Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P9239-cd

From Nobel Prize(s) to Safety Risk Management: Lessons learnt from 2018 Uber collision for their application to autonomous train systems

Sajeev Kumar Appicharla

Member, IET, Member INCOSE,UK. E-mail: appicharlak@yahoo.co.uk

Lessons learnt from the Case Study Analysis of the 2018 Uber Automated Vehicle Collision are presented in the paper. The System for Investigation of Railway Interfaces (SIRI) Cybernetic Risk Model is used for modeling and analysis of the NTSB Report NTSB/HAR-19/03. Elements of effective safety risk management system compose the SIRI risk model.

The research and development of autonomous driving systems in the automotive sector is driving the trend towards autonomous train systems as well. However, railway safety risk management has been facing a crisis since the promulgation of the 2004 EU Safety Directive. This directive demands a cultural shift away from current engineering safety management practices despite the good statistical record of railways as a safe transport mode. Using techniques such as Failure Mode, Effects, (and Criticality) Analysis and Bowtie analysis do not support identification of systematic errors at the higher levels of socio-technical system that is involved in monitoring and certifying the autonomous train systems. Further, the traditional risk assessment process used in the rail sector does not address decision mistakes noted in the decision making under uncertainty literature. Past research on accident case studies indicates that organizational and management factors that contribute to fallible decisions are not included in the risk assessment process. Thus, the hypothesis of underestimation of significant hazards and risks such as autonomous train system collision hazard and over-estimation of failure of human safety supervisor of autonomous train system is proposed.

One benefit of the paper is to contribute to reflection on the part of systems engineers to help them plan, design, develop and operate safe autonomous train systems and related signaling systems as well.

Keywords: Complex socio-technical systems, Decision Making under Uncertainty, Heuristics & Biases, Safety Risk management, Safety Culture, Autonomous Train Systems.

1. Summary Description of the collision

"On March 18, 2018, at 9:58 p.m., an automated test vehicle, based on a modified 2017 Volvo XC90 sport utility vehicle (SUV), struck a female pedestrian walking across the northbound lanes of N. Mill Avenue in Tempe, Arizona. The SUV was operated by the Advanced Technologies Group of Uber Technologies, Inc., which had modified the vehicle with a proprietary developmental automated driving system (ADS). A female operator occupied the driver's seat of the SUV, which was being controlled by the ADS. The road was dry and was illuminated by street lighting (NTSB 2020) (Crash summary).

The probable cause of the crash in Tempe, Arizona, was the failure of the vehicle operator to monitor the driving environment and the operation of the automated driving system because she was visually distracted throughout the trip by her personal cell phone. Contributing to the crash were the Uber Advanced Technologies Group's (1) inadequate safety risk assessment procedures, (2) ineffective oversight of vehicle operators, and (3) lack of adequate mechanisms for addressing operators' automation complacency—all a consequence of its inadequate safety culture (ibid) (Probable Cause)(section 3.2).

Further factors contributing to the crash were (1) the impaired pedestrian's crossing of N. Mill Avenue outside a crosswalk, and (2) the Arizona Department of Transportation's insufficient oversight of automated vehicle testing(ibid) (Probable Cause)(section 3.2). Further, to the above probable cause, and contributory factors, the NTSB investigation identified the following two safety issues: Uber ATG's inadequate safety culture and the need for safety risk management requirements for testing automated vehicles on public roads(ibid)

(Probable Cause)(3.2). As a result of its investigation, the National Transportation Safety Board made recommendations to the National Highway Traffic Safety Administration, the state of Arizona, the American Association of Motor Vehicle Administrators, and the Uber Technologies, Inc., ATG." (ibid) (section 4). In summary, the NTSB concluded 19 findings(ibid)(3.1) (pp.58). Factors such as (1) driver licensing, experience, or knowledge of the ADS operation; (2) vehicle operator substance impairment or fatigue; or (3) mechanical condition of the vehicle did not contribute to the accident(ibid) (Executive Summary). It is to be noted that Recommendations made by NTSB (see clause 4) used in an implicit manner the 1997 socio-technical system (STS) Risk Management Framework (RMF)(Rasmussen.1997).

A significant contribution of this paper is to identify high level latent failure conditions in the form of lessons learnt. For the Artificial Intelligence (AI) systems, traditional System Engineering SE concepts of system definition, verification, validation, tests need to be modified(Rand Report, 2018) (Box 3.1). Since the AI systems are defined as an engineered systems that do not understand; they need human design choices, engineering, and oversight as per(IEC 22989, 2022) (see section 5), (Appicharla, 2024a). The developments in the AI software in terms of ML OPS for assurance purpose need to be accounted as well (Zeller et al 2023). Due to length constraint, this paper refers readers to the author's previous papers for details of several concepts and their application.

The paper is organized thus. Section 2 presents the methodology. Section 3 presents the lessons learnt. The section 4 briefly discusses the rail AI literature and examines it from the perspective of section 3 lessons.

2. The Accident Analysis Methodology

The System for Investigation of Railway Interfaces (SIRI) Cybernetic Risk Model(2017) is used as a SE methodology for accident analysis and proactive risk assessment of complex socio-technical systems. In 2005, the methodology was developed at Rail Safety Standards Board(RSSB),London against an internal research brief. The Model uses the hybrid Swiss Cheese Model (SCM) advanced by (Reason, 1990b) and Management Oversight & Risk Tree (MORT) manual, chart, and terminology promoted by (Kingston, et al 2009a) using the lens of 1997 STS-RMF. This way it is feasible e to integrate three perspectives of a firm, namely human, organizational, and technological (HOT) factors (with their interactions and their error models) from Systems Thinking discipline perspective (Mitroff and Linstone, 1993), (Rasmussen and Svedung, 2000) (section 7), (Appicharla,2023a). In academia as well as in business it is accepted that the main vulnerabilities in industrial safety come from human and organisational factors. Despite this acceptance, the HOT perspective is not, in general, integrated into system safety as part of the SE (Appicharla 2006a) cited in (Appicharla, 2023a). The model is described in greater detail in (Appicharla, 2024a, b)(see Figure 1).

The use of multiple methods, models and techniques enables accident analyst(s)/risk assessor(s) to overcome complexity for risk assessment purposes, represent all stakeholders and their contribution (in the form of decisions) to safety risk performance and see the problem of fallacies and biases using a fault tree representation as developed by (Kingston, et al 2009a) such that it can be flexibly scaled to cover the whole STS as noted by (Rasmussen.1997)(Figure 1) in a reliable manner. Benefits of accident analysis were described together with biases that may attend such analysis in (Appicharla,2024a).

SE Handbook states that cognitive biases are mental errors in judgment under uncertainty caused by our simplified information processing strategies (sometimes called heuristics) and are consistent and predictable as per (Kahneman, 2012), (INCOSE 2023) (1.4.2). Further, the Handbook states that ccognitive biases can contribute to incidents, failures, or disasters as a result of distorted decision making and can lead to undesirable outcomes(ibid). The 1997 RMF cannot be used for performing safety risk assessment as it is because the framework does not support with hazard analysis in a detailed manner(Appicharla, 2024a). (Appicharla, 2006a) cited in (Appicharla, 2024a) may be consulted for the details on know how to conduct pro-active risk assessments in complex STSs.

The qualitative control system approach to risk management by (Rasmussen and Svedung, 2000) enables us to extend it to include the Nobel Prize winning work on judgment and decision-

making to draw attention to inadequate safety risk rationality. The role of judgemental heuristics and biases (H&B) in generating systematic errors is explained by (Kahneman,2012)(pp109-448). The deviations from rational decision-making criteria according to Subjective Expected Utility Theory are considered as disturbances in control systems (Reason, 1990b). Nobel laureate H.A. Simon's concept that organisation's decisions are satisficing rather than maximising and are based on 'bounded rationality' is considered as contributing to inadequate decision making as well(Rasmussen,1997). The control system representation may be seen in (Appicharla, 2024a,b)(see Figure 1). The author's perspective is that the multi-method model explains how unsafe outcomes result from less than adequate (LTA) interaction between elements of effective safety risk management system. These elements are: first, specification of requirements through compliance with (SE) and related standards and statutory obligations; second, business policy with its integration of risk related policies; third, safety risk management (policy and its implementation) in the form of "as low as reasonably practicable"(ALARP) decision-making; and fourth, how use of heuristics induce biases in decision making on risk management at various levels of a socio-technical system including learning from accidents to control the system of interest (SOI). (Appicharla, 2024a,b) explains graphically, how these elements are integrated dynamically. Further, heuristics of representativeness, availability, and anchoring & adjustment and biases they induce in judgements and decision-making on risk assessments are described as well(Appicharla, 2024a,b).

"Systems Thinking" is a philosophy currently prevalent within the SE discipline that is applied to understand and improve performance and safety in complex STSs (Stanton, et al 2019). Seven tenets of accident causation that this philosophy embodies are consideration of multiple causal/contributory factors, emergent properties arising from interactions, communication and feedback, consideration of all stakeholders (including social actors) and their contribution to safety risk, migration to unsafe zones (the safety drift model), vertical flow of information and a combination of triggering events/conditions identified from the study of road traffic collisions by (Stanton, et al 2019). H&B are added as an eighth tenet.

3. Lessons learnt from 2018 Uber collision

The graphical Concept of Operations (ConOps) is a part of System Definition and is developed in an analogous manner to the (Sequential Timed Event Plotting (STEP) analysis or the Events Causal Factors Analysis (Stanton, et al 2019),(Appicharla, 2024b). System definition requires a deep understanding of the domain interactions, that must be mapped to the new AI operational context (Johnson and Fang 2019). (INCOSE 2023) and (Appicharla, 2024a, b) may be consulted for more details.

Symbolically, the accident scenario(s) of the popular Swiss Cheese Model representation or the Energy Barrier Trace Analysis (ETBA) in the MORT Procedure or the 1997(STS)-(RMF) can be formulated, thus:

$$\forall\ System\_hazard(SB1) \cap loss\ of\ lives(SB2) \cap LTA\ Barriers\ \&\ Controls\ (SB3) \cong SA1\ Accident(s) \quad (1)$$

Where SB1 is potentially harmful energy flow or environmental Condition; SB2. vulnerable People or Objects; SB3 -LTA barriers and controls and SA1 is the Accident (potential or real) as per the terms for the Procedure for MORT Analysis(Appicharla, 2024a, b). The above equations can be used to relate safety property to Trustworthiness. The above (Eq. (1)) can help formulate a Bayesian risk assessment with appropriate data input is to be noted(Kahneman, 2012)(pp. 166). Other equations relating to the dynamic driving task are stated in (Appicharla, 2024a, b).

Applying the flowchart of the MORT Procedure for barrier analysis as per (Kingston, et al 2009a)(section 3.2), the Cybernetic risk model as per (Appicharla, 2024a, b). and keeping in view the related MORT Fault tree Chart(2009), we obtain the following results from (NTSB 2020):

Lesson 1: The MORT Code SB1: potentially harmful energy flow or environmental Condition branch LTA: LTA judgment and decision making (JDM) by AV /ADS designers:

With hindsight bias, it is argued the NTSB failed to establish the failure of the Deep Neural Network (DNN) as a contributory cause. The AV /ADS designers foresaw an inadequate Object and Event Detection and Response (OEDR) function and relied on safety operator to mitigate the inadequacy by designing the ADS "action suppression" function not to apply emergency braking and did not alert the safety operator

either. The system design did not include consideration for jaywalking pedestrians (NTSB) (1.5.6.1). When the SUV was operated in autonomous mode (controlled by the ATG's ADS), all the Volvo ADAS components were automatically deactivated(1.6.3).The NTSB concludes that the Uber ATG did not adequately manage the anticipated safety risk of its ADS's functional limitations, including the system's inability to correctly classify and predict the path of the pedestrian crossing the road midblock(ibid) (2.2.1.1). (Reason, 1990b) drew attention to the three ironies of automation highlighted by Bainbridge. One of them is where operator is left to undertake tasks that designer felt that cannot be automated. Lack of awareness of safety risk management processes on the part of ADS designers led them to deploy an inadequate ADS. Thus, it is concluded LTA safety assurance system led to potentially harmful energy flow is concluded due to LTA system configuration.

Omission bias of ironies of automation due to availability heuristic on the part of the AV /ADS designers is the lesson learnt(Kahneman,2012) (Appendix A), (Appicharla, 2024b).

Lesson 2: MORT Code SB2. vulnerable People or Objects & SA.2 Stabilization and Restoration. a1. Non-functional Energy' b3. Control of exposure LTA. LTA JDM by HF experts/ SE experts:

The NTSB concludes that the pedestrian's unsafe behavior in crossing the street in front of the approaching vehicle at night and at a location without a crosswalk violated Arizona statutes and was possibly due to diminished perception and judgment resulting from drug use(2.1.2). The pedestrian used the median that was an X-shaped, red-brick configuration, which at the time of the crash had the appearance of a pathway (ibid) (Figure 1). When interviewed by NTSB, ATG vehicle operators reported occasionally encountering pedestrians crossing a road midblock, and ATG's training of vehicle operators included preparation for hazardous situations such as jaywalking pedestrians (2.1.2.1). Further, underestimation of risk of AV collision is due to non-simulation of safety operator -ADS interaction(Macrae, 2022). Automation complacency led to vigilance failure on the part of the safety operator, and ATG's LTA oversight of vehicle operators (including removal of second operator) contributed as well(2.1.2.1; 2.2.2.2).

Overestimation of failure of safety operator /supervisor due to oversimplification of causality (Reason, 1990b) and underestimation of the probability of disjunctive event of LTA ADS function or Confirmation bias of the safety operator error due to anchoring and adjustment heuristic by Human Factors experts/ SE engineers is the lesson learnt (Kahneman, 2012), (Appicharla, 2024b).

Lesson 3: MORT Code SB3 -LTA barriers and controls SC1-5: control of work and process -LTA Safety Standards LTA: a1. Technical Information LTA branch: d6. Research LTA? LTA JDM making by SE engineers:

(Stanton, et al 2019) stated that lack of international and national standards for automation design and testing meant that Uber had no technical guidance for appropriate interfaces, safety standards, or testing regimes.

The NHTSA guidance 2.0 listed 12 safety-related areas but contained little specific information on how to achieve those safety goals-for example, training vehicle operators, ensuring oversight, or evaluating whether an ADS has reached a level of safe functionality. Moreover, submitting a safety self-assessment report is voluntary, and NHTSA does not publish an evaluation of the reports to determine the extent to which developers follow the automated vehicle guidance(NTSB 2020)(2.3.2.1).

(Koopman, 2023) stated that the safety standard UL 4600 fills the gaps in the ISO 26262 and ISO 21448 pair of safety standards by providing a comprehensive umbrella standard for AV system-level safety. Software and SE Processes must be both defined and followed to ensure the sufficient quality of not only the software but also engineering analysis and other work products as per the safety standard UL 4600. This output of such definitional activity is called the Systems Engineering Master Plan (SEMP) (INCOSE 2023) (2.3.4.1). As noted earlier, (INCOSE 2023) calls for mitigation of cognitive biases and (Koopman, 2023) calls for avoiding training data bias and ensuring sufficiency of data samples for rare events. At the time of the crash, ATG did not have a corporate safety division or a dedicated safety manager responsible solely for assessing the risk of testing the ADS on public roads (NTSB 2020)(1.8.2),(Macrae, 2022). (NTSB 2019a) provides an example of how SE

process can contribute to aviation accidents despite the Federal Aviation Administration (FAA) safety management system (SMS). Therefore, resident pathogens in SE process need to be mitigated (Appicharla,2023a).

LTA Object and Event Detection and Response (OEDR) and related documentation is a contributing factor to the risk. In the precrash scenario, there was no SMS considered by the ATG (NTSB 2020)(2.2.3).The NTSB did not investigate any ML algorithms for wrong classification of objects detected and did not state if the failure to detect an obstacle was a result of omissions in the training data. Nor did the NTSB request the ATG to provide such information. In control system terms, whether the error in the classification output due to the feedback or feed-forward operation in relation to deep learning or the data fusion filter (such as Kalman filter) or why classification flickered is not stated(Appicharla, 2024b). The Uber AV accident emerged from structurally interlinked failures that interacted across different parts and at different scales of the system, encompassing the design of the ADS, the role of vehicle operators, the decisions of engineers and managers,the processes of vehicle testing, and the actions of other road users and regulators(Macrae, 2022). Capability and commitment to cognise potential AI hazards by safety engineers, human factors, and systems engineers requires improvement (Rasmussen.1997)(Appicharla 2024 a,b).

Confirmation bias and underestimation of risk of technology due to lack of definition of software and SE processes, SMS and safety standards due to anchoring and adjustment heuristic is the lesson learnt (Kahneman, 2012).

Lesson 4 -do- MA3. Risk Management System LTA; MB3. Risk Analysis Process LTA; a1. Concepts and Requirements LTA: b3. Risk Analysis Criteria LTA: c5. Other Analytical Methods LTA:

The methods of Failure Mode, Effects, (and Criticality) (FMEAC) Analysis and Bow-tie analysis do not meet the criteria of tenets of Systems Thinking. Other AI accident analysis methods of STAMP,FRAM, Fault tree and others were assessed against three criteria by HF experts(Stanton, et al 2019). (Appicharla,2024a) discussed other AI accident analysis methods as well. AI systems should be robust, secure and safe throughout their entire life cycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk(IEC 22989)(Appicharla, 2024a). Selection of right methodology is an imperative is the lesson learnt such that safety risk in the AI system can be identifed, analysed and adquate control action is taken.

Lessons 5 -do- b5. Use of Previous Accident/Incident Information LTA? LTA JDM in the accident analysis process :

Uber conducted an internal safety review and commissioned an external safety review from a team led by a former leader of the NTSB. The findings and recommendations of these reviews were published alongside Uber's annual public safety report (Macrae, 2022). However, in terms of organization learning, lessons learnt from other accident reports were not incorporated in safety assurance process, a vital component of SMS(Appicharla, 2024b). The NTSB has examined automation complacency in the operation of vehicles with Level 2 automation capabilities—vehicles that can maintain control and respond to slowing traffic but require constant driver monitoring due to their limited capabilities. The NTSB determined that the probable cause of crashes that occurred in Florida, in May 2016 and in California, in January 2018 included driver inattention and overreliance on vehicle automation (NTSB 2020) (2.2.2.1).

Confirmation bias of human error hypothesis and neglect of counter evidence of under estimation of ADS induced error risk due to anchoring and adjustment heuristic is the lesson (Kahneman, 2012).

Lessons 6: MORT Code: MA1. Policy LTA: MA3 Risk Management System LTA . MB3. Risk Analysis Process LTA : Are the individual decision-makers (staff, management, and regulators) properly informed about the system status in terms comparable to the objectives?

A comprehension gap can develop between apparently safe performance and less visible indicators of growing risk vulnerability such that organizational leaders proceed under a false sense of security. Lengthy periods in which bad events are absent create the (false) impression that the system, the organization, and its technologies are operating safely. Therefore, it is imperative that evaluation of current work practice with respect to the safety objectives and adequacy of control structures and functions to manage these

objectives is conducted (Rasmussen and Svedung 2000), (Starbuck, and Farjoun, 2005)(pp.63).

Optimism bias and strategic misrepresentation bias can induce overconfidence and comprehension gap with respect to quality of decision process in managing safety and performance is the lesson learnt (Kahneman, 2012).

Lesson 7: MORT Code: MA3. Risk Management System LTA; LTA Understanding of Safety Drift model:

The complex STS may drift into failure zone by eroding the defenses in the face of production pressures, and this is a well-documented pattern (Starbuck and Farjoun 2005), (Macrae, 2022). (Hendrycks, et al 2021) argue that competitive dynamics surrounding machine learning-ML's development may pressure companies and regulators to take shortcuts on safety. The economic imperative to compete on costs with Airbus resulted in LTA safety culture perspective, and organisation dynamics drove the decision(s) towards setting up of the latent failure pathway to Boeing 737 Max 8 crahses(Appicharla, 2023a).

Thus, systems engineers need to improve their culture to upgrade their traditional SE tasks to make safety a super-ordinate goal is the lesson learnt(Reason, 1990b).

Lessons 8: MB3. Risk Analysis Process LTA: a2. Design and Development LTA: b9. Human Factors (Ergonomics) Review LTA: c28. Did not Establish Human Task Requirements: d4. Design of Controls LTA: LTA Interaction between elements for safety risk management, omission of HOT Factors:

In addition to the companies developing AVs, there are many other actors actively seeking to understand and attend to AV safety (Rand Report, 2018). The SEMP needs to include processes to learn requirements from all these stakeholders. Lack of a SEMP led to inadequate, mono-causal explanation relying upon perceived similarity between single cause and its effect due to representativeness heuristic (Reason, 1990b)(pp. 91), techno-solutionism bias (NIST 2022a), data inadequacy regarding human error (94%) being a cause (Johnson and Fang 2019), omission of HOT Factors and LTA interactions between design team and other teams at the Uber ATG is the lesson learnt.

Lessons 9: -do- a1. Concepts and Requirements LTA/ a2. Design and Development LTA: (Kingston et al, 2009). LTA Safety Culture:

(NTSB 2020), (footnote 63) defines the safety culture. Safety cases called structured arguments with falsifiable claims supported by evidence, which identify potential hazards, describe mitigations, show that systems will not cross certain red lines, and model possible outcomes to assess risk are one of the possible governance mechanisms as per Nobel laureates Kahneman and Hinton and other AI experts as well (Bengio, et al 2023). Encoding human goals and intent like safety culture is challenging(Hendrycks, et al 2021). However, logical fallacies (e.g. planning fallacy, defence in depth fallacy, McNamara fallacy to name a few) and biases (e.g. confirmation bias, Omission bias, out of sight put of mind bias, to name a few) that may undermine the arguments of the safety case and these need to be considered(Starbuck, and Farjoun, 2005).

Three components of culture, namely, Cognition, Commitment and Competence to deal with potential hazards and integrate them together in a matrix form with Principles, Policies , Procedures, and Practices to help improve the safety management systems is another imperative (Appicharla. 2024a). The four primary components of an SMS—as advocated by the FAA and adopted industrywide, including by the ground transportation are not sufficient for manging AV safety. Because the rapid evolution and introduction of Artificial Intelligence (AI) and Machine Learning (ML) into SE further increases complexity of verifiability, safety, and trust of self-learning and evolving systems(INCOSE, 2023) (1.2). Therefore, the SMS needs to consider bias risks in its safety data in addition to the requirements noted in the PAS 1881 safety standard and other challenges noted in (Appicharla, 2024a,b) are to be addressed as well. (Appicharla, 2024a,b) presented optimism bias in the case of 1954 Nobel laureate Max Born. The choice experiment in economics of risk by future Nobel laureate M. Allias revealed bias in the form of certainty effect on the part of three future Nobel laureates and an expert statistician as well violating the norms of economic rationality (Kahneman, 2012)(pp.313). Thus, the hypothesis of insensitivity to variations of risks among small probabilities is advanced. Therefore, all AI stakeholders need to improve to their JDM process to address LTA safety risk rationality is

the lesson learnt (NIST. 2022a), (Kahneman, 2012).

4. Discussions regarding rail sector

Scrutiny of rail sector AI documents by the UIC, EU Rail research and the IEC standard 22989 revealed the differences in terms between the rail domain and the AI discipline (Appicharla,2024c). Thus, the hypothesis that conceptual clarity is lacking on the part of rail sector is established.

(European Union Agency for Railways 2024), the EU wide SE authority, admits that the EU railway sector maintains a culture of silo management, leading to fragmentation: with each technical discipline having its own approach to common issues(e.g. acceptable levels of risk); each country/national railway company/infrastructure manager defining its own system architecture; and the emerging importance of data and digitalisation has the potential to also create new barriers to the Single European Railway Area. Thus, the hypothesis of lack of SE process and its definition as per the UL 4600 standards is established.

The European railway legislative framework (European Commission, 2018, 2020)(an amendment of 2004 Safety Directive) introduce explicit requirements to take a systematic approach to supporting human performance and managing human and organisational factors within the SMS (Accou and Carpinelli 2022). The difficulties in managing emergent properties of inter-operability and safety are not addressed by the traditional approaches in the EN 50126-1 SE standard as HOT factors are not considered in the System definition(Appicharla,2024c). (Tonk et al, 2024) consider human and organizational factors but do not state how the modeling and safety analysis of such interactions is carried out.

(Zeller et al, 2023) note the challenge that ADS also in rail will operate in an open world, which is difficult to specify a-priori and is prone to changes during its lifecycle, hence requires agile MLOps cycles including testing & validation in the field. The integration of SE lifecycle, the safety assurance lifecycle, and the data & ML lifecycle is considered but the authors fail to address HOF interactions, risk assessment biases, tailoring of SE processes, learning lessons and "catastrophic forgetting "hazard (IEC 22989) (Appicharla, 2024a,b). (Schnitzer, et al 2024) lists AI hazards but not all safety risk factors identified in section 3 are discussed.

(ASTRail 2019) performed the moving block hazard analysis without a system definition, without incorporating the HOT factors and without using the STS lens. Thus, hazards and their contributing factors remain unknown.

(RSSB 2024) Guidance stresses the importance of the system definition for mandatory risk assessment but does not address HOT, ConOps and other SE concepts such safety drift, and tenets for accident causation listed in this paper. Further, pitfalls of traditional hazard identification methods, e.g. , FMEA and Bowtie to address latent failure conditions in complex STSs is not stated either. Logical fallacies, and biases that may undercut the safety case arguments such as techno-solutionism bias, under and overestimation of safety risks, LTA SEP definition, LTA safety standards, LTA Alarp decision making, LTA System Thinking, LTA JDM processes, and LTA hazards analysis to name a few are not discussed by any of the above. The latent failure conditions revealed in the section 3 are seen in the rail sector as well. Thus, the hypothesis of cultural shift as desired by the legislative framework since 2004 is not seen in the rail sector(Appicharla, 2023a, 2024 a,b,c).

5. Conclusion

The latent failure conditions elicited from the Uber ADS collision were shown to be present in the rail sector as well. It is hoped that system engineers will reflect on the latent failure conditions highlighted and improve their SE processes to address them in the rail and AV sectors as well.

Acknowledgement

Gratitude is expressed towards family members and authors who shared their works freely with the author. Thanks are due to the anonymous reviewers to help improve quality and readability of the paper.

References

Accou, B & Carpinelli, F. 2022.. "Systematically investigating human and organisational factors in complex socio-technical systems by using the "SAfety FRactal ANalysis" method." *Applied ergonomics* 100 (103662):1-9.

Appicharla, S.K. 2024a. Accident Case Study Analysis of Developmental Automated

Driving System Collison. Accessed Jan 9th, 2025. https://bit.ly/3XbAWsz

Appicharla, S.K. 2024b. "Accident Case Study Analysis of 2018 Developmental Automated Driving System Collison." Accessed Jan 9th, 2025. https://bit.ly/3PqzTSc .

Appicharla,S. 2024c. "Less than Adequate Application of Systems Thinking Concepts in the Rail Sector." August 29th. Accessed Septmber 4th, 2024. https://bit.ly/4g8nHBr .

Appicharla.S.2023a. "The Boeing 737 MAX 8 Crashes, System-based Approach to Safety —A Different Perspective.".Accessed February 6th, 2023. https://scsc.uk/journal/index.php/scsj/article/view/18 .

ASTRail. 2019. *D2.2 - Moving Block signalling system Hazard Analysis.* safety related, The ASTRAIL Consortium. Accessed Jan 13th, 2025. https://projects.shift2rail.org/s2r_ip2_n.aspx?p=S2R_ASTRAIL .

Bengio, Y, et al. 2023. *Managing AI Risks in an Era of Rapid Progress.* 1-11. Accessed Feb 22nd 2025. https://arxiv.org/abs/2310.17688

European Union Agency for Railways. 2024. "A compelling vision for 'the target railway system'." Accessed December 12th, 2024. https://bit.ly/3VZdKOz .

Hendrycks, D, Carlini, N., Schulman, J and Steinhardt. J 2021. "*Unsolved* problems in ML safety, Pre-print." Accessed Jan 15th, 2025. https://arxiv.org/abs/2109.13916 .

INCOSE. 2023. *International Council on Systems Engineering Handbook, 5th Edition.* Walden D.D et al(Eds). John Wiley & Sons Ltd. Accessed October 13th, 2023. https://bit.ly/3IgJeYQ .

Johnson, N and Fang, X. 2019. "Three reasons why: Framing the challenges of assuring ai." In *Computer Safety, Reliability, and Security. SAFECOMP 2019. Lecture* Notes *in Computer Science*, by A., Troubitsyna, E., Gashi, I., Schoitsch, E., Bitsch, F. (eds) Romanovsky, 281-287. Accessed February 17th, 2024. https://bit.ly/4ibCCuL

Kahneman, D. 2012. *Thinking Fast and Slow.* London: Penguin Group.

Kingston, J., F.et al. 2009a. "The Management Oversight and Risk Tree User Maual and Chart." December 20th. Accessed May 7th, 2022. https://bit.ly/3vTTWzi .

Koopman, P. (2023, May). Ul 4600: what to include in an autonomous vehicle safety case. *COMPUTER*, 101-104.

Macrae,C.2022. "Learning from the failure of autonomous and intelligent systems: Accidents, safety, and sociotechnical sources of risk." *Risk analysis* 42 (9): 1999-2025.

Mitroff, I. I., & Linstone, H. A. 1993. *The Unbounded Mind: Breaking the Chains of Traditional Business Thinking.* New York: Oxford University Press.

NIST. 2022a. "Towards a standard for identifying and *managing* bias in artificial intelligence." March. Accessed May 31, 2024. https://doi.org/10.6028/NIST.SP.1270 .

NTSB. 2020. "Reissued Report-Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018. NTSB/HAR-19/03." June 26th. Accessed February 18th, 2024. https://bit.ly/3QtNJn6

NTSB. 2019a. CERTIFICATION *SPECIALIST'S REPORT NTSB ID No.: DCA19RA017.*, 35. Accessed August 20th, https://data.ntsb.gov/Docket/?NTSBNumber=DCA19RA017

Rand Report. 2018. *Measuring automated vehicle safety: Forging a* framework..Accessed June 20th, 2024. https://bit.ly/3Qv3wSF

Rasmussen, J., and Svedung, I. 2000. *Proactive Risk Management in a Dynamic Society.* Karlstad,Sweden: Swedish Rescue Services Agency. Accessed February 25th, 2024. https://bit.ly/48wR0bX.

Reason, J. 1990b. *Human Error.* 17th. New York: Cambridge University Press.

RSSB- Rail Safety and Standards Board. 2024. *GE /GN 8646: Guidance on the Common Safety* Method *for Risk Evaluation and Assessment Iss 1.1.* London: Rail Safety and Standards Board.

Schnitzer, R., Hapfelmeier, A., Gaube, S., & Zillner, S. 2024. "AI Hazard Management: A framework for the systematic management of root causes for AI risks." Accessed Jan 6th, 2025. https://arxiv.org/abs/2310.16727 .

Stanton, N. A., Salmon, P.M,Walker, G.H and Stanton. M.2019. "Models and methods for collision analysis: A comparison study based on the Uber collision with a pedestrian." *Safety Science* 120: 117-128.

Starbuck W.H and Farjoun.M.(Eds). 2005. *Organization at the limit: lessons from the columbia disaster.* Blackwell Publishing.

The NRI Foundation. 2014. "ECFA+" June 30th. Accessed March 6th, 2024. https://www.nri.eu.com/NRI4.pdf .

Tonk, A., Chelouati, M., Boussif, A., Beugin, J., & El Koursi, M. 2023. "A safety assurance methodology for autonomous trains." *Transportation research procedia* 3016-3023.

Zeller, M., Waschulzik, T., Schmid, R., and Bahlmann, C. 2023. "Towards a safe MLOps Process for the Continuous Development and Safety Assurance of ML-based Systems in the Railway Domain." 6th July. Accessed Feb 25th, 2025. https://arxiv.org/abs/2307.02867.