. The draft IEEE Guide for Architectural Framework and Application of Federated Machine Learning [43], defines a machine learning framework that allows a collective model to be constructed from data that is distributed across data owners, while meeting applicable privacy, security and regulatory requirements. The draft IEEE guide includes:

- description and definition of federated learning;
- the types of federated learning and the application scenarios to which each type applies;
- performance evaluation of federated learning; and
- associated regulatory requirements.

The rail industry should consider collectively defining the architectural framework and application guidelines for federated machine learning.

3.2.4 Fail-safe

A fundamental concept for signalling and train control is ‘fail-safe’. This is where the system is designed such that a fault (failure mode) leads to a safe state. Autonomous and Semi-Autonomous Systems for train control will need to be designed to be fail-safe. The IEEE are developing a draft standard for the Fail-Safe Design of Autonomous and Semi-Autonomous Systems [42]. The draft IEEE standard:

- establishes a practical, technical baseline of specific methodologies and tools for the development, implementation, and use of effective fail-safe mechanisms in autonomous and semiautonomous systems;
- defines procedures for measuring, testing, and certifying a system's ability to fail safely on a scale from weak to strong, and instructions for improvement in the case of unsatisfactory performance; and
- provides a basis for developers, as well as users and regulators, to design fail-safe mechanisms in a robust, transparent, and accountable manner.

These fail-safe design methodologies or equivalent will need to be applied to autonomous and semi-autonomous train controls systems.

4 AUTONOMOUS AND SEMI-AUTONOMOUS LEVELS OF AUTOMATION

The excepted definitions for metro automation (as described in section 1.1) need to be expanded to cover the characteristics of automatic, semi-autonomous and autonomous operations on open networks. The numerous potential system permutations cannot be aligned and categorised into to the four existing Grades of Automation. It is also not practical to define a definitive GoA for each possibility. There are however, six distinct levels, these are:

- Level A: No automation
- Level B: Automatic (protected manual driving, ATP)
- Level C: Automatic (driver present, AoE);
- Level D: Semi-autonomous (with driver present);
- Level E: Semi-autonomous (remote driver available); and
- Level F: Autonomous (no driver or remote driver available).

The levels are illustrated in Figure 5.

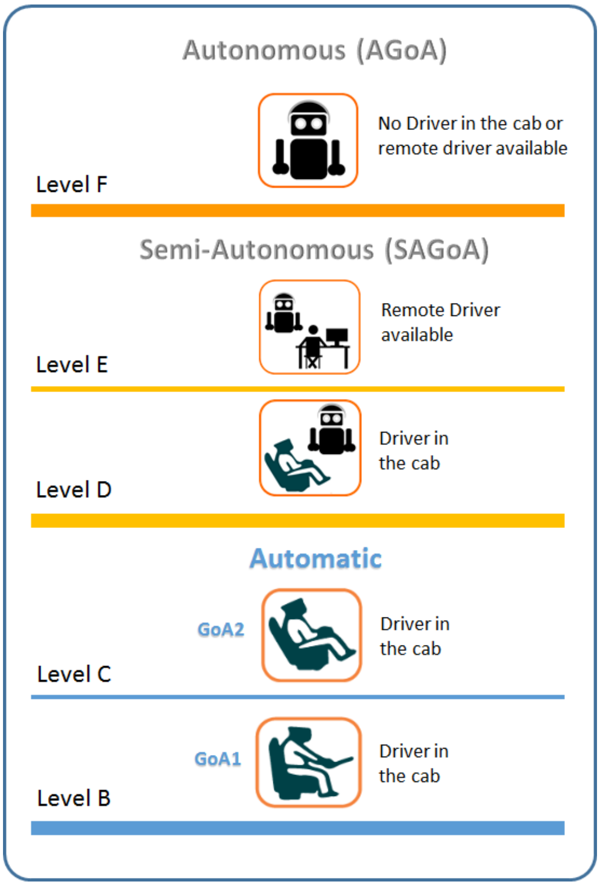


Figure 5 – Levels of Automation for Open Rail Networks (Level A is not shown)

5 CONCLUSIONS

Across all sectors, the world of work is undergoing a radical transformation as automation and AI/ML technologies are adopted promising a productivity and safety revolution. Since the first industrial revolution business have been restlessly seeking opportunities to concede routine physical tasks to machines. Humans have always sought to develop ‘tools’ to make activities easier, safer, more efficient and convenient.

Robots and computers are increasingly capable of accomplishing complex work that requires cognitive abilities as well as simple routine tasks. This approach, can provide consistent outcomes, eliminate human error and poor choices (or at least reduce SFAIRP) and operate at the extreme of system capabilities. Automation will undoubtedly be disruptive across all sectors and the rail sector is no different.

In terms of digital train control, ETCS can provide the foundation and safety critical core for automation at all levels (B to F). AoE can provide the driving function to optimise driving performance. It is expected that this arrangement (Level C) will become the norm for ETCS applications over the coming years. Supplemented, with PSD or PEB in high frequency applications. This investment can be protected and future proofed by building progressive semi-autonomous and autonomous applications on this foundation.

The introduction of each semi-autonomous safety assistance system with a driver present (Level D) will provide an incremental safety improvement. Many of these systems are already being deployed as supplementary system to more basic manually driven train control technologies (for example, obstacle detection). It is also expected that semi-autonomous applications without a driver, using a remote driver for specify tasks (Level E), will start to be deployed in the mid-2020s. By the mid-2030s, fully autonomous in-service train operations (Level F) are likely to become common place.

In a role reversal, it is likely that technology from open network automation will also be adopted in metro (segregated) applications. These systems could provide Safety Assistance systems that further mitigate the residual risks associated with potential intrusions and infrastructure defects. There is also potential to improve metro performance using AI/ML and to provide additional safeguards and mitigate software failures in pre-programmed software.

Railway roles will continue to change and rail organisations will need to carefully manage a people centric transition. What history has taught us is that people can readily adapt to the evolution of the jobs market when this is supported by proactive interventions that can empower individuals to make the labour market transitions and skills changes they need. In addition to the obvious evolution of the role of the train driver. The deployment of autonomous trains will introduce new highly-skilled job roles, for example, new operational roles such as, remote-control drivers and supervisors for automatic trains, as well as technical and engineering roles to maintain and manage the digital systems. This is in addition to the management, technical and engineering roles and opportunities for the development and implementation of autonomous rail systems.

We know from history that change is inevitable and the march of technology and innovation causes relentless disruption. The received wisdom is that it is more productive to embrace technology and to use it to advantage rather than resist it.

Logically, from a SFAIRP argument, within five to ten years, as it becomes practicable, versus state of art, Safety Assistance systems using autonomous technologies are likely to become expected, and possibly mandated through regulatory/legislative imperative. Due to the competitive advantages, the race to bring this technology to market is on. This situation requires next generation regulatory strategy and oversight to ensure proper development. Key to all of this will be the assurance that potential CEV errors (ethics) of the AI elements of the systems are acceptable and can react safely to all real-time events that may be encountered. Other challenges will be related to gaining the benefits of ensemble and federated learning in a competitive commercial environment.

Additional professional training will be required in the development and separately in the operation of AI risk management for trains. Professionals will need to be trained in the fundamentals and principles for managing AI risks. The countries that initiate these education advances will likely economically benefit compared to other nations.

There is a need to update ISO31000 and EN50126 and its companion documents to match the more socio-technological interactive applications of AI, or at least to develop a guidance handbook, much like what was developed for climate change. More needs to be done about this, like the IEC is now doing.

Specific tests have to be developed for the AI to demonstrate the level of competence to operate systems and/or provide critical information for human operators upon which to act. The next evolution of safety cases with AI HF and Stakeholder Management Plans are required, along with appropriate regulatory support.

National and international tolerability criteria need to be developed for AI in use, just like any other control equipment. Would it be possible to consider well developed and tested AI with functional safety integrity levels? These aspects need to be wisely considered over the next few years. Some specific societal engagement strategies may be required for AI operating trains, likely requiring unique approaches due to the diversity and capability of application.

REFERENCES

[1] NTC, Safety Assurance for Automated Driving Systems: Decision Regulation Impact Statement (RIS) [Internet]. 2018 November. Available from: https://www.ntc.gov.au/Media/Reports/(A7B4A10F-22A5-2832-1A11-6E23E1BB7762).pdf

[2] CENELEC EN 50126, Railway Applications – The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS)

[3] CENELEC EN 51028, Railway Applications – Communications, signalling and processing systems: Software for railway control and protection systems

[4] CENELEC EN 50129, Railway Applications – Communications, signalling and processing systems: Safety Related Electronic Systems for Signalling

[5] IEC Blog, New International Standard will offer risk management framework for AI, 2019 March. Available from: https://blog.iec.ch/2019/03/new-international-standard-will-offer-risk-management-framework-for-ai/

[6] IEC 62290-1, Railway applications – Urban guided transport management and command/control systems – Part 1: System principles and fundamental concepts

[7] UITP, World Report on Metro Automation. 2018 May.

[8] Wang Y, Zhang M, Ma J, Zhou X, Survey on Driverless Train Operation for Urban Rail Transit Systems. 2016 December

[9] Karvonen H, Aaltonen I, Mikael Wahlström M, Salo L, Savioja P, Norros L, Hidden roles of the train driver: A challenge for metro automation. 2011

[10] RSSB, Rule Book Train Driver Manual. 2015 December

[11] McKay M, Fatal Consequences: Obstructive Sleep Apnea in a Train Engineer. 2015 December.

[12] NTSB, National Transportation Safety Board, Railroad Accident Brief, Long Island Rail Road Passenger Train Strikes Platform in Atlantic Terminal. 2018 February

[13] NTSB, National Transportation Safety Board, Railroad Accident Brief: New Jersey Transit Train Strikes Wall in Hoboken Terminal. 2018 February

[14] Subset-023, ETCS Glossary of Terms and Abbreviations

[15] Subset-026, ETCS, System Requirements Specification

[16] Draft Subset-125, AoE, System Requirements Specification

[17] Draft Subset-126, AoE, FFFIS ATO trackside to ATO on-board

[18] Draft Subset-130, AoE, FIS ATO on-board to ETCS on-board

[19] Draft Subset-131, AoE, FIS ATO to TMS

[20] Draft Subset-132, AoE, ATO trackside to adjacent ATO trackside

[21] Draft Subset-139, AoE, FIS ATO on-board to Rolling stock systems

[22] Shift2Rail Annual Activity Report 2017. 2018 June

[23] OJEU, Commission Regulation (EU) 2016/919, on the technical specification for interoperability relating to the ‘control-command and signalling’ subsystems of the rail system in the European Union. 2016 May

[24] OJEU, Directive (EU) 2016/797 of the European Parliament and of the Council on the interoperability of the rail system within the European Union. 2016 May

[25] Milburn D, Digital Train Control and the Fourth Industrial Revolution, AusRail 2019

[26] Goebel, Randy; Poole, David L.; Mackworth, Alan K. (1997). Computational intelligence: A logical approach, Oxford University Press. p. 1. ISBN 978-0-19-510270-3. Archived (PDF) from the original on 7 March 2008.

[27] Erskine M, Artificial Intelligence and Autonomous Vehicles, Emerging Needs for Human Factors and Stakeholder Engagement, WEC 2019

[28] LessWrongWiki, Coherent Extrapolated Volition, https://wiki.lesswrong.com/wiki/Coherent_Extrapolated_Volition

[29] Pettenqill J., Segal's Law, 16S rRNA gene sequencing, and the perils of foodborne pathogen detection within the American Gut Project, NCBI, https://www.ncbi.nlm.nih.gov/pubmed/28652935

[30] Johnson, C.W., The Increasing Risks of Risk Assessment: On the Rise of Artificial Intelligence and Non-Determinism in Safety - Critical Systems, 2018. School of Computing Science, University of Glasgow. http://www.dcs.gla.ac.uk/~johnson/papers/SCSC_18.pdf

[31] Muehlhauser, L., Transparency in Safety-Critical Systems, Machine Intelligence Research Institute. August 2013. https://intelligence.org/2013/08/25/transparency-in-safety-critical-systems/

[32] Call for Contributions, First International Workshop on Artificial Intelligence Safety Engineering (WAISE 2018), http://www.es.mdh.se/safecomp2018/workshops/WAISE-CfP_SafeComp2018.pdf

[33] Kanioura Dr, A., Critical Mass: Managing AI's unstoppable progress, Sep 2018, https://www.accenture.com/us-en/insights/digital/critical-mass-managing-ai-unstoppable-progress

[34] Rodriguez, J., 6 Types of Artificial Intelligence Environments, Medium.com, Jan 2017, https://medium.com/@jrodthoughts/6-types-of-artificial-intelligence-environments-825e3c47d998

[35] Morris, C., Tesla's massive accumulation of autopilot miles, Inside EV's, Jul 2018, https://insideevs.com/tesla-autopilot-miles/

[36] IEC 61508-3:2010, Page 48, Table A.2 Software Design and Development – Software architecture design.

[37] EPRS | European Parliamentary Research Service, Artificial intelligence in transport, Current and future developments, opportunities and challenges. PE 635.609. March 2019. http://www.europarl.europa.eu/RegData/etudes/BRIE/2019/635609/EPRS_BRI(2019)635609_EN.pdf

[38] Simmons A, Furness N, The main line ATO journey, IRSE News, Issue 251, January 2019.

[39] Railway Gazette, SNCF targets autonomous trains in five years, September 2018. https://www.railwaygazette.com/news/traction-rolling-stock/single-view/view/sncf-targets-autonomous-trains-in-five-years.html

[40] Smart Rail World, DB tells staff and unions to prepare for driverless operations by 2021, June 2016.

[41] Rio Tinto, How did the world's biggest robot end up here? https://www.riotinto.com/ourcommitment/spotlight-18130_25692.aspx

[42] Draft Development - IEEE P7009, Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems. http://sites.ieee.org/sagroups-7009/

[43] Draft Development - IEEE P3652.1, Guide for Architectural Framework and Application of Federated Machine Learning. https://sagroups.ieee.org/3652-1/