ARTICLE :

AI evidence and the future of motor vehicle accident disputes

By Jo-Ann Pattinson

Introduction

The following considers disputes involving automated vehicle (AV) collisions and the potential inconsistencies or issues surrounding the proposed transparency frameworks for AI, their practical application and how these translate into existing rules and standards relating to evidence presented before UK Courts. To facilitate the discussion, we consider a hypothetical scenario where we apply legal standards, rules of evidence and methods of explaining AI to a plausible incident involving AV. We examine how issues may be dealt with by; Claimants Defendants and the Courts, and the impact this may have on individuals or parties to the dispute, and upon the result of the dispute itself.

At the time of writing, no AVs have been type-approved for driving on UK roads. However, for the purposes of this paper, we assume that in the near future, it will be the case that AVs will be developed for the UK market with sufficient performance capabilities to be classified as an automated vehicle in accordance with the Automated Vehicles Act (2024). It is further assumed that the self-driving capabilities will be facilitated by using 'deep learning' AI.

The term "AI" as it is used in AVs is differentiated from traditional software systems which may execute an operation in a vehicle based on a rule made by a natural person. AI refers to a system which is able to produce outputs or decisions based on inferences or predictions derived from models or algorithms trained on data[1]. By deep learning AI, we refer to a complex class of artificial neural networks (ANN) which are capable of iteratively learning and making predictions based on examples provided by data. Artificial neural networks can learn in both a supervised, or unsupervised manner. Where learning is unsupervised, the model is trained from unlabelled examples[2]. When an AV or a class of AVs is designed for self-driving in an open environment, the AI involved is likely to utilise deep learning as opposed to a process-based decision-making model.

The degree to which the AI system will operate and adapt, free from human oversight or input, will entirely depend on the individual system. Further, the degree to which the outputs, actions and decisions made by AVs in self-driving modes can be explained or interpreted in a way that is understandable in a meaningful way, so that it may be considered by the Courts, will depend upon the opacity and complexity of the system, and the tools used to explain it. Explainable AI (XAI) refers to tools used to describe the expected impact and outputs of an AI model[3].

The types and capabilities of AI used in AVs will vary between manufacturers, and will also be dependant on local laws governing AV use. For example, several states in the USA permit vehicles operate without a driver behind the wheel in certain circumstances or on certain stretches of road[4]. In contrast, the UK does not yet have driverless vehicles being sold to the public for personal use on public roads, however it has initiated the legal groundwork for

[1] Razavi, Saman (2021) 'Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling' Environmental Modelling and Software (144) 105159 (October 2021) https://doi.org/10.1016/j.envsoft.2021.105159
[2] EU AI Act (2024) <https://www.euaiact.com>
[3] Dwivedi, Rudresh et al (2023) 'Explainable AI (XAI): Core Ideas, Techniques and Solutions' ACM Computing Surveys 2023-09 55(9) 1-33 (Article 194)
[4] Banner J (2023) Are Self Driving Vehicles Legal in My State? Motortrend 6 January 2023 https://www.motortrend.com/features/state-laws-autonomous-self-driving-driverless-cars-vehicles-legal/

vehicles fitted with automated lane keeping systems on public roads[5], and has recently passed the Automated Vehicles Act (2024). The revolutionary aspect of this framework being introduced in the UK is that, for the first time, drivers of these vehicles will be permitted for some part of the journey, to remove their hands from the wheel and take their eyes off the road. While at the time of writing no such vehicle has been approved, regulation exists in anticipation of these types of vehicles being manufactured for the UK market. The UK government defines an AV as a vehicle that is 'designed or adapted to be capable, in at least some circumstances or situations, of driving safely themselves'[6]. Where an AV enters a situation which exceeds its operational design domain (ODD), it may transfer the operational responsibility back to the user-in-charge (UiC). The transfer should be signalled by haptic, audible and visual signals, as described by UN Regulation 157[7]. As a public road is a dynamic environment, for a vehicle to be driving itself without human input, an AI system must be in place to make decisions and execute actions instead of the driver. These decisions and actions will be based on predictions or inferences made by a model or algorithm trained on data. In the event the AV is involved in a collision, there are obvious opportunities for dispute, which may result in a criminal or civil court matter involving AVs.

The regulation of AI is rapidly developing at a national and international level. Both national and international regulatory instruments relevant to AI broadly includes requirements that AI be; secure, robust, transparent, fair, and accountable[8]. Systems including AVs which utilise AI should be trained on high quality data, validated, and tested[9]. It is widely agreed that AI should be 'human-centric', while ensuring a high level of protection for health, safety and fundamental rights - sentiments that feature heavily in statutes purporting to regulate AI and the data or information it generates. Principles such as ensuring that AI is ethically sound, and compatible with societal well-being, technically robust, and subject to human agency and oversight are a fundamental component of the developing governance of AI.

The relevance of this to AVs is that when a collision occurs, in particular when it involves damage, injury or death, we should be able to explain how that accident occurred. This is essential for improving the safety of AVs and for the purposes of protecting human life. However, it remains to be seen whether the methods developed for this will also be used to determine liability for damage caused by an AV collision. The issue of how an AV came to be involved in a collision, and how its design or manufacture may have contributed to this, is a central discussion point that we will expand upon below in a learning scenario. This is designed to test the interaction between; civil litigation involving motor vehicle disputes, the law of evidence, the transparency of AI, and the framework for claims which has been developed in the UK in anticipation of automated vehicles being sold and driven on public roads in the near future.

Method

The UK legal system is built upon a rich body of common law where real cases and scenarios have been examined, and judges have applied law to the facts. In the case of AV disputes, the law and technology are developing, and a body of knowledge or precedent using, developing and interpreting UK laws is yet to be created. Consequently, we must learn by simulation. In this paper, a learning scenario is proposed as a thought experiment to bridge the existing gap between empirical facts[10] and theoretical concepts, to increase our understanding of the impact of proposed legal frameworks. Learning scenarios are useful in evaluation, to provide context and a reference point to assist in discussing how the technical, social and legal systems developing for AVs will interact and impact those involved.

---

[5] United Nations Economic Commission for Europe, Addenda to the 1958 Agreement (Regulations 141-160), (United Nations Economic Commission for Europe, 2021) < https://unece.org/transport/vehicle-regulations-wp29/standards/addenda-1958-agreement-regulations-141-160>

[6] Automated and Electric Vehicles Act 2018 UK <https://www.legislation.gov.uk/ukpga/2018/18/contents>

[7]

[8] Artificial Intelligence (Regulation) Bill https://researchbriefings.files.parliament.uk/documents/LLN-2024-0016/LLN-2024-0016.pdf

[9] European Parliament (2024) Artificial Intelligence Act P9_TA (2024) 0138 https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf page 14

[10] Stewart Michael T (2018) 'How thought experiments increase understanding' Brown JR (ed) The Routledge Companion to Thought Experiments (Taylor and Francis Routledge Handbooks)

Table 1 below describes a learning scenario which simulates a set of circumstances involving an AV collision and subsequent claim. The learning scenario has been used in place of actual cases. In the learning scenario, a vehicle driving itself (but with a user-in-charge on board) collides with a bollard on an approach to roadworks. Roadworks are stated to be outside of the vehicle's ODD. The accident also occurred on a rainy day.

The full discussion of how this scenario may be dealt with by a Court will occur after we set out all the relevant laws and legal standards relating to evidence in the UK. We will also discuss the legal landscape of product liability as it is developing in the UK and the EU, and what can be achieved by XAI.

Court as a stakeholder- what type of AV disputes will come before a Court?

Types of Claim (UK)

The following will discuss the different type of claims that may be made for financial compensation following an AV collision causing injury or damage, those being; (1) an insurance claim pursuant to the Automated and Electric Vehicles Act (2018) (AEVA) and subsequent actions which may result if an insurer denies such a claim, (2) issues surrounding a claim under product liability law, and (3) requirements for a claim under the common law for negligence.

Insurance Claim

The principle type of claim involving an AV following an incident or collision will be an insurance claim pursuant to the AEVA, which states that where an accident '*is caused by an automated vehicle driving itself.. it is insured…and a person is injured or suffers damage'* then the insurer is liable for that damage[11]. This provision is designed so that an insured injured party need not prove a case in Court in order to receive financial compensation.

However, if the insurer denies the claim, then proving a case in Court is exactly what the party will be forced to do. If the UiC or other road user has claimed that the vehicle was driving and at fault, and the insurer makes a decision that the vehicle was either (1) not driving or (2) driving but not at fault, then the claim must be pursued through the Courts. Further, even if an insurer accepts a claim, there may be elements of financial loss or damage which may not be payable under the insurance policy.

If the insurer denies (or partly denies) the claim

Section 3 of the AEVA sets out what occurs where the insurer is liable; however, the injured party contributed to the incident by their own negligent actions. In this case, the liability should be reduced in accordance with the Law Reform (Contributory Negligence) Act 1945 (LRCNA). The relevant provision of LRCNA states:[12]

'*the damages recoverable in respect thereof shall be reduced to such extent as the court thinks just and equitable having regard to the claimant's share in the responsibility for the damage'*

Deciding 'the extent' is likely to be contentious with the insurer seeking to apportion as much blame to the insured as possible. Where the parties do not agree, this can only be resolved by a court.

Another type of claim that may result in a contentious hearing may be triggered by Section 4, which limits the insurer's liability where an insured person failed to install software updates or where the insured person made unauthorised software alternations. As software updates are commonly installed automatically, it is unclear whether a dispute may arise where an insured failed to notice an automatic update had failed to install and whether an insurer would attempt to rely on Section 4 in these circumstances. Additionally, an insured may have accidentally manually uploaded an incorrect version of the software, invalidating his insurance despite acting in good faith.

Product liability claims

Traditional vehicles (i.e., those not driven by AI) can be considered a 'product' under consumer protection law. In the UK, someone who suffers injury due to a faulty product has two avenues of recourse; one being to claim under the Consumer Protection Act (1987) ("CPA"), and the other is to claim under the common law (judge-made law) that a manufacturer or developer breached their duty of care by negligently creating a defective product causing harm which was reasonably foreseeable and preventable. Claims made under the common law require a Claimant to

[11] Automated and Electric Vehicles Act 2018 UK <https://www.legislation.gov.uk/ukpga/2018/18/contents>
[12] Law Reform (Contributory Negligence) Act (1945) <https://www.legislation.gov.uk/ukpga/Geo6/8-9/28/contents>

prove negligence and that the damage was caused by the negligence. Whereas under the CPA, there is strict liability for damage caused by a defective product[13]. That is, the Claimant does not have to prove the Defendant was negligent when producing the product; only damage or injury was caused. When a defective or unsafe traditional vehicle is involved in an accident in the UK and causes death, injury or damage, manufacturers are liable under the CPA. However, AI is not considered to be a product as defined in this legislation. Consequently, if damage, injury or death is caused by defective AI in an AV, there can be no claim under the CPA; this is explained further below.

Changes are in the pipeline internationally in product liability law, which may impact the validity of claims made in relation to AVs and AI, as will be expanded upon below.

At the time of writing, in the UK and the EU, product liability law protecting consumers is based on the Product Liability Directive (85/374/EEC) (Existing PLD), which establishes a liability regime against manufacturers and suppliers for defective products[14]. In the UK, the CPA brought the Existing PLD into domestic law. However, as referred to above, AI does not neatly fit into the CPA's definition of a product, which includes 'goods or electricity'. A new EU Directive replacing the Existing PLD is expected to be adopted in in 2024. The new Product liability directive (New PLD) will extend product liability to protect individuals in EU member states (who may or may not be consumers) and will include provisions confirming software and AI is a product and subject to the terms of legislation enacted under the New PLD. The type of damage which may trigger compensation will include death, personal injury, psychological injury, and damage to property and destruction of data. The New PLD also indicates that the requirements set out in the EU AI Act should be considered when determining if a product is defective. Consequently, in EU member states, if AI, which comprises part of an AV, is defective and causes injury or death, a claim may be possible under the new PLD.

However, for the UK, since Brexit, we do not yet know whether the CPA will be amended in line with the New PLD. The former UK Government signalled a 'pro-innovation' framework with little change to legislation[15]. At the time of writing, the new UK government has provided some indication further regulation relevant to AI can be expected.

"*Labour will ensure the safe development and use of AI models by introducing binding regulation on the handful of companies developing the most powerful AI models*"[16]. However, this does not mean that the current government will enact legislation incorporating product liability in the same fashion as the EU. As it stands, the only path for claimants to take in the meantime, if they have suffered damage or loss as a result of defective AI, is to sue for negligence under the common law.

Negligence

Negligence is a concept established and developed in the UK by the common law or judge-made law. Negligence occurs when a person's actions, which impact others, fail to reach the standard expected of a 'reasonable person'. If a person fails to act as a reasonably competent person would, and these actions cause foreseeable damage, that person may be held liable for any resultant damage or loss[17].

This concept extended to manufacturers being liable for negligence in the seminal case *Donaghue v Stevenson*[18]. This established the principle that a manufacturer must exercise reasonable care to avoid causing foreseeable harm to a person (or class of person) using or coming into contact with their product. A manufacturer failing to carry out that duty is liable for injury or damage related to that failure. When considering whether a manufacturer was negligent, a Court may consider the likelihood of the injury or damage, any relevant industry standards, and the seriousness of the injury relative to the effort and cost to prevent it.

In the case of *Donaghue v Stevenson*, Mrs O'Donoghue became unwell after drinking a bottle of ginger beer, which contained the remains of a decomposing snail. Upon inspection, it was found that the packing area of the ginger

---

[13] Consumer Protection Act (1987) < https://www.legislation.gov.uk/ukpga/1987/43>
[14] Product Liability Directive (85/374/EEC) < https://eur-lex.europa.eu/eli/dir/1985/374/oj>
[15] Office for Artificial Intelligence (2023) 'A pro-innovation approach to AI regulation' https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
[16] Labour: Labour Manifesto Our Plan to Change Britain (13 June 2024) < https://labour.org.uk/updates/stories/labour-manifesto-2024-sign-up/>
[17] *Hall v Brooklands Auto-Racing Club* [1933] 1KB 205
[18] *Donaghue v Stevenson* [1932] AC 562

beer factory contained open bottles and silver snail trails. Consequently, the 'system of working' fell below the expected standard. Had the factory been impeccably clean and systems in place to prevent snails, yet due to an unforeseen and unlikely circumstance, the snail ended up in Mrs Donoghue's ginger beer, her claim would have failed. If the system of working was to the expected standard, a negligence claim could not succeed despite her illness caused by the snail.

It is common for manufacturers to limit their liability for injury or damage which occurs outside of their control. Manufacturers often stipulate how a product or device should be used in order to avoid incidents causing damage or injury. For example, directions on a bottle of medicine about safe dosage or a warning that face protection should be used when operating a dangerous machine. These may be referred to as an 'exclusion clause'. In these circumstances, if the users of these products failed to follow directions and suffered an injury as a result, the manufacturer would not be liable. However, there are limitations to the effect of exclusion clauses, which may render them ineffective[19]. If a product is faulty and it was the fault that caused the injury, then directions being issued with a product would not protect the manufacturer from negligence. It is not possible to 'contract out' of negligence. Manufacturers will always have a duty of care to all members of the public who may come into contact with their product, and normally, the emphasis when assessing duty of care is on the 'system of working' or the process the manufacturer took to design and create the product.

However, while this is the case with any manufactured product in the UK, the situation may be different when it comes to AVs, which we explore below.

A New Reasonable AV Standard?

In 2024, the Automated Vehicles Act (2024)[20] defined the standard of safety required of an automated vehicle driving itself.

Section 2(2) states that a vehicle driving itself should operate to;

"*a level of safety equivalent to or higher than, a careful and competent human driver*'.

This has created a quasi 'reasonable AV standard' i.e., a reasonable person standard, for the car. The Automated Vehicles Act (2024) states a safety standard that the manufacturer must achieve, which is output-based. The standard is that of a vehicle which behaves like it is being driven by a careful and competent human driver. In this case, it is arguably unnecessary to examine the process, design or manufacture of the vehicle, as the standard focuses on how the vehicle performs. However, there are many instances where it may be necessary for evidence to be presented to a Court about the design and inner workings of an AV and how that impacted the vehicle's safety. If a claim is made about an AV driving itself by an injured UiC, and the insurer denies the claim on the basis that the vehicle was being driven outside of its operational design domain (ODD) or that the UiC was at fault, then the UiC would have no choice but to pursue the matter in Court. In order to respond or address allegations in a counterclaim that the UiC was at fault, and in order to prove the claim that the AV did not drive at least as safely as a competent human driver, the UiC may need to approach the claim in a manner similar to a traditional negligence claim. Under the traditional principles of negligence established by *Donaghue v Stevenson*, if a manufacturer was negligent when designing and producing a product which caused damage or injury, the person injured could be entitled to financial compensation. In order to do this, the UiC would need to access information related to the AV's perception, planning and control, and vehicle recorder data. In the current legal and technical framework, there are several issues that create potential problems for Claimants and Courts relating to output and performance data, as well as perception, planning, and control data, as set out below.

---

[19] *Andrews Bros Ltd v Singer & Co Ltd* [1934] 1 KB 17

[20] Automated Vehicles Act (2024) <https://www.legislation.gov.uk/ukpga/2024/10/contents>

Vehicle data recorders

Under UN Regulation 157 and Regulation 160[21] all vehicles must be fitted with an event data recorder (EDR) 'black box' which records limited data parameters such as speed, acceleration, braking, steering positions, seatbelt use and airbag deployment. EDR is only triggered by an event such as a crash or an airbag activation. All vehicles fitted with automated lane-keeping systems must be fitted with EDR and a data storage system (DSSAD), which records the operator status, i.e., whether (according to the vehicle) the human or the vehicle system was in control of the vehicle at the material time. We will refer to the data stored by EDR and DSSAD as "compulsory data parameters".

The type of data which may be recorded by vehicles but not stored by law includes exterior/interior video of the vehicle, external/internal audio, location data, the vehicle identification number or any data relating to driver alertness, such as eye movement, hand movement, seat occupancy data. We will refer to these parameters as "additional data parameters". While the data stored and recorded by EDR and DSSAD may be useful in determining liability in a crash, it may not provide enough data to establish the exact movements of the UiC, the environmental conditions in which the crash occurred, nor any external entity, which may have also contributed to the crash.

If a Claimant is seeking evidence relating to whether a vehicle drove itself to the standard of a competent and careful human, that the vehicle was operating within its ODD and that the UiC was not at fault, the very narrow compulsory data parameters may not be useful. The alternative options for a Claimant would be to call for the disclosure of the additional recorded data or to access XAI data. To assist with improving the transparency of AI systems and XAI, an industry standard has been developed. However, as we will explain, even if manufacturers follow the standard, the AI used in AVs may still be opaque or uninterpretable to a stakeholder like a Court.

Standard for transparency

The IEEE Standard for Transparency of Autonomous Systems[22] was developed for the objective assessment of autonomous systems and contemplates a recommended level of transparency for different stakeholders. For certification agencies, investigators and courts of law, the standard recommends a "Level 5" version of transparency.

This includes EDR capable of recording a time-stamped log of key system inputs and outputs, and storing decision-making logic or mechanism, behind each high-level decision. However, the standard acknowledges that for complex artificial neural networks (ANNs) which is necessary for a sophisticated self-driving AV, the determination of the reasons for decisions (ANN outputs for a given set of inputs) is not possible. However, it recommends that "*the system should periodically send the complete set of ANN connection strengths to the EDR in order to allow incident investigators to reconstruct the ANN in an effort to reproduce the sequence of outputs leading up to the incident*".

Recording periodic connection strengths with XAI methods, such as gradient weight class activation mapping, and providing decision visualisation, may assist in reconstructing the sequences leading to an incident. This is not the same as providing a record of how decisions were made. However, as the decision-making process of deep neural networks is opaque, it may be the only information available[23].

We will explore why there is difficulty in recording key system inputs, outputs, and decision-making logic, in the chapter below.

Industry standards for how a system should function is not normally a compulsory legal requirement for technical systems, unless these are required contractually. However, industry standards serve as a visible benchmark for best practice and can be considered by a Court in matters where a Court is asked to make a decision on whether a person or company has been negligent and not fulfilled their duty of care to the requisite standard.

---

[21] United Nations Economic Commission for Europe, Addenda to the 1958 Agreement (Regulations 141-160), (United Nations Economic Commission for Europe, 2021) < https://unece.org/transport/vehicle-regulations-wp29/standards/addenda-1958-agreement-regulations-141-160>

[22] IEEE P7001/2021 : A Standard or Transparency of Autonomous Systems <https://standards.ieee.org/ieee/7001/6929/>

[23] Winfield A, Joritka M (2018) 'Ethical governance is essential to building trust in robotics and artificial intelligence systems' Philos Trqans A Math Phys Eng Sci Oct 15: 376 (2133) doi: 10.1098/rsta.2018.0085. PMID: 30323000; PMCID: PMC6191667 < https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0085>

Explainable AI (XAI) Methods: Tools, Skills, and Bias

As mentioned earlier, it is assumed that when AVs enter the UK road space, they will be driven or operated by trained AI models, able to make predictions and decisions. However, exactly how an AI-driven system makes predictions, or why it arrives at a particular decision or course of action, is often unclear.

Explanation methods for AI algorithms are commonly divided into two categories: ante-hoc and post-hoc explanations. Ante-hoc methods are ingrained into the AI model during its design and development phase, ensuring their decision-making mechanisms remain transparent and interpretable. Certain methods in this category are intrinsically interpretable due to their simple nature, comprising decision trees or sparse linear models. However, these methods are unlikely to be applicable to sophisticated AI responsible for operating an AV. Post-hoc methods, on the other hand, are implemented after the AI model has been trained and deployed, explaining the decisions made by the AI model without modifying its architecture or training process. Ante-hoc explanations derive directly from internal representations, accurately reflecting the decision process. In contrast, post-hoc methods might not provide explanations with the same level of precision, as they rely on estimating explanations from input-output variations rather than accessing the internal decision-making mechanisms. Both methods produce explanations in the form of natural language and visuals, such as heat maps, saliency maps, graphs, and plots. In the context of AV, both ante-hoc and post-hoc methods are applied to varying extents to explain and visualize AI decisions across driving operations such as perception, localization, planning, and control.

Perception involves sensing an operational environment, encompassing two primary tasks: road surface extraction, which involves segmenting images into distinct regions, and on-road object detection through sensors including vision, LiDAR, RADAR, and ultrasonic technology. Various methods of explainability have been proposed to address early identification of potential traffic accidents, risk assessment in challenging road conditions, segmentation under hazy weather, and glare detection using vision data. Explanations are explored for LiDAR and RADAR data to detect 3D objects and sensor input uncertainty for possible deceptive attacks. Once the AV perceives its surroundings and gets its precise localization, it plans the trajectory from the initial point to the destination. Explanations pertaining to this operation primarily focus on route planning, which encompasses selecting an optimal route, and behaviour planning, which involves predicting and interacting with other road users sharing the same trajectory. These explanations are often presented through bird's-eye view visuals, highlighting the most influential elements in trajectory planning.

Control operation takes charge of executing planned motions. Primarily managed by feedback controllers, this function involves continuous interaction with sensors to guide the vehicle's trajectory throughout the journey. Explanations for the control and navigation operation, encompassing tasks such as longitudinal control (speed regulation) and lateral control (steering operation), are coupled with perception and planning processes. These explanations are often displayed as visual heatmaps, highlighting the most influential elements within the scene, accompanied by text-based explanations, e.g., "Stopping because cyclist stopped on my lane''.

XAI Tools

There are numerous off-the-shelf explainability tools in the form of open-source libraries, platforms, and visual analytic systems that can be plugged into ML models to explain a model's behaviour. These tools help analyse the model's internal representations to explain individual predictions, revealing key features shaping a model's decisions and their impact on the output. They differ in their functionality, usability, and features (e.g., ante-hoc/post-hoc explainability), programming language (e.g., Python, R, Java, C++), and compatibility with the development environment. Here are some noteworthy examples:

SHAP DeepExplainer: Python library based on the Shapley values concept from cooperative game theory.

LIME: Post-hoc explanation method for explaining individual predictions in a human-interpretable manner.

Scikit-learn: Machine learning library includes built-in model interpretation and explanation functionality.

Google What-if tool: An interactive visual exploration tool that helps users investigate and understand machine learning models.

AIX360: Toolkit developed by IBM Research that provides a comprehensive set of algorithms and tools for explaining machine learning models.

ELI5: A Python library that offers several techniques for model interpretation, including feature importance, permutation importance, and visualization of decision trees.

Shapash: A Python library that automates the process of explaining machine learning models and generating interactive reports.

Skater: A Python library that provides a set of tools for model interpretation and explainability.

EconML: A Python library developed by Microsoft Research that provides tools for causal inference using machine learning techniques.

Certain tools have the functionality to generate documentation automatically, facilitating the reporting and sharing of explanations in an understandable form. Additionally, some tools allow the reports to be shared in multiple formats, including PDF, HTML, and e-mail. While these tools may streamline the process and save resources by furnishing readily accessible explainability for AI systems, they also raise liability concerns, particularly when the system lacks adequate certification.

Skills of expert

As helpful as they are, many of these tools are primarily designed for experts or individuals with technical backgrounds who seek in-depth insights into the underlying mechanisms of the AI's decisions. This helps experts to fine-tune models, thereby enhancing accuracy and efficiency. However, the extent to which non-technical users benefit from these explanations remains unclear. Only a handful of tools, such as AIX360, tailor explanations to diverse user profiles, providing non-technical descriptions through natural language or highlighting important regions in an image. Some of the tools offer guidelines to assist users in selecting a suitable algorithm for their specific use case, along with examples demonstrating how users can use the tool to complete the task.

While these tools may generate explanations and simplify information for non-technical users, utilising them demands expert skills. A robust understanding of machine learning concepts, algorithms, and techniques, as well as knowledge of statistics, is paramount when working with these tools. Proficiency in programming languages and familiarity with specific explainability frameworks such as SHAP and LIME are also essential for implementing these tools. Finally, understanding the domain-specific challenges and requirements of autonomous vehicles is essential for effectively contextualizing and applying explainability techniques in this field.

Bias

Explanations provided by XAI techniques also have the potential to suffer from various biases. If the underlying AI model itself is biased due to biased training data or flawed algorithms, the explanations generated by XAI techniques will reflect these biases. For instance, if a model trained on biased data the explanations will inherently rationalize these biased predictions.

The algorithms used to generate explanations might introduce their own biases. Post-hoc XAI methods prioritize features that are statistically significant, potentially overlooking subtle but influential factors. This selective representation can create a skewed perception of the model's reasoning process, overemphasizing some features while downplaying others. Similarly, methods that rely on explanations by simplifications can omit important nuances and context, leading to explanations that are not fully accurate or that misrepresent the model's decision-making process.

Moreover, the presentation of explanations can significantly impact the explainee's perception. The choice of words, emphasis, and framing can introduce biases. Visual explanations, for example, may emphasize certain aspects, influencing how users understand the model's behaviour. Additionally, the explainee interpreting explanations may introduce biases. This can occur when users misinterpret the explanations based on their preconceptions, cognitive biases, or lack of domain knowledge, leading to inaccurate conclusions.

These issues are relevant for considering how information, predictions and explanations derived from XAI could be used as evidence in Court, which we explore below.

Issues relating to presenting AI in evidence

When disputes are being heard before a Court, the general rule about evidence is that all relevant evidence is admissible. However, the weight given to any evidence is determined by its reliability. There are many types of evidence which are inherently unreliable. Judges are responsible for determining what evidence is reliable and, otherwise, determining the weight to give different types of evidence.

A Claimant is more likely to prove their case on the balance of probabilities[24] using the best available evidence. This tends to be evidence which is accurate and verifiable.

Untrustworthy evidence should be excluded from consideration or given less weight. Untrustworthy evidence includes that cannot be tested or confronted through a mechanism such as cross-examination. In circumstances where an expert is called to interpret data from an AV there are several potential problems that may impact the reliability of the evidence. For example if;

- the AV system and AI operating the AV are unique, it may be that the manufacturer or AI developer are the only parties able to give evidence about their own system. In this case, experts in AI and AV engaged by the Claimant's solicitors are at a disadvantage when analysing it and cross-examining the Defendant's expert.
- AI decisions relating to perception, planning, and control have been carried out in an opaque 'black box' mechanism, then the full information will not be available.
- the data is not captured at the time of the incident but after the crash reports investigating decisions made by the algorithm may not be accurate.
- the Defendant defends disclosure requests about releasing certain information due to commercial sensitivity or other reasons, such as volume of data.
- Explanations are given, but they are not adequate or substantive enough for the Court to determine the legal question placed before it.

As discussed above, Post-hoc AI explanations are methods that provide explanations based on estimations of how different factors influence the AI's decision/action; it is not a record of what it did.

In this way, these types of explanations are similar to hearsay[25]. Hearsay generally refers to a statement made "*otherwise than by a person while giving oral evidence in proceedings, which is tendered as evidence of the matters stated*"[26] It is evidence that a witness has been told or heard from another source rather than having personal knowledge of what is stated. A post-hoc explanation may fall into this category of evidence. It is not based on actual knowledge but on access to data that could itself be incomplete, inaccurate or misinterpreted.

Transparency and Ambiguity (quality of explanation)

The UK Government has recognised the problem of opaque AI and how this presents a problem when determining accountability. In a policy paper, the Department for Science, Innovation & Technology, together with the Centre for Data Ethics and Innovation, stated:

'Some opacity may be expected depending on the AI system used, but what matters is that there is sufficient explainability for accountability similar to the level of explainability currently needed to hold human drivers accountable[27].

It is argued that the solution proposed in this statement is not possible in the context of the evidence rules. As explained above, Courts hold people accountable on the basis of reliable evidence. A post-hoc explanation of what,

[24] The Court is able to determine it was more probable than not that a particular fact occurred, ie that the Defendant was negligent.

[25] Seng D, Mason, S (2021) 'Artificial Intelligence and Evidence' Singapore Academy of Law Journal (33) 246
[26] Civil Evidence Act (1995) 33(1) < https://www.legislation.gov.uk/ukpga/1995/38>
[27] Centre for Data Ethics and Innovation 'Responsible innovation in self driving vehicles' https://www.gov.uk/government/publications/responsible-innovation-in-self-driving-vehicles/responsible-innovation-in-self-driving-vehicles#explainability

when, how and why complex AI made decisions, are only estimation of what it did. It is not obvious whether Courts will consider these types of estimations of themselves to prove something to the requisite standard, which is on the balance of probabilities in civil cases and beyond a reasonable doubt in criminal cases.

Rebuttable Common law principle of computer evidence

The UK jurisdiction presents a further potential complication to Courts determining the reliability of XAI, which is that there is a rebuttable presumption that data produced by a computer is accurate and reliable.

Originally, section 69 of the Police and Criminal Evidence Act 1984[28] required that in order to adduce evidence from computer records, evidence also had to be produced to show that the computer was operating properly. Section 60 of the Youth Justice and Criminal Evidence Act 1999[29] removed this stipulation. Therefore, the situation reverted to a common law presumption that a 'mechanical instrument' producing an evidential record was working properly at the material time, and therefore, the record produced by the computer is admissible unless evidence is produced to prove otherwise.

The obvious difficulty from the point of view of the party not in possession of the computer is that without access, knowledge and expertise, there is no simple method to ascertain enough information to rebut the presumption that a computer record is accurate. A notorious case in point (or cases) being the numerous claims involving the Post Office and the Horizon scandal, where the faulty Horizon system incorrectly produced data to show that numerous sub-post-masters all over the UK had been negligently or fraudulently operating their business, resulting in catastrophic consequences for individuals concerned[30].

Where it is difficult to challenge digital evidence, and where there may be a lack of expertise at the bench of XAI, there is the risk that a Court could place a greater weight on such evidence, not understanding the potential problems with presenting an accurate version of events.

EU AI liability Directive

Issues relating to AI and evidence such as these have driven the draft (EU) AI Liability Directive[31] ("AILD Proposal"). This draft legislation proposes that instead of the Claimant proving that damage was caused by AI, the rebuttable starting point will be that the damage was caused by the AI, and it will be up to the Defendant to prove that the AI was safe and/or did not cause the damage claimed. While the Directive is yet to be passed, it is uncertain how the UK will approach this issue. The White Paper on AI Regulation[32] states that there will be no change in the short term to the liability imposed on AI developers in the UK. AI Developers can be sued pursuant to existing laws (for example, under common law negligence), but there is no immediate intention of removing the burden of proof on a Claimant to prove that an AI developer has been negligent.

Leaning Scenario

Law/Statute	Example Circumstances
	Learning Scenario
S3 Automated and Electric	A UiC operating a SAE level 3 AV was driving itself when it collided with a bollard on an approach to roadworks. The collision occurred approximately 10m ahead of the roadworks. The vehicle's manual states that the self-driving feature should not be used

[28] Police and Criminal Evidence Act (1984) S 69 < https://www.legislation.gov.uk/ukpga/1984/60/section/69>
[29] Youth Justice and Criminal Evidence Act (1999) S60 < https://www.legislation.gov.uk/ukpga/1999/23/contents>
[30] Marshall P, (2020-03) 'The harm that judges do- misunderstanding computer evidence: Mr Castleton's story: an affront to the public conscience' Digital Evidence and Electronic Signature Law Review
[31] Artificial Intelligence Liability Directive: Proposal for a Directive of the Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf
[32] Office for Artificial Intelligence (2023) 'A pro-innovation approach to AI regulation' https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Vehicles Act (2018)	whilst travelling through roadworks. This is reiterated on the in-vehicle display when the operator turned on the AV. The collision occurred on a rainy and overcast day. Damage was caused to the AV as well as minor injuries to the UiC. The AV had slowed but had not stopped or alerted the UiC to a potential crash. Signs indicating 'roadworks ahead' had been displayed 1km before the roadworks, these were very dirty from being splashed by mud and water from passing cars. The UiC had been reading a book at the time of the crash. A claim was made by the UiC for damage to the vehicle and minor injuries. The insurer responded that the insured UiC was 80% responsible for the accident as the roadworks placed the crash outside of the vehicle's ODD. The insured UiC disputes the insurer's decision on liability and argues that the vehicle was not travelling through the roadworks at the time of the accident. The UiC also argues that the AV failed to operate at least as safely as a careful and competent human driver when it failed to recognise the signs warning of roadworks ahead and alert the driver, deal with the bollard, or come to a safe stop.

Table 1: Learning Scenario

Learning Scenario Discussion

This learning scenario considers a case where it is alleged both the vehicle and another party (i.e., the UiC) are at fault. The UiC would likely claim that the manufacturer or AI provider was negligent in that it did not produce a vehicle which met the requisite standard, being: it failed to operate to the standard of a careful and competent driver, as it did not avoid the crash. In answer to this, the manufacturer/insurer has alleged a counterclaim that the UiC contributed to the crash by operating the vehicle outside of its ODD, i.e., in roadworks and has asserted that the manufacturer/insurer liability ought to be reduced by 80%.

Claim by UiC against the manufacturer The manufacturer of the AV owed the UiC a duty of care when it produced the AV, to produce an AV to the required safety standard expected of an AV manufacturer, which is a vehicle that can operate as safely as a careful and competent driver when driving itself. The manufacturer failed in its duty of care to produce an AV which operated to the requisite standard. The AV failed to operate like a careful and competent driver when: It failed to recognise roadworks signs. It failed to warn the UiC of roadworks. It failed to avoid the crash. As a result of the manufacturer's failure to produce a safe vehicle, the vehicle crashed, causing damage to the vehicle and injury to the UiC. The UiC claims for injury and loss of earnings caused by the crash. **Claim by UiC against Insurer** The UiC was operating a vehicle which at all material times was covered by a contract of insurance to cover injury and damage in the evident of crash. The AV crashed while driving itself. The UiC suffered an injury as a result of the crash and seeks damages in accordance with the Automated and Electric Vehicles Act 2018.
Counterclaim by Manufacturer and Defence: The manufacturer denies that the AV was driving itself at the time of the crash.

In the alternative, the manufacturer states that the collision occurred when the vehicle was being operated outside of the vehicle's ODD. The manufacturer denies that it was negligent when producing the vehicle and states that the crash was caused by the negligence of the UiC. **In the alternative**, the manufacturer denies that it was negligent when producing the vehicle and states that if it was found that the AV was driving itself, the rain and overcast conditions obscured the bollard and the road signs and that even a careful and competent driver would not have avoided the crash. The UiC failed in their duty of care and contributed to the crash by operating the vehicle outside of its ODD by allowing the AV to operate in roadworks. The ODD is defined by the operating manual and to which the UiC acknowledged on the in-vehicle interactive screen when turning on the vehicle. But for the negligence of the UiC, the crash would not have occurred, and the manufacturer/insurer claims that in the event the AV was driving at the time of the incident, its liability for the crash should be reduced by 80% In the further alternative, the manufacturer states that any damage or injury caused by the crash was unavoidable due to the weather conditions, and all damages are payable by the insurer in accordance with the AEVA. **Defence by Insurer** The Insurer admits that the AV was covered by a contract of insurance at the time of the incident. The Insurer neither admits nor denies that the AV crashed while driving itself. In the event that the AV was driving itself, the Insurer denies liability under the contract of insurance because the UiC negligently or wilfully operated the vehicle outside of its ODD and caused the crash. The Insurer states that the UiC was negligent in operating the vehicle in roadworks and that this reduces liability under the contract of insurance to 20% of the total claim. **In addition or in the alternative,** if Court finds that the manufacturer was negligent in producing a vehicle which does not meeting the safety standard, then the Insurer claims 100% contribution from the manufacturer for any liability it pays to the UiC.

Table 2: Example pleading setting out the claims of the parties

As the insurer and manufacturer have denied the claim, the UiC must institute court proceedings. The UiC claimant has been left to prove every part of the claim, even that the AV was driving itself at the time of the crash. This type of complex claim is difficult to resolve without expert legal and technical advice and a mechanism to formalise that resolution. It is likely this scenario would involve civil court proceedings.

We have set out the potential questions which might be addressed in court proceedings:

Potential Questions of fact to be determined by the Court	**Was the AV driving itself at the time of the crash/?** **Did the AV issue any transition warning to the UiC?** **Did the AV operate as safely as a careful and competent driver?** Would a careful and competent driver have seen the bollard in the rainy and overcast conditions? Would a careful and competent driver have been able to avoid the collision with the bollard? Did the AV recognise the bollard, what relevant decisions were made by the AV about the vehicle, and when

	Did the AV recognise the roadworks signs, what relevant decisions were made about the road-signs or roadworks and when? **Was the approach to the roadworks within the AV's ODD?** What is the legal status of the manufacturer's warning in the manual and in-vehicle notification that the AV should not be operated in roadworks?
Potential relevant Evidence addressing the issues above	EDR and DSSAD data Video of the vehicle's approach to roadworks Video of other vehicles approach to the roadworks Data related to perception, planning and control[33] Whether the UiC was aware of roadworks on the route The AV manual and information communicated to the UiC on the in-vehicle screen before the journey commenced.

Table 3: Potential questions of fact to be considered by the Court and relevant evidence

The main claim of the UiC is that the AV failed to drive at the requisite standard, as the AV crashed into the bollard/failed to avoid the roadworks. However, both the insurer and the manufacturer have denied that the AV was driving itself and then argued in the alternative that if the AV was driving itself, the crash would have occurred 'outside of the vehicle's ODD'. They do not deny that the crash occurred; they only state what caused the crash. We will start with this first.

Was the AV driving itself at the time of the crash? Did the AV issue any transition warning to the UiC?

The disclosure of EDR and DSSAD data should resolve the issue of whether the AV was driving itself and or whether any transition warnings were given. However, just to obtain this data, the UiC may be forced to engage lawyers and request the disclosure. If, for any reason, the data recorded by the EDR and DSSAD does not reflect the UiC account of the crash (i.e., that the AV was driving itself), then the UiC would have to engage an expert to challenge/examine the data and also seek the disclosure of additional data to corroborate their account. If the UiC cannot prove that the AV was driving itself at the time of the crash, the UiC's claim will fail. We will assume, in line with the learning scenario, that the UiC was able to obtain enough data from the EDR and the DSSAD to prove that the AV was driving itself at the time of the crash and that no warning or transition signals were given to the UiC.

Was the approach to the roadworks within the AV's ODD?

As set out above, a manufacturer is able to specify the way in which their product or device should be used to avoid injury or damage. The AV in question was not designed for operating in roadworks which was stated in the manual and the in-vehicle HMI.

In the event the UiC is able to prove that the AV was driving itself at the time of the crash, we will assume from the learning scenario that the UiC had legitimately engaged the AV system at the beginning of the journey in accordance with the ODD and the manual, and was permitted to take their eyes from the road and hands from the wheel to read a book. When the vehicle is driving itself, UN Regulation 157 and the Automated Vehicles Act (2024) permit a UiC to remove their hands from the wheel and their eyes from the road. In these circumstances, the UiC was entitled to expect to be alerted when the AV was about to leave its ODD when it approached the roadworks by way of haptic, audio and visual signals, in accordance with UN Regulation 157 and the Automated Vehicles Act 2024. It is foreseeable that an AV which starts out on a journey within its ODD may encounter a situation that exceeds its ODD and that if it does not hand over operational responsibility to the UiC, the AV will crash and cause injury or damage.

---

[33] Malik et all (2022) 'How to autonomous vehicles decide' Sensors Vol 23 Issue 1 <https://www.mdpi.com/1424-8220/23/1/317>

In these circumstances, it is unreasonable to expect a UiC to look out for danger, including roadworks. The learning scenario states that no warning signals were given. As a matter of law, we would argue that the manufacturer cannot rely on an exclusion clause omitting roadworks from the ODD where the AV did not warn the UiC that the vehicle was no longer in its ODD. We consider that it is likely the Court would make a determination that the manufacturer cannot rely on the exclusion clause about roadworks, as the AV was driving itself, and it did not warn the UiC of the roadworks or transition operation of the driving to the UiC. Where the Court has made a finding like this, the insurer should reverse its decision and accept the claim.

Did the AV operate like a careful and competent driver?

This is a question of fact to be determined by a Court. If the Court has determined that the AV was driving itself, then the relevance of this part is that if the manufacturer was negligent in producing the vehicle, the insurer would be able to claim a contribution from the manufacturer for what it paid to the UiC.

In order to determine the answer to this question, the Court must consider whether a careful and competent human driver would have avoided the collision (which occurred on a rainy and overcast day). Therefore, data and documents must be disclosed to reveal the cause of the crash.

A lawyer acting for the UiC could request disclosure of documents/data (which includes computer records) of all AV recording devices.

Data which has been stored in accordance with EDR and DSSAD, such as speed, steering inputs, whether the automated system was engaged, whether airbags were deployed, and whether the vehicle braked i.e., the compulsory data parameters should be recorded and stored by the manufacturer. In addition to these parameters, the manufacturer may also have in its possession video and audio data, location data, seat occupancy data, eye movement data, i.e., the additional data parameters. Lawyers acting for the UiC, if they were well-funded, may seek to access both the compulsory data parameters and the additional data parameters. While some of the compulsory data parameters may provide useful evidence, these will not be sufficient to answer all questions of fact before the court described in Table 3. In addition, if well-funded, lawyers for the UiC may subpoena video footage from surrounding vehicles and digital infrastructure to inspect the collision, as well as seek eye-witness statements to obtain as much information as possible. This would be a time consuming and costly exercise.

The additional data parameters (in particular video and location data) could shed further light on how the accident occurred, in particular: how heavy it was raining, the visibility of the road/surroundings and the roadworks signs or otherwise evidence of roadworks. However, access to this data may be resisted by the manufacturer. Disclosure may be resisted on the basis that it is a third party (such as a developer) that has control of the data, that the time and cost of producing the data is disproportionate, or that the data has been destroyed or lost (as it is not compulsory to store additional data).

As the UiC has alleged that the vehicle did not drive like a careful and competent driver, the lawyers for the UiC may also request documents/data relating to AI decision-making in the lead up to the collision (i.e., data relating to perception, planning, and control). The IEEE Standard on Transparency suggests that, at a minimum, manufacturers should be logging periodic connection strengths, providing a mechanism for visualising the regions of input which have been prioritised when making decisions/outputs. However, the way in which this type of data is collected, interpreted and presented may vary vastly among manufacturers.

To assist with this, it is likely that a lawyer acting for the manufacturer may call an expert who is knowledgeable about the AI system and the method of XAI to give evidence about the meaning of the data/information produced by the XAI. In an adversarial system, it is to be expected that a witness from a party to the dispute is likely to give evidence to support that party. It is the role of the opposing party to test that evidence with their own expert.

Again, this would be a costly and time-consuming exercise. Every vehicle and AI system is different, and consequently we do not know what or how much information would be provided or what format it would take. If the AI system in question is unique or new, it may be difficult or impossible for the lawyer for the UiC to engage a suitably qualified and competent expert witness to challenge the manufacturer’s witness. In the event the disclosure documents provide a Post-hoc explanation of the AI (XAI) which indicates (for example) that due to rain interference, the roadworks was detected too late to respond effectively, it may be very difficult to interrogate this explanation on the information/data provided, with a third-party expert not familiar with that AI system, or at all. If the Claimant was unable to pay for their own expert to challenge the post-hoc explanation, there is a greater chance this evidence may be accepted by the Court.

As we have stated, during a hearing, the rules of evidence allow the Court to examine all relevant evidence, but that more reliable evidence will be given more weight. The Court's opinion about the evidence presented and explanations given by experts will affect the weight. Due to the rebuttable presumption that calculations made by a computer are correct, if the UiC wishes to challenge the XAI or any of the data which tends to contradict their version of events, the Court will require convincing evidence, expertise and argument.

Where Courts are provided with education and explanation about the usefulness of XAI, it may be that Courts will not make decisions based solely on explanations given by these models. If all data is made available, ie compulsory recorded data, additional recorded data, and any data that may be subpoenaed from third parties such as other vehicles, digital road infrastructure devices and eye-witnesses, it is hopeful that the true course of events can be made out. However, to do this will require enormous expense and time. A UiC is less likely to have resources to accomplish this.

If the Court decides, as a matter of fact, based on the data and evidence discussed above, that the AV did not operate as safely as a careful and competent driver, then it may be found that the manufacturer was at fault by failing in their duty of care to produce a vehicle to the required safety standard. Consequently, this would mean that while the insurer would have to pay the claim of the UiC, the insurer could make a claim against the manufacturer for any financial loss.

Potential Alternatives

A potential alternative framework may be for the safety standard to be raised to a level where the AV is expected to avoid all collisions, and otherwise, to create a rebuttable presumption that where an AV is involved in a crash, the AV is responsible. An AV may be involved in a crash despite operating at the required safety standard (ie the crash was unavoidable). However, the advantage the manufacturer has in this scenario is that they will be able to prove this, as they will already have access to the data.

Under the current framework, there is an obvious disparity or imbalance of power between Claimants alleging that AI caused harm and parties which own or control the data and or source code for the AI. The fact that this does and will continue to exist has caused legal commentators not only in the UK but internationally to propose alternative approaches, such as reversing the burden of proof be reversed, or to introduce a strict-liability scheme, where the Claimant still has to prove that damage occurred, but does not have to prove that negligence caused the damage[34]. However, these types of schemes are sometimes avoided by national governments as 'anti-innovation' as this will mean that in the first instance, AVs and their insurers will be obligated to pay for any damage or loss when an AV is involved in a crash, and this will include crashes that may have not been the fault of the AV. Potential liability and costs associated with such a scheme may cause AV manufacturers to choose not to launch shared-operational responsibility vehicles in the UK.

Conclusion

The above learning scenario demonstrates that although the AEVA purports to simplify and streamline claims relating to AV collisions so that those injured can obtain compensation quickly, there will often be scenarios where complex and expensive litigation is required. Manufacturers will specify via the HMI and the vehicle manual about circumstances where it is unsafe to operate the AV in self-driving mode and otherwise specify legal responsibilities of the UiC. Where the UiC fails to adhere to the warnings given via HMI and the manual, this may result in an insurance claim being denied. When an insurance claim is denied or party denied, then parties must turn to the courts to resolve the matter. Litigation is expensive, time consuming and places a large burden upon Claimants. In our learning scenario, the UiC's claim that the AV failed to operate at the safety standard expected, did not end at the crash and an insurance claim. The UiC's initial claim, followed by the allegations regarding the UiC's negligence in the counter-claim, required a dissection of data relating to who/what was in control of the vehicle and potentially data related to AV planning, perception and control. The compulsory data parameters which may be provided under

[34] Buiten M, A. de Streel, M Peitz (2023) 'The law and economics of Ai liability' Computer Law and Security Review 48, 105794 < https://www.sciencedirect.com/science/article/pii/S026736492300005

ERD and DSSAD may not be enough to explain how an accident occurred or who was at fault, and currently manufacturers are not obligated to record or store any additional parameters (such as video) which would be useful in an investigation. Where a Claimant is arguing that an AV did not drive at the required safety standard, and the compulsory data does not reveal how the incident occurred, XAI may provide the means to show how the AI operated at the time of the crash including why particular actions or decisions were made. However, XAI for complex artificial neural networks may not be interpretable to a standard that a Court could rely. XAI may not provide a loyal or transparent answer for what occurred, as XAI is only a tool to provide an explanation which occurred, and to be useful in Court proceedings, the explanation will require extrapolation by an expert. There is no guaranteed availability of another suitably qualified expert in that system for an opposing party to challenge any weaknesses in the explanation. In our learning scenario, the vehicle data ultimately helped the UiC to prove that the AV was driving itself and no warnings were given. However, in the face of an initial denial of the claim, the UiC was forced to institute Court proceedings to prove this. The learning scenario also demonstrated how difficult it may be to determine whether the AV operated at the standard of 'as safely as a careful and competent driver'.

If instead, the safety standard of operation for AVs required an AV driving itself, not to crash or be involved in a collision at all, then this would place the burden on the manufacturer to prove that a collision was unavoidable or not their fault. A manufacturer would have ready access to data with which to demonstrate this to an insurer. The current framework requires that the road user or UiC Claimant, attempt to prove their claim against an AV manufacturer without ready access to data.

Ultimately, as these cases are yet to be litigated in the UK, it remains to be seen what Courts will make of the standard of safety expected of an AV, how reliable they will consider XAI. This will be determined by the level of explanation given by counsel and experts, and general education of the judiciary on AI.

Jo-Ann Pattinson, Research and Impact Fellow, Institute for Transport Studies, University of Leeds. Jo-Ann is a specialist in the regulation of new technology and emerging transport trends, decarbonisation and expert in legal issues in automated and autonomous systems. Jo-Ann is a UK and Australian qualified solicitor and socio-legal researcher.