
Proceedings Paper:
Ryan Conmy, Philippa Mary orcid.org/0000-0003-1307-5207, Porter, Zoe, Habli, Ibrahim orcid.org/0000-0003-2736-8238 et al. (1 more author) (2023) Safety engineering, role responsibility and lessons from the Uber ATG Tempe Accident. In: First International Symposium on Trustworthy Autonomous Systems (TAS '23). Association for Computing Machinery, Inc .

https://doi.org/10.1145/3597512.3599718

Safety engineering, role responsibility and lessons from the Uber ATG Tempe Accident

Anonymous Author(s)***

ABSTRACT

Safety critical autonomous systems (SCAS) require a safety assurance case (SAC) to justify why they are considered acceptably safe to use, despite the residual risk associated with their operation. Reducing risk is an overarching principle of all safety critical systems development and operation. The SAC should demonstrate that the risk is tolerable and has been reduced as far as possible, through robust design and operational controls. As a SCAS may not have an operator, safety engineers have a more direct responsibility for operational decisions. Following an accident it may be useful to understand which engineering decisions causally contributed to it, and roles responsible for those decisions. This paper contains a review of how different senses of responsibility (role, moral, legal and causal) apply to SCAS engineering and operation. We use this to illustrate how considering role responsibility can help support a defensible SAC, and potentially improve system safety practice. Our findings are illustrated with an analysis the Uber/Tempe Arizona fatal collision accident report. We found that existing safety practice may not identify all role responsibilities in a way that supports causal safety analysis. This paper is intended for the whole TAS community, but with an emphasis on safety professionals.

CCS CONCEPTS

• **Hardware → Safety critical systems**; • **Software and its engineering → Software safety**; • **Computer systems organization → Dependable and fault-tolerant systems and networks**.

KEYWORDS

autonomous systems, safety, responsibility

ACM Reference Format:

Anonymous Author(s). 2023. Safety engineering, role responsibility and lessons from the Uber ATG Tempe Accident. In *Proceedings of First International Symposium on Trustworthy Autonomous Systems (Conference acronym 'XX)*. ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Safety Critical Autonomous Systems (SCAS), such as autonomous vehicles, inspection drones and medical diagnosis systems are being rapidly developed and deployed in the real world. SCAS provide many technical challenges, particularly when they include Machine Learning (ML) components. Additionally, when there is limited or no interaction with an operator or Human in/on The Loop, responsibility for decisions which lead to accidents or hazards are more directly linked to design decisions made by the those responsible [4].

During development of a SCAS, safety engineers will identify issues which lead to design decisions which are intended to reduce the risk associated with the system, either to make it less likely to occur or to mitigate the severity of outcome. They will produce a safety assurance case (SAC) or safety justification which is intended to show the residual risk is acceptable, and, in certain cases, provide some protection against blame for harmful outcomes [22]. Nevertheless, the nature of the system will mean that risks remain, despite due diligence from system developers. Following an accident it may be useful to understand which design decisions causally contributed to it, and who was responsible for those decisions. This is not necessarily to blame or punish those involved, but to prevent mistakes happening again.

There are a number of different senses of responsibility, including role, moral, legal and causal [9]. We argue that understanding these in a safety system engineering context can help to improve the construction of a robust safety case which explicitly considers different roles. We show how an existing safety analysis method (bow-tie) would not have been sufficient for predictive safety role and risk analysis when compared to the findings of the fatal Uber ATG crash in Tempe, Arizona [16], as a number of roles and causal factors are not specified. Based on our findings we have developed a SAC structure which incorporates roles explicitly, and forms a basis for future research on causal safety analysis including roles.

In section 2 we describe different senses of responsibility and their relationship with safety assurance approaches. Then we present a framework in section 3, highlighting some key responsibility questions and gaps. Section 4 contains our case study using the framework. Section 5 contains the outline role SAC structure. Finally, there is a discussion and considerations for future work in section 6.

2 BACKGROUND AND RELATED LITERATURE

In this section we present background information, first providing an overview of typical safety engineering practice relating to risk reduction and engineering tasks. We expand this by considering classical definitions of responsibility and how they might apply.

2.1 Safety engineering practice

Safety-critical systems are defined as those whose failure, under certain conditions, can lead to harm to humans or the environment. There are many different examples, such as medical devices, nuclear power plants, defence systems, cars and aircraft. Safety-critical autonomous systems include autonomous vehicles, inspection drones, and medical diagnosis systems. Every system is subject to differing

Conference acronym 'XX, July 11-12, 2023, Edinburgh, UK

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

legal and regulatory regimes, depending on domain, type, country of use and even degree of autonomy. Rather than attempt to cover all of these in this paper we have summarised some common practices which will broadly apply when developing a SCAS.

The manufacturers and operators of a SCAS will need to provide a SAC or safety justification that the system is *acceptably safe* [22]. The SAC may require external scrutiny, e.g., from a regulator. For the purposes of this paper we define *acceptably safe* to mean that the risks associated with operating the system in a specific environment are both within tolerable bounds, and have been reduced as far as possible or practicable through a combination of design and operational measures. This is consistent with practice for medical device assurance [12] and UK health and safety legislation [1] amongst others. In other words, even though there will still be some residual risk associated with the system, justified measures have been taken to minimise this. It is not enough to just show risk is tolerable, if there are mechanisms to reduce risk further that are practical and proportional to implement, these should also be used. If they are not used, this should be justified.

For example, an aircraft has the potential for a catastrophic crash killing everyone on board, but will have been designed to reduce this risk (e.g., through use of good quality physical components and robust software testing). Further, it should be operated in a way which reduces risk further (e.g., through regular maintenance inspections and use of pilots with up to date training and relevant experience). It is, however, impractical for the aircraft not to leave the ground. A complication for autonomous systems is that there will be limited interaction with a human operator to act as an agent who can reduce risk on the ground. For example, an autonomous drone will make safety-critical decisions itself, rather than have them made by human remote operator. This means design decisions have a more direct impact on the safety of operational decision [4].

At every stage of the development of a SCAS, design decisions will be made which will impact on safety risk [8]. For example, for an autonomous system with Machine Learning (ML) components, decisions will be made on the depth and range of training data used, the acceptable level of ML performance, and verification and test coverage, and even setup of the testing regime [10]. Even with a very robust training regime, latent failures will remain in the system. This is not through lack of engineering diligence, but due to the complexity of the system, making it impossible to fully analyse, and due to inherent uncertainties of ML which typically cannot be completely or formally specified and verified [4]. An additional problem is that it can be hard to determine exactly which element of the ML training has led to a failure. Accidents may also be caused by environmental factors out of an engineer's or operator's control, e.g., bird strike on an aircraft. It is still expected that the engineers would consider these and how to manage them as part of their design process.

We illustrate in Figure 1 how a failure in an ML component can propagate to the system boundary, causing a hazard. Methods will be applied by engineers at each stage of development to reduce the occurrence and severity of consequence. During operation this continues, such as by maintenance to replace broken parts or by braking to reduce speed of impact thus reducing accident severity. Part of the purpose of a SAC is to justify, in advance, that all due diligence was applied at these stages to ensure risks are tolerable and

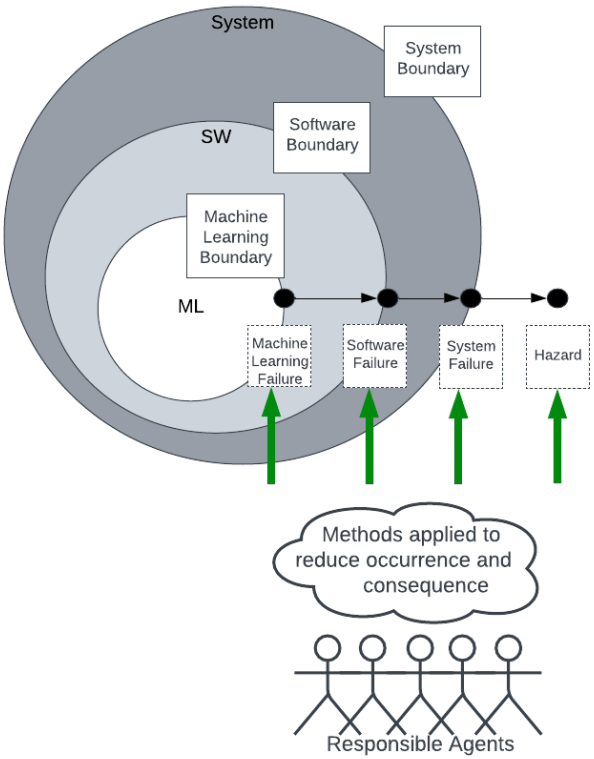

Figure 1: Responsible agents influencing occurrence of failures, adapted from [10]

reduced as far as possible. However, the concept of responsibility is usually implicit rather than explicit. We argue that considering the different senses of responsibility explicitly can help improve the safety process, and support a defensible position for engineers of a SCAS. In the next section we discuss in more depth how different senses of responsibility apply in the SCAS domain.

2.2 Responsibility in safety engineering context

Drawing on Hart's classic taxonomy of the different senses of responsibility [9], we outline five senses of the term "responsibility". These are:

(1) Role Responsibility
(2) Causal Responsibility
(3) Legal responsibility
(4) Moral responsibility
(5) Capacity Responsibility.

Role responsibility refers to the tasks, specific duties, and obligations that attach to particular roles. For example, a software engineer would have duties and tasks to develop a Machine Learning (ML) component to embody safety requirements. An operator working with a SCAS will have duties involving monitoring its performance and intervening if necessary.

Causal responsibility is another way of referring to causation. If one is causally responsible for something, one is just a cause or it or

a salient causal factor in its coming about. Events and artefacts, as much as people, can be causally responsible for things. For example, an ML subsystem may be causally responsible for the dangerous manoeuvre of the SCAS. There are different models which address when a factor can be considered a causal factor, such as reaching a certain threshold, or proportion of contribution, or remoteness from contributions [23].

Legal responsibility typically means legal liability which includes being required to pay financial compensation, or be subject to a legal order, or face punishment [19]. Civil liability concerns actions and practices which could harm others but which are not criminal. There are different types of civil liability, for example vicarious liability where one person is liable for the actions of another, as well as strict liability where the person is liable irrespective of fault. A necessary condition of civil liability is that the defendant has either a duty laid down in legislation (a legal obligation) or a common law duty towards the claimant. Legal obligations are also a kind of legal responsibility. Some of the role responsibilities of people in the SCAS development process may be legal obligations or duties.

Moral responsibility, concerns whether an agent deserves either blame or praise for an outcome. This can overlap with the other senses of responsibility, such as role responsibility. Causal responsibility is generally taken to be necessary but not sufficient for moral responsibility. In other words, although an agent could be causally responsible for an outcome, it is not necessarily fair to blame them for it. As noted, a design decision (or lack of decision) made by an agent could have contributed to an accident involving a SCAS. Consider the example of a large (100,000+) database of road sign images used to train an ML based image classifier for an autonomous car. The curators of the database did not expect or intend it to be used in a safety-critical environment and there are many images which have been incorrectly labelled through simple human error. It was infeasible for the engineers using the training data to perform a complete manual check on a database of that size. An accident is caused, in part, due to the failure to recognise a stop sign from a certain angle which can be causally linked to poorly labelled training data. It is unclear the degree to which the curators could be considered blameworthy, if at all, despite their causal contribution. Instead, we might blame the ML engineers for using an inappropriate source of data.

Related aspects for safety engineering are those of safety culture, including a just culture [7][14][17]. A just culture is one in which incident reporting is encouraged within an organisation, encouraging transparency without fear of blame. Instead the focus should be on reviewing and understanding mistakes and avoiding them in the future. The main aim of the safety culture as a whole is to embed risk awareness within an organisation, encouraging openness, promoting learning and discussion of safety matters while avoiding issues such as complacency, pressure to deliver, and a blame culture. A more nuanced understanding of causal contribution and moral responsibility is planned for future work.

Capacity responsibility relates to a person's mental capacity to be legally liable or morally responsible (for example, relating to someone who might be suffering from "diminished responsibility" in a murder trial). Though it is an interesting philosophical question whether a SCAS could ever have capacity responsibility, it is beyond the scope of our paper to address this question. We assume, in line with the current orthodoxy, that they do not currently have capacity responsibility and cannot themselves therefore be morally responsible.

One related aspect to these five senses is forward-looking responsibility (i.e., responsibility for bringing about or preventing an outcome) versus backward-looking responsibility (i.e., responsibility for an outcome after it has occurred). A SAC is generally used to clarify forward-looking responsibility for ensuring that the SCAS will continue to be safe to operate. Evidence in the SAC can be used to help identify backward-looking responsibility after an incident or accident. It could be used to ascertain whether all actors did what they should have done, to the appropriate level.

Another useful concept for our review is "the problem of many-hands" by Thompson [24][25]. This is summarised as the problem of determining responsibility for decisions made in a complex organisation. When many hands are involved in decision making it can mean that it is difficult, even in principle, to ascribe responsibility for outcomes. In the development of a SCAS there will be multiple developers, engineers, suppliers of components and data, project managers, stakeholders, regulators etc. and it is infeasible to document all decisions they make which contribute to safety.

The problem of many hands, and the difficulties in identifying responsible agents for machine learning in a non-safety environment is explored in [6] by Cooper et al., noting many of the different responsible agents. Cooper notes that the inevitability of ML bugs can be used to excuse responsibility for non-safety systems. We note that for SCAS this isn't the case, as reduction of bug occurrence and impact is required to reduce risk, and should be documented. However, bugs will still remain despite this. Both Thompson and Cooper argue that responsibility should be designed into organisations while systems are developed. Building role responsibility concepts into safety engineering analysis, and justifying them in a safety case, are one way we propose to do this for SCAS.

3 APPLYING RESPONSIBILITY CONCEPTS TO SCAS ENGINEERING AND OPERATION

In this section we consider briefly how the different senses of responsibility relate to each other, specifically when considering safety-critical systems engineering and operation for a SCAS.

The main concept we consider is that of role responsibility. Roles are assumed to include both individuals and organisations. The roles may have specific duties defined. Where there are duties defined, these will include both legally mandated duties (e.g., driver needing a licence) and other duties. Formally specified legal duties may be directly considered when identifying liability. However, many duties although not legally mandated may still reduce risk, such as an engineer following good design practice. Examples of good practice can be found in [12] for medical devices, [21] for avionics, and [11] for automotive. They provide suggested engineering measures to reduce risk, which should be applied proportionally dependent on risk. Hence, when followed, they could be considered to reduce the extent to which an agent in a role could be held accountable after an incident. We describe these as *compliance duties*.

Causal analysis is a means by which causal responsibility and contribution to risk is identified, either backwards-looking following an accident, or forwards-looking to determine potential causes of hazards. As noted, determining level of causal contribution may be complex, especially where a role and duties were tangentially related to the outcome. Additionally, the problem of many hands may complicate this further. However, causal safety analysis of how hazards may arise for a SCAS, including failures of its components and directly interacting agents, is an essential part of any safety-critical development process. There are many different analysis methods which are too numerous to describe here (see section 4 for one example). Their purpose is to consider the many different technical ways in which a hazard could occur, i.e. causes, and how to manage these. Analyses concentrate on technical failures and typically don't consider all of the agents/roles which could be causally responsible for these, particularly from development decisions. For example, a circuit board has physical failures and analysis should review the impact of this, and means to reduce likelihood, such as redundancy. However, the causal contributions from agents behind the physical failure may not be probed in much depth.

An enhanced version of causal safety analysis, using the responsibility concepts in section 2, could help to determine what and who is potentially causally contributing to a hazard, supporting causal contribution and, potentially, a (moral) responsibility analysis if required. For example, a SCAS may be the cause of an undesirable outcome but has no legal personhood so cannot be held morally responsible or liable. An engineer could have been causally responsible for a bug in an ML component, but if they developed it following appropriate guidance it may not be considered fair to hold them accountable. Alternatively, an operator performing their role and duties with due diligence could argue they are not blameworthy for an accident.

In summary, roles relating to the operation and engineering of a SCAS have duties. Performing their duties diligently reduces risk relating to some causes (likelihood and severity) and enables role-holders to show they have reduced risk as much as possible, thereby reducing their exposure to being held morally responsible, even where risk remains of the related causal factor leading to an accident. In the SAC we must provide assurance about these duties, demonstrating that they ensured risk is tolerable and reduced as far as possible. A key aspect of this is showing the risk reduction is appropriate and proportional to the risk contribution. This is represented in Figure 2.

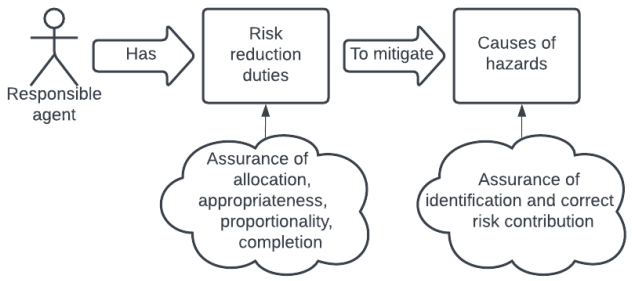

Figure 2: Roles, duties, risk reduction and assurance

3.1 Example roles

Based on this framework we consider three different examples of safety engineering roles which relate to designing and operating a SCAS. This is not an exhaustive list of roles, e.g., regulators, supplier or duty holders may have a role. Each role may have specific duties which impact on whether they are accountable or liable. Remembering that our aim is to support role assurance, we assume that these duties provide means to reduce the risk an adverse outcome, hence prevent unfair judgements following such an outcome. We note duties may intersect or differ across different domains and jurisdictions.

Operational role – this refers to any human in/on the loop who interacts with the SCAS in a way which can directly intervene in, or manage, its functions in-service. Examples are a safety driver in an autonomous car, remote operator of an inspection robot, clinician using an ML based diagnosis system. Depending on the domain, they may have legal or compliance duties to consider. For example, the highway code or medical ethics codes of conduct. For some SCAS this role may not exist. A key assurance issue we do not cover in this paper will be ensuring the SCAS/operator interface and handover of role responsibility is fit for purpose [15]. Another issue is to ensure the duties continue to be performed as required during operation.

Engineering role – this role refers to individuals and organisations who both design and develop the SCAS and provide operational oversight. For example, an ML engineer may produce a classifier, safety engineers may develop hazard analysis, and technicians may monitor in-service performance logs. Again, they may have both legal and compliance duties. For example, there may be legal frameworks such as the health and safety at work act [1] or environmental legislation which must be considered in design and operation. The SAC should demonstrate that these legal duties are fulfilled.

Compliance duties for engineering roles refer to guidance and standards which provide what is known as good practice for developing and operating safety-critical systems. Examples are DO-178C for avionics [21], ISO-14971 [12] for medical devices, or ISO-26262 [11] for automotive. These may not be legally mandated. Their purpose is to recommend engineering methods to reduce the occurrence of undesirable behaviour (for example, using specific software testing methods). Thus, following compliance guidance should alleviate whether an individual could be considered morally blameworthy following an incident. Further, even where there is no legal requirement to follow a particular type of compliance document, not following them could be considered negligent which may impact on liability. A key issue for SCAS is the lack of compliance guidance for ML components.

Third-party role – this refers to people, organisations or other systems which could be causally responsible following an incident or accident but who have no formal duties relating to operating the SCAS. It could also refer to agents who are subject to damage from the SCAS. For example, pedestrians involved in an autonomous car accident could have behaved in an unpredictable way, an offshore wind farm could be damaged by a remote inspection drone, or a third-party service such as telecommunications could have been faulty. We have presented these explicitly, as their role in causal

chain may be important when determining moral accountability and liability. Safety engineers should consider third-party behaviour as a factor when identifying issues and risks.

4 CASE STUDY

In this section we use a case study to examine the core concepts of causal safety analysis with role responsibility. In 2018 there was a fatal collision of the Uber Advanced Technologies Group (ATG) vehicle with an automated driving system (ADS) and a pedestrian pushing a bicycle across a highway in Tempe, Arizona. The National Transportation Safety Board (NTSB) have issued an accident report [16] describing contributory factors, highlighting a number of areas where there were safety shortfalls. Uber ATG have since published a report describing where they have improved their safety processes to address some of the shortfalls identified [3]. A detailed sociotechnical analysis of the accident can be found in [13] which points at many undesirable organisational factors such as undue pressure to deliver and update a working solution, learning lag when reviewing in service data due to too much information, push for technical capability constraints to prioritise normal performance over safety, and over-reliance on the safety drivers reliability over long periods of time monitoring for hazards. These three reports were the main sources of information for our analysis.

Uber ATG were the manufacturer responsible for developing the ADS, which was an adaptation to an SUV from a third-party supplier (Volvo). The ADS used a number of ML components to detect and classify objects in the vehicles path, and also to predict those objects trajectories. The findings of the accident report noted that the classifier failed to recognise the pedestrian with a bicycle, and due to that failure the ADS could not predict their path correctly to take emergency action.

The operator role was a safety driver who was trained to disengage the ADS in emergency situations and take avoidance action. The safety driver was found to be distracted for the accident. The impacted third party role was the pedestrian, who unfortunately lost their life. A number of other regulatory roles are highlighted in the report.

We first show a typical causal safety analysis which illustrates key safety roles and risk reduction mechanisms, as described in the accident report. We then show an alternative view, including all roles from the accident report, and different senses of responsibility as introduced in sections 2 and 3.

4.1 Safety analysis using bow-tie method

In section 3 we described the concept of causal safety analysis. One example of this is bow-tie analysis, which is used in rail [20] and civil aviation [5], as well as other domains. Bow-ties model different *threats* which may cause an undesirable *top-event*, and the various methods or *barriers* which are used to either prevent it, or mitigate the severity of the *outcome*. One notable difference for bow-ties over other safety analysis methods is that role responsibility can be explicitly documented for each barrier, along with other annotations, such as effectiveness or type (procedural or design). This is why we have chosen it for our case study, having made a comparison with other methods such as Fault Tree Analysis and Functional Hazard Analysis. One weakness with the method is there is no formal ordering of the barriers, although this can be useful as it doesn't constrain the safety analyst's model.

Our bow-tie model of this incident is shown in Figure 3. The model is not a specific accident analysis, but is instead a model of all the technical safety measures identified in the accident report [16], and their shortcomings. These measures existed whether the accident had occurred or not. The top-event (circle in the middle) is *pedestrian in the path of the AV*, which is linked to the hazard (above the event) *loss of safe distance between the AV and pedestrian*. Note that a hazard is defined as an undesirable situation which can lead to an accident, and the top-event defines an event which can lead to the hazard - for some systems there may be many such top-events. There are multiple ways to model hazards and top-events based on engineering judgement, in this case we have put the top-event which could have been prevented by barriers on the left hand side, so that mitigations of its severity are on the right[1]. This is consistent with other reviews of the accident, such as [13].

The *threat* (blue box on left hand side) is *pedestrian crossing major highway* and there are a number of preventative *barriers* in place to stop reaching the situation that the pedestrian is in the path of the Autonomous Vehicle (AV). These include object detection and classification, the safety driver taking preventative action, and braking or maneuvering by the Automated Driving System (ADS). There are a number of *escalation factors* (yellow boxes) which impact the effectiveness of the barriers.

The first issue highlighted in the accident report [16] is that the classifier (the first barrier) did not identify the bicycle until too late to prevent the top event. It cycled between different incorrect classifications. This could be considered a result of ineffective training by the engineering team (either through poor data selection or verification). Note that the report doesn't specify or name individuals, therefore we refer to Uber ATG the organisation as the role responsible agent. Another set of technical barriers are the object path predictor, alert to safety driver that there is a potential hazard, and calculation of whether this can be avoided. However, there were a number of escalation factors associated with these, including suppression of warnings due to the number of false alarms. Trade-offs between functionality and safety are an inevitable part of safety engineering, however in this case they were not well justified [13] and were a causal factor in the accident according to the NTSB report [16].

Intervention by the safety driver was possible both before and after the top-event, as they could have intervened on first sight of the pedestrian, without waiting for the ADS to respond. This barrier therefore appears on both sides of the bow-tie. However, there were a number of escalating factors preventing this being effective, including the driver being distracted, and a lack of monitoring by Uber ATG to ensure the safety driver was performing their role as needed. Additionally, a second safety driver was no longer used for trials, following updates to the interface by Uber ATG to make interaction and intervention easier for a single operator. This meant tasks of monitoring and intervention that were previously shared were now demanded of a single driver.

[1]The outcome of one bow-tie can be linked to be the threat of another forming a more complex model.

After the event there are still mitigation barriers which could have been applied to reduce the severity of the *outcome* of *Fatal collision/severe injury* (red boxes on right hand side). This includes emergency braking (part of the SUV but deactivated by Uber ATG's engineers) as well the safety driver.

Three roles are explicitly shown, Uber ATG (this would include engineering staff, managers and many agents in the company), the safety driver, and the emergency services. If it was possible for the emergency services to arrive in time, they could have potentially treated an injured party and reduced severity of outcome. In this situation, their presence made no difference. Implicit is the role of the pedestrian as a third party (the *threat*), and the second safety driver who was no longer in the vehicle, i.e. this is a potential role, and it being dropped should have been explicitly justified.

The sociotechnical analysis [13] and accident report [16] both identify other causal roles including regulators, and pressure to deliver from company directors. Therefore, although a bow-tie does represent some roles, we cannot assume it is complete for a causal analysis of the incident, and hence for a safety justification considering responsibility. Nor could a SAC based on the bow-tie be used to support a complete backwards looking causal analysis of the accident. In our experience other safety analysis methods are likely to also be incomplete with respect to role/causal responsibility analysis, and potentially less complete than bow-ties as roles are not explicit.

Our bow-tie model also includes a judgement of the effectiveness of the barrier (indicated by red/amber/green on the barrier top). Whilst we do not explore this aspect further in the paper, potentially the effectiveness of a barrier could be considered in causal analysis when determining scale/size of causal contribution. However, a barrier having low-effectiveness is not necessarily a reason why it wouldn't be incorporated, particularly if it was a small, low cost modification which could reduce the risk even slightly. This is in keeping with the principle of reducing risk as far as possible or practicable.

It is of note that most of the safety barriers in the bow-tie are related to the roles of Uber ATG rather than the safety driver. This also may not be indicative of scale/size of causal contribution. From a legal perspective, only the safety driver is being pursued for legal-liability responsibility.

4.2 Mapping roles and responsibility types to findings

As not all roles were identified in the bow-tie, we have performed an alternative analysis based on the core findings from the accident report, considering the senses of responsibility for each finding and their roles, using the discussion from sections 2 and 3. We identified specified roles and considered where they had compliance and legal duties. We have not considered any judgment of size or impact of each of the causal responsibility findings, only whether the accident report listed them as contributory or not.

The analysis is in Table 1 and each of columns has been populated as follows:

Finding – we list each of the findings listed in the NTSB report executive summary.

Related roles – this specifies the specific individuals or organisations which have direct role relationships with the causes in the findings. As we have limited information on the development team for the ADS we list the umbrella organisation of Uber ATG for both engineering and safety monitoring roles. Additional roles of lawmakers and regulators were added to our review. The report also made note of some issues which were potential causal factors in the outcome (i.e., the fatality), but where these did not impact on the severity of outcome hence are not contributory causal factors. These included emergency services who were considered to perform as required.

Compliance duties refer to guidelines which we or the NTSB report, consider to be potentially helpful to manage and mitigate the outcome, this includes guidelines for specific use of the vehicle, as well as wider safety guidelines considered applicable to automotive vehicles. A rigorous analysis of relevant compliance documents would be required when building an assurance case, along with evidence of how they are adhered to. Alternatively, justifications for compliance shortfalls could be provided. This is common practice in safety engineering e.g., where expected safety activities are not relevant to a particular system or are compensated for with alternatives.

Legal duties refer to any specific legal requirements relating to a role which should have been considered prior to, or during operation, in other words in the forwards looking sense. In this case we have referred only to the legal duties which were listed in the NTSB report or are highlighted by ongoing legal action.

Causal contribution and moral responsibility is a complex issue. In section 2 we noted that causal contribution was a necessary but not sufficient indicator of moral responsibility. Further, there are many different models to determine causal contribution such as those described by Tadros in [23]. In the table we have included causal contribution as a choice of *Yes*, *Possible* or *None* based solely on the judgements in the accident report, but the threshold or measure by which the actor is considered causally, and morally, responsible will vary considerably. On a simplistic level, if an actor in a role has agency and were causally responsible, they could potentially be held morally responsible (even if only partly) for the outcome. However, a more nuanced analysis would need to consider contribution carefully, and further consider the relation between causal contribution and moral responsibility. As noted, all SCAS have a degree of residual risk associated with their operation.

Liability relates to legal liability from the backwards looking perspective, as judged by legal actions relating to the accident. For a SAC the relevant legal duties relating to SCAS safety would need to be systematically identified with evidence of how they will be or have been discharged, or justifications provided for any shortfalls. Where a causal contribution was found by the report we have noted if there is or was a possible liability case. We are not legal experts and we do not explore the findings further.

This review has highlighted an inter-relationship between roles as the duties they perform can impact on the duties of other roles, for example, Uber ATG as part of their safety activities needed to produce compliance guidance (i.e., training for their role) for the safety driver to follow. Also, the oversight from Uber ATG of ongoing performance of their safety drivers was found to be lacking. Considering role responsibility and relationship to other roles and

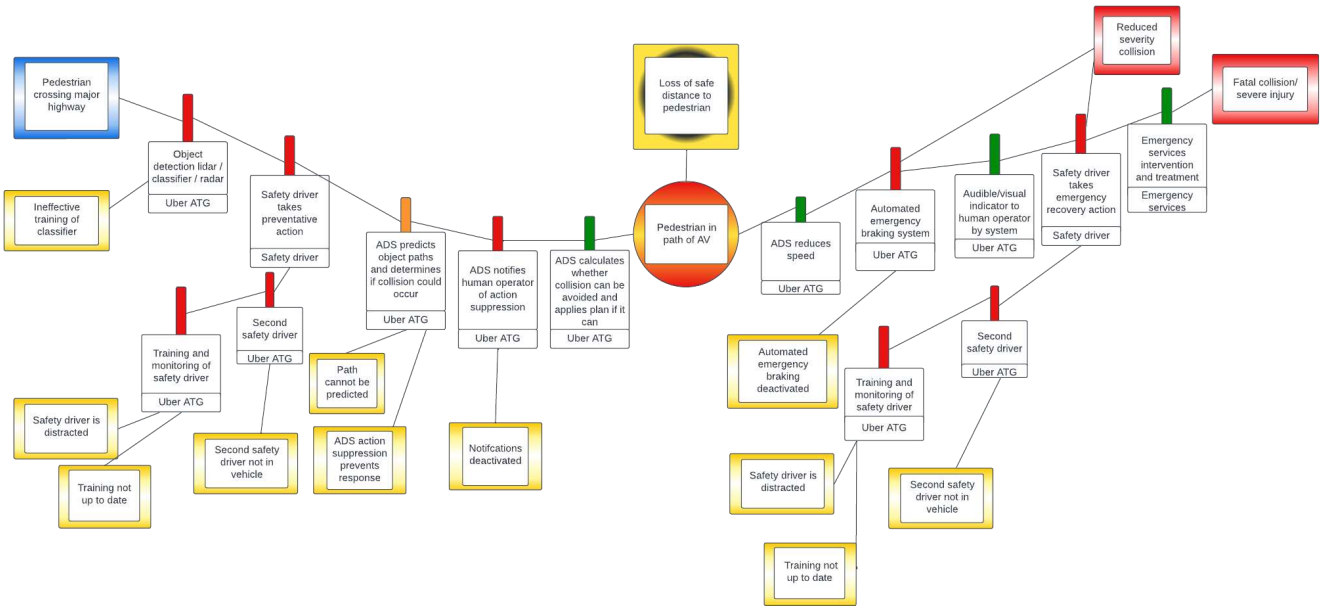

Figure 3: Bow-tie model showing safety barriers for the ADS

their duties could provide useful insight for role safety analysis. In other words, an actor in a role performing safety related duties can contribute not just directly to system risk, but also transfer risk to another role. This is another factor that complicates considerations of causal contribution and moral responsibility and which needs careful review and thought. Not all of these risks would be safety risks, for example they may include risks to reputation or personal autonomy. Ongoing work looking at developing an ethics assurance case would consider these aspects [18].

Similarly, the lack of regulatory oversight meant no specific compliance duty was required from Uber ATG to produce system safety assessments. Whilst typical practice would assume SAC or some similar safety justification was produced, it was not required and the inaction of the regulator impacted on the behaviour of Uber ATG. The consequence is that residual risk associated with the system was never scrutinised or challenged sufficiently. If there had been regulatory oversight or safety practice guidelines to follow, this would have appeared in our bow-tie as a barrier. The fact that there wasn't is arguably an escalation factor, but a barrier that doesn't exist is typically not something which that would be included. Considering whether to do so may be an avenue for extending our role safety analysis.

One issue to explore would be the scope of concern of the organisation performing the safety analysis and producing a SAC has, or is required to have. For example, a manufacturer would be expected to perform safety analysis of their system, but they do not typically develop regulatory frameworks, although they may contribute or adhere to them. Regulations may lead to limitations on the scope of roles and or type of risk mitigation duties identified by the safety analysis. We will consider the impact of scope of concern as part of future work.

Some of the issues are not yet resolved at time of writing, such as the liability of the safety driver in the incident. Liability of Uber ATG was not tested due to a private settlement with the family of the pedestrian.

5 SAFETY ASSURANCE CASES FOR ROLE RESPONSIBILITY

Having undertaken our case study we have identified that traditional safety analysis may not include different roles and their duties which impact on safety, such as those identified for the Uber ATG Tempe accident. Further, we noted the interaction between roles which meant that some depended on the outputs of others. In this section we consider a SAC structure to capture safety contribution to risk from the differing roles, and their inter dependencies. This argument could help frame the required elements more clearly, and thus support the next stages of our research. The question of scope and allocation of responsibility for elements of the case itself is planned for future work.

5.1 Role safety assurance case

A SAC is defined as *"a reasoned and compelling argument, supported by a body of evidence, that a system, service or organisation will operate as intended for a defined application in a defined environment"*[2]. For example, a claim might be that all safety requirements have been tested for a piece of software. Supporting evidence would include the results of a set of tests which are traceable to related safety requirements and cover them all. Additional evidence and claims would demonstrate the set of safety requirements was complete and valid. A typical SAC starts with a top claim that *System is acceptably*

Table 1: Uber ATG Tempe accident responsibility analysis based on the findings of [16]

Finding	Related Roles	Compliance duties	Legal Duties	Causal Contrb	Liability
Driver licensing and experience	Safety driver	N/A	Yes	None	None
Driver training	Safety driver	Compliance to training	None specific to safety driver role	None	None
Driver impairment	Safety driver	N/A	Yes	None	None
Driver attentiveness	Safety driver	Compliance to training	Potentially	Yes	Legal case is not resolved
Monitoring driver attentiveness	Uber ATG	None identified	None	Yes	Not tested – settlement with pedestrian's family
Emergency response	Emergency services	N/A	Yes	None	None
Vehicle condition	Uber ATG	None identified	Yes	None	None
Pedestrian behaviour	Third party	N/A	Yes – should not have crossed road	Yes	None
Pedestrian impairment	Third party	N/A	Not known	Possible	None
Uber ATG limited risk analysis of ADS and experimental systems	Uber ATG	Established risk assessment principles would apply, nothing specific for ML components	None known	Yes	Not tested – settlement with pedestrian's family
Braking action suppression	Uber ATG	Established risk assessment principles would apply	None known	Yes	Not tested – settlement with pedestrian's family
Uber ATG deactivated automated braking system	Uber ATG	Established risk assessment principles would apply, this issue has been addressed [3]	None known	Yes	Not tested – settlement with pedestrian's family
Removal of second driver	Uber ATG	Established risk assessment principles would apply	None known	Yes	Not tested – settlement with pedestrian's family
Lack of ATG safety culture	Uber ATG	Established good practice could apply, now being addressed [3]	None known	Yes	Not tested – settlement with pedestrian's family
Lack of regulatory oversight	NHTSA/ Regulator	Not available	Not available	Yes	None
Lack of state legal oversight	Arizona law	Not available	Not available	Yes	None
Lack of state or federal oversight specifically for AVs	Regulator/ lawmakers	Not available	Not available	Yes	None

safe. This is deconstructed into a series of claims about identification of hazards, safety requirements, design and implementation, and operational procedures.

We have used the Goal Structuring Notation (GSN) [2] to express the argument section of the case, and commented on the type of evidence required. The symbols shown in Figure 4 are as follows. Rectangles represent claims (e.g. *G1*) we wish to make about the system, and parallelograms (e.g. *Decompose over parts of claim*) describe the way in which the top-claim is logically decomposed into more detailed claims. Where there is text in braces e.g. *{Risk}*

this will be instantiated at a later date with specific information where known. Lozenge elements to the side of *G1* show contextual information, some of which is contained in other parts of the argument, e.g., we assume definitions of risk would be found in the main safety argument so could be referenced from there.

As a starting point for our argument we propose the fragment shown in Figure 4. This has the top claim *G1 - All roles associated with causes of {Risk} for {SCAS} have been identified where possible, and their risk contribution reduced to be {acceptably safe}*. This claim is deconstructed into its constituent parts. We will need to define risks, the SCAS (which may or may not be a physical system, and could include related services such as telecommunications and operators of that system) and what we mean by acceptably safe, e.g., tolerable and reduced as far as possible. Our case is scoped to cover safety risks, i.e., we are not currently considering security, societal risk, ethics and similar issues which also are of concern for SCAS.

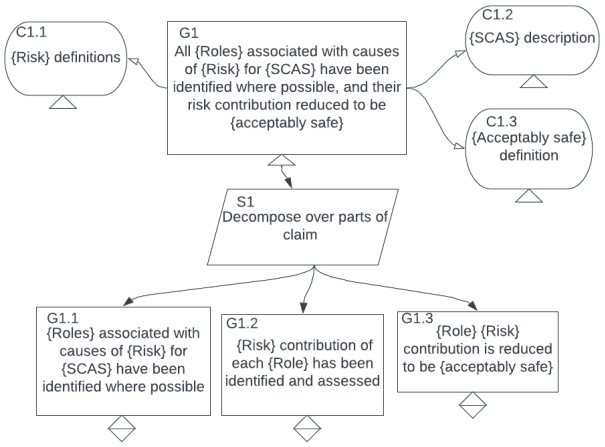

Figure 4: Fragment of SAC for safety role contribution

For the first sub-goal (G1.1) we wish to identify the roles associated with causes of risks for the system. Our case study identified that existing safety analyses are unlikely to be sufficient for this task. Developing richer safety analysis methods which both identify causes of risk and the relevant roles, duties, and scope of these is needed as evidence to support this claim. We have a caveat to this claim that these are identified where *where possible*. There may be causes of risk for which no related roles are identified or even exist (for example, some environmental risks may fall into this category). Further, there is the problem of many-hands discussed earlier in the paper, as it may be difficult to locate specific agents and roles.

Note that roles may be part of the design process, or part of the operational management of the system. One thing that is especially important for a SCAS is the continuing roles during operation which may now sit with a manufacturer, rather than human agents.

We would expect that all reasonable means to identify risk/cause associations from the various roles would be undertaken, however this needs to be proportionate with the risk. For a safety system with relatively low criticality (e.g., which could cause only minor injuries at worst), this analysis would probably not need the depth of that for a system with potentially catastrophic outcomes. Justification for the depth of the analysis is required in the safety argument. Another important point is that although we have said a role could be associated with the risk, they may not be necessarily be responsible for managing or contributing to the risk, directly or indirectly. Nevertheless, if the way a role is undertaken (including any duties associated with that role) could potentially have a causal contribution it should be identified.

Once we have identified each of the roles which are associated with risks, we then need to consider the role contribution to risks (G1.2). We would expect that each risk and role contribution must be considered. This allows us to ensure at least minimum coverage of all the possible pairs. However, it should be noted that roles may contribute (directly or indirectly) to multiple risks, and it should not be assumed they are independent. As shown in the case study, some roles will depend upon the output or performance of others. The means by which we do this, e.g. by a role transparent causal safety analysis, is the next stage in our research.

Having determined each of the {Role}{Risk} causal contributions, we need to ensure those are managed to be {acceptably safe} (G1.3). This section of the argument will be more complex and refer to claims in the main safety argument around measures which have, or have not been, chosen to reduce each risk. The latter will be an important claim as it relates to residual risk associated with the SCAS, and justifies why some risk reduction measures were not possible to implement or are impracticable. In other words, even where there is causal contribution from a role holder after an incident, they may not be morally responsible. One area that may or may not evidenced in an existing safety argument is whether duties have been performed (e.g., developed or AI using good practice), or are continuing to be performed (e.g., monitoring of the safety driver or of the SCAS safety performance) by a particular role.

We also propose a claim that risk exposure from one role to another is acceptably safe would be needed. This could be fulfilled by a similar argument to that found in the ethics case in [18], although the scope of that argument is wider and different than ours, which is focused on safety roles.

6 SUMMARY AND FUTURE WORK

This section contains a summary of our findings and proposed directions for future work.

Every SCAS has residual risk. Existing safety engineering practice should identify and reduce that risk to acceptable levels. Means to reduce risk are aligned with different engineering, operating and other roles, but for a SCAS this is more likely to be a design or manufacturer role. To support a SAC, we should demonstrate that all safety-related duties have been, and are being, performed by relevant role-holders. Further, following an incident we need to identify the causal factors, and related duties and roles. This is not necessarily to blame or punish those involved, but to prevent any mistakes happening again and to learn and improve safety where possible.

Existing safety analysis methods may not sufficiently identify role responsibility for causal factors leading to safety risks. We illustrated this using an example bow-tie of the safety barriers identified

by the Uber ATG Tempe accident report [16] and compared this with full findings of that report. Several potential causal contributions to the accident would not have been found. The case study also highlighted that there are complex inter-relationships between different roles which will impact on assurance activities and duties. We further noted that a role's causal contribution to risk and related links to causal contribution and moral and legal responsibility for a SCAS, are complex and require further research.

We have developed a template argument structure which provides a set of claims which could be used within a SAC to identify all roles and their potential contributions to causes of risk, and their mitigation. This provides a starting point for developing a role transparent safety analysis method which identifies roles as well as causes of risk. This can both help to improve safety as well as help agents with role responsibilities that include duties for reducing risk justify their decisions and actions, and show that they are not necessarily blameworthy despite causal contribution to an accident.

Our work is potentially generalisable to other safety-critical systems, but we are specifically interested in SCAS due to the more direct link between design decisions and operating decisions which could cause an accident.

REFERENCES

[1] UK General Public Acts. 1974. Health and Safety at Work etc. Act 1974. https://www.legislation.gov.uk/ukpga/1974/37/contents.

[2] ACWG. 2021. *Goal Structuring Notation Community Standard.* Technical Report SCSC-141C v3.0. Safety Critical Systems Club. https://scsc.uk/scsc-141C

[3] Uber ATG. 2018. Safety Report Supplement Internal and External Safety Reviews. https://aurora-dev.cdn.prismic.io/aurora-dev/4f3e03fe-7d41-4bed-a998-7b2d918b9579_UberATGSupplementSafetyReview2018.pdf

[4] Simon Burton, Ibrahim Habli, Tom Lawton, John McDermid, Phillip Morgan, and Zoe Porter. 2020. Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective. *Artificial Intelligence* 279 (2020), 103201.

[5] Civil Aviation Authority(CAA). 2023. Introduction to bowtie. https://www.caa.co.uk/safety-initiatives-and-resources/working-with-industry/bowtie/about-bowtie/introduction-to-bowtie/. Online - accessed February 2023.

[6] A. Feder Cooper, Emanuel Moss, Benjamin Laufer, and Helen Nissenbaum. 2022. Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 864–876. https://doi.org/10.1145/3531146.3533150

[7] Sidney Dekker. 2018. *Just Culture: Restoring Trust and Accountability in Your Organization, Third Ed.* CRC Press.

[8] Ibrahim Habli, Tim Kelly, Kevin NJ Macnish, C Megone, Mark Nicholson, and Andrew Rae. 2015. The Ethics of Acceptable Safety. In *23rd Safety-critical Systems Symposium, SSS 2015.*

[9] H. L. A. Hart. 2008. 210POSTSCRIPT: RESPONSIBILITY AND RETRIBUTION. In *Punishment and Responsibility: Essays in the Philosophy of Law.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199534777.003.0009 arXiv:https://academic.oup.com/book/0/chapter/160311870/chapter-pdf/38995148/acprof-9780199534777-chapter-9.pdf

[10] Richard Hawkins, Colin Paterson, Chiara Picardi, Yan Jia, Radu Calinescu, and Ibrahim Habli. 2021. Guidance on the assurance of machine learning in autonomous systems (AMLAS).

[11] ISO. 2018. ISO-26262 Road Vehicles Functional Safety.

[12] ISO. 2019. ISO-14971 Medical devices. Application of risk management to medical devices.

[13] Carl Macrae. 2022. Learning from the Failure of Autonomous and Intelligent Systems: Accidents, Safety, and Sociotechnical Sources of Risk. *Risk Analysis* 42, 9 (2022), 1999–2025. https://doi.org/10.1111/risa.13850 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.13850

[14] UK Maritime and Coastguard Agency. 2022. Improving Safety and Organisational Performance Through A Just Culture. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/286139/just_culture.pdf.

[15] Helen E Monkhouse, Ibrahim Habli, and John McDermid. 2020. An enhanced vehicle control model for assessing highly automated driving safety. *Reliability Engineering & System Safety* 202 (2020), 107061.

[16] National Transportation Safety Board. 2019. Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018, NTSB/HAR-19/03. https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf

[17] Tor-Olav Nævestad, Ingeborg Storesund Hesjevoll, and Ross Owen Phillips. 2018. How can we improve safety culture in transport organizations? A review of interventions, effects and influencing factors. *Transportation Research Part F: Traffic Psychology and Behaviour* 54 (2018), 28–46. https://doi.org/10.1016/j.trf.2018.01.002

[18] Zoe Porter, Ibrahim Habli, John McDermid, and Marten Kaas. 2022. A Principles-based Ethical Assurance Argument for AI and Autonomous Systems. https://doi.org/10.48550/ARXIV.2203.15370

[19] Z. Porter, A. Zimmermann, P. Morgan, J. McDermid, T. Lawton, and I. Habli. 2022. Distinguishing two features of accountability for AI technologies. *Nature of Machine Intelligence* 4 (Sept 2022), 734–736. Issue 9. https://doi.org/10.1038/s42256-022-00533-0

[20] Rail Safety and Standards Board (RSSB). 2021. Bowties in rail - case studies. https://www.rssb.co.uk/safety-and-health/guidance-and-good-practice/bowties/bowties-in-rail-case-studies. Online - accessed February 2023.

[21] RTCA/EUROCAE. 2011. DO-178C Software Considerations in Airborne Systems and Equipment Certification.

[22] Mark A Sujan, Ibrahim Habli, Tim P Kelly, Simone Pozzi, and Christopher W Johnson. 2016. Should healthcare providers do safety cases? Lessons from a cross-industry review of safety case practices. *Safety science* 84 (2016), 181–189.

[23] Victor Tadros. 2018. Causal Contributions and Liability. *Ethics* 128, 2 (2018), 402–431. https://doi.org/10.1086/694275 arXiv:https://doi.org/10.1086/694275

[24] Dennis F. Thompson. 1980. Moral Responsibility of Public Officials: The Problem of Many Hands. *American Political Science Review* 74, 4 (1980), 905–916. https://doi.org/10.2307/1954312

[25] Dennis F. Thompson. 2017. *Designing Responsibility: The Problem of Many Hands in Complex Organizations.* Cambridge University Press, 32–56. https://doi.org/10.1017/9780511844317.003

Received 08 March 2023