Final draft

Regulating the safety of autonomous vehicles using artificial intelligence

Roger Kemp, Lancaster University 21 November 2018

1. Introduction

Many politicians are enthusiastic for the introduction of autonomous vehicles. Proponents claim they will all but eliminate road traffic accidents and provide mobility to those unable to drive; an economic consultancy claims they would increase the UK's GDP by £50bn per annum.[1] However, few people seem to have analysed how risks will be managed and regulated.

The term "autonomous vehicles" (AV) can be used for a variety of different levels of automation: **Level 1** covers functions such as adaptive cruise control or parallel parking. A **Level 2** system can manage steady state driving on a clearly defined route, such as a motorway, or inching forward in a traffic jam. The driver is still in charge, monitors what's going on and has to be able to resume control at a moment's notice. In the next stage of automation, **Level 3,** the driver can leave driving to the vehicle software, but has to be ready to take over, at short notice, if the computer decides it cannot cope. In **Level 4** the system can manage driving in known conditions, such as in a clearly delineated urban area. Finally, in **Level 5**, the vehicle can undertake end-to-end driving anywhere and under all conditions. It is only really levels 4 and 5 that count as truly autonomous.

2. Road traffic regulations

In most European countries, including the UK, the regulations for road traffic are based on the 1968 Vienna Convention. Article 8 requires:

1. Every moving vehicle or combination of vehicles shall have a driver.
2. [not relevant – concerns animals]
3. Every driver shall possess the necessary physical and mental ability and be in a fit physical and mental condition to drive.
4. Every driver of a power-driven vehicle (except accompanied learner drivers) shall possess the knowledge and skill necessary for driving the vehicle.
5. Every driver shall at all times be able to control his vehicle or to guide his animals.

Annex 5 covers *Technical Conditions Concerning Motor Vehicles and Trailers.* This lists various requirements for vehicles including braking and lighting. The Convention is written on the basis that the vehicle has to meet defined standards and that safety in operation is the responsibility of the driver.

Autonomous vehicles of Level 3 and below will have a driver compliant with Article 8. AVs of Levels 4 and 5 do not have a driver and thus there is a question over who carries the responsibility that would otherwise be the driver's.[2]

3. Safety regulation of electronically-controlled transport systems

In the UK, all hazardous activities are regulated. For example, before a nuclear power plant can be commissioned, the Office for Nuclear Regulation assesses the licensees safety cases to ensure that

[1] *Connected and Autonomous Vehicles – The UK Economic Opportunity*, kpmg.co.uk, March 2015, with introduction by Mike Hawes, Chief Executive, Society of Motor Manufacturers and Traders (SMMT)

[2] Kemp R J: *Autonomous vehicles – who will be liable for accidents?* Digital Evidence and Electronic Signature Law Review, 15 (2018) Pages 33 - 47

the hazards have been understood and are properly controlled. During operation, they check that licensees comply with their license conditions through planned inspections. Most industrial sites have to comply with *The Health and Safety at Work Act 1974*, which is the primary legislation covering occupational health and safety in Great Britain. The Health and Safety Executive, with other enforcing authorities, is responsible for enforcing the Act and a number of other Acts and Statutory Instruments relevant to the working environment. Although framed as legislation referring to risks at work, the scope has been extended to cover many other types of business and activities (including educating students in universities).

Different transport modes have different safety regulatory regimes. UK railways, tramways, maglev systems and similar are required to comply with *The Railways and Other Guided Transport Systems (Safety) Regulations 2006* (ROGS); air travel in Europe has to comply with the European Aviation Safety Agency (EASA) regulations, and so on.

An autonomous vehicle of Level 4 or 5 does not have a driver. In terms of the control system and human involvement, it is much closer to an automatically controlled tramcar or magnetically-levitated people mover than to a conventional road vehicle. UK regulations for guided vehicles, such as people movers, are very different to those for road vehicles but there is no logical reason why an automated road-based transport system should be treated differently to an automated guided transport system or to any other computer-controlled system; the risk profile is closer to a guided system than to a manually-driven car.

The risks created by widespread adoption of autonomous road vehicles would be at least as great as those intrinsic to the railway network. In the latter case, trains are generally constrained by the infrastructure to stay within the railway boundary and so the main risk is of one train running into another. This is a risk that has been understood for many years and the mechanical structure of trains is designed to stringent safety standards, such as resisting a buffer load of 200 tonnes without damage. The infrastructure is also designed to trap out-of-control trains with catch-points and similar arrangements. Underground railways have readily accessible emergency buttons on platforms that allow passengers or staff to stop trains.

By contrast, there is a negligible physical barrier between a computer-driven car and pavements, pedestrian areas and similar places and there is no easy way of bystanders witnessing a problem developing doing anything about it.[3] A fault that caused a car on a city street to accelerate uncontrollably would be likely to cause serious injuries and, possibly, multiple fatalities. Connected autonomous vehicles (CAVs), such as fleets of HGVs running in convoy on motorways, represent a potential hazard considerably greater than can be found on most rail networks and they could cause a major pile-up with consequences at least as serious as the 1999 Ladbroke Grove rail crash.[4]

The acceptance processes for new vehicles on UK rail networks are more stringent than in most other European countries. The process varies, depending on the network but all generally require a proof-of-safety underwritten by an independent body to be submitted to a regulator. For a straightforward manually-driven multiple unit train this can run to several large volumes with filing cabinets full of supporting documents. Computer systems that are considered “safety critical” (i.e. where a fault could cause an accident) are treated with caution and suppliers are expected to provide a safety justification including, where relevant, validation of the software.

The cost of producing the proof-of-safety for the electrical and electronic systems of a train can run into £ millions and the process can take more than a year. But a train is several orders of magnitude[5]

[3] A readily accessible means of stopping a car from the outside would offer significant opportunities for criminal activity.

[4] The report into the accident can be found at http://www.railwaysarchive.co.uk/docsummary.php?docID=38

[5] In engineering, an “order of magnitude” refers to a factor of 10. So one order of magnitude represents 10 times, two orders of magnitude 100 times, and so on.

simpler than an autonomous vehicle and the cost and time of producing a proof-of-safety for the latter would be proportionately greater.

All (fully or partially) automated transport systems have a safety regulator. The Office of Rail and Road (ORR) is the independent safety and economic regulator for Britain's railways. It also monitors the road network but does not have a lead role in vehicle safety. The Civil Aviation Authority is the UK's aviation regulator responsible for ensuring, inter alia, that the aviation industry meets the highest safety standards. It is a public corporation, established by Parliament in 1972 as an independent specialist aviation regulator. Most other automatic transport systems, such as fairground rides, lifts, moving walkways and escalators, are regulated by the Health and Safety Executive.

In the EU, road vehicles are subject to type approval for them to be allowed to run throughout Member States. (Presumably some equivalent arrangement will be put in place after Brexit.) Type approval describes the process applied by national authorities to certify that a model of a vehicle meets all EU safety, environmental and conformity of production requirements before authorising it to be placed on the EU market. The manufacturer makes available a dozen or more pre-production cars which are tested and, if all relevant requirements are met, the national authority delivers an EU vehicle type approval to the manufacturer authorising the sale of the vehicle type in the EU.

The process of type approval is only required to demonstrate that a vehicle meets technical requirements, largely based on Annex 5 of the Vienna Convention, discussed earlier. For an autonomous vehicle, there is no equivalent approval process or safety regulator to approve that the electronic equivalent of the driver meets appropriate standards.

The approval process for road vehicles is largely based on testing prototypes. This is a logical approach for their mechanical systems. However, it has been known for half a century that testing is inappropriate for validating safety-critical software. In 1969, Professor Sir Tony Hoare wrote *"One can construct convincing proofs quite readily of the ultimate futility of exhaustive testing of a program and even of testing by sampling."* In the same year, Edsger Dijkstra wrote *"Testing shows the presence, not the absence of bugs."* [6] Any regulatory authority for autonomous vehicles would need to approach the task in the same analytical manner that the ORR or CAA approach approval of software systems for rail or aviation, but for a systems that is perhaps two orders of magnitude more complicated than either a computer-based signalling system or an autopilot. This would be a large professional organisation. (For comparison, the CAA has 930 employees and the ORR 290.) So far, the UK government has taken a decision not to start setting-up a regulator for autonomous vehicles.[7]

4. Interaction of AVs with real people

Much of the foregoing argument is about the safety of AV systems by themselves. However, they will need to interact with human drivers and those interactions could cause accidents. Many stand-up comics have talked about a car driver who has never had an accident but claims to have witnessed them regularly. Could an AV also suffer no crashes itself but the cause of any number?

There will be some vehicles that are never driven by computer. Police armed-response vehicles, fire engines, the Prime Minster's motorcade and local contractors driving from job to job spring to mind but there will be many more. In addition, the penetration of AVs into society will, inevitably, be slow.

[6] Both cited by Professor Martyn Thomas in a lecture at Gresham College.

[7] Letter from Baroness Sugg, Minister in the Department for Transport, dated 7 March 2018 to a House of Lords Committee.

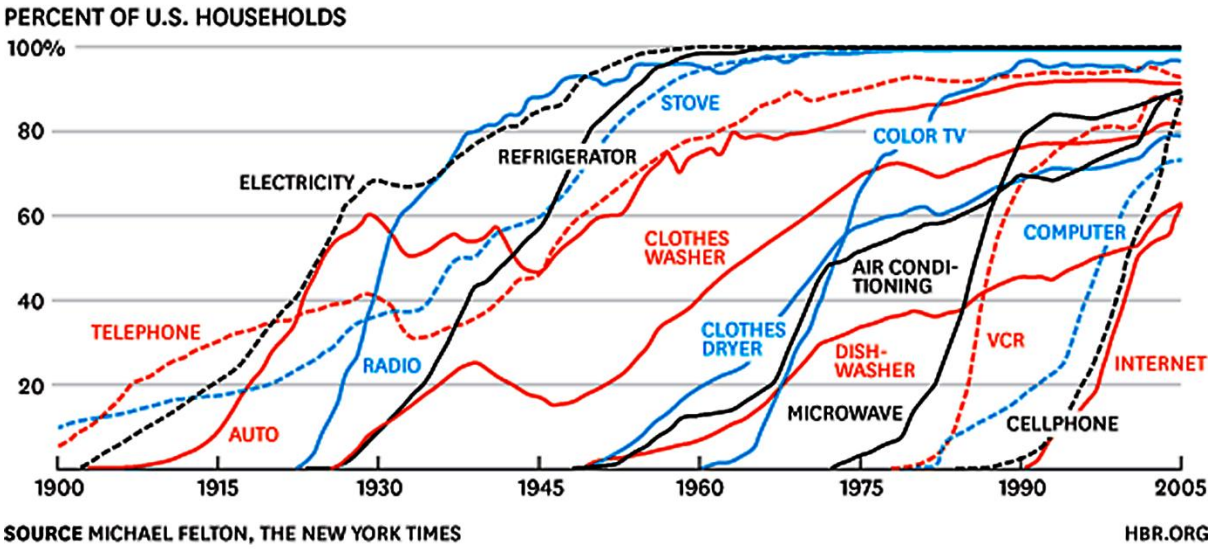

Figure 1: Take-up of new technology

Figure 1 shows the take-up of new technology in the USA. A few technologies, particularly those that are reasonably cheap and provide a service that didn't exist previously, such as cell phones or video recorders, were taken-up in less than 15 years. Others, such as washing machines or the telephone, took more than 70 years to be widely adopted.

Autonomous vehicles, if accepted, are likely to become available progressively across the country, as different areas are mapped in detail. Some sectors of society will be resistant to their introduction and they will be expensive; thus they are unlikely to achieve widespread adoption for several decades from their introduction. Taking account of the number of vehicles that are never likely to be driven by computer and the probable slow take-up of the technology, AVs will have to be able to operate in an environment in which most other vehicles are driven manually.

5. Coexistence between human and computer-driven vehicles

When a human driver is undertaking a manoeuvre, they undertake frequent risk assessments (although they probably don't think of them as such). One of the first mental activities a driver takes when approaching a roundabout, turning out of a side road or preparing to join the traffic on a motorway is to update their situational awareness.

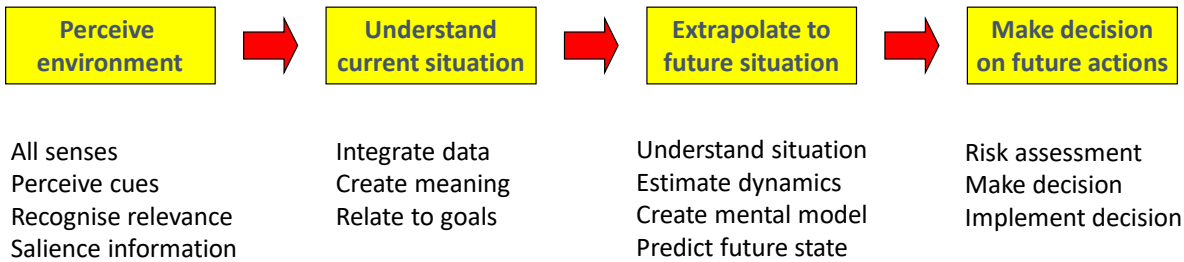

Figure 2: Establishing situational awareness

Firstly they perceive their environment – primarily visual evidence but often supported by audible evidence (e.g. children shouting to each other, horns, sirens, squealing tyres, the sound of revving engines) and possibly even olfactory evidence (such as the smell of burnt rubber). Secondly their brain integrates these inputs to form a meaningful mental image of the current situation. Thirdly,

they make a mental extrapolation of the situation to provide a scenario of what is likely to happen – whether a child is about to run into the road; whether a car will actually turn left, as it is indicating, or continue on the main road, and so on. Finally, they take a decision.

With experience, a driver learns to read the signals of what other road users are likely to do, which allows them to make an informed decision, for example, whether to accelerate into motorway traffic or hang back and ease in behind another vehicle in the nearside lane. Other drivers, observing the manoeuvre, may modify their own strategy, either empathetically or aggressively. Human drivers become good at assessing what other humans are likely to do: they may be less good at assessing what a computer will do. Drivers are also aware that driving styles vary regionally and modify their expectations, so the aggressive style one might anticipate on Birmingham's inner ring road at rush hour is not what one would expect on Sunday afternoons on the Mull of Kintyre; a computer might not be so flexible.

Almost certainly, AVs will be programmed to comply strictly with the law. If sensors are well-developed, they are unlikely to misinterpret road signals or hit other vehicles. But human drivers might misinterpret what the AV is planning to do, resulting in an accident. Thus, although the safety record for computer-driven cars may be good, they could be vicariously responsible for far more accidents than officially recorded.

It is not obvious how any regulatory system could assess a computer's empathy or consideration for other road users, as opposed to strict compliance with regulations, but this is a characteristic assessed, if only informally, in the driving test for their human equivalents. Under UK law, you are guilty of a careless driving offence if you drive a mechanically propelled vehicle on a road or other public place without due care and attention, or *without reasonable consideration for other persons using the road or place*. It is interesting to surmise what the equivalent regulation for an AV might be.

6. The concept of acceptable risk

All motorists take risks. As discussed in the previous section, a car driver approaching a busy roundabout will observe the other traffic and will undertake a risk analysis to decide when to join the traffic flow. A driver who minimises risk by waiting until there are no other vehicles on the roundabout, is likely to be there a long time and to seriously inconvenience other road users.

Motorists also make a risk assessment when deciding how far to drive behind the car in front. On European railways, the signalling system is designed to ensure that, if a train stops suddenly (for example by running into an immovable object), the following train will be able to brake safely even assuming worst-case adhesion conditions. Were car drivers to use equivalent criteria, they would leave more than 100 metres between cars travelling at 120 km/h (thus reducing the capacity of motorways to a quarter of present traffic flows). But most drivers take the risk that the car ahead of them will not stop dead and that the adhesion between their tyres and the road will not suddenly drop to a value equivalent to spilt diesel on black ice. Such risk assessments, even if they turn out to be unfounded, are generally accepted by regulatory authorities and opinion-formers.

Safety-critical automatic control systems are designed to eliminate risk. The servomechanisms to lower control rods into a nuclear reactor (and thus reduce its output) are designed to be able to fulfil their task under the worst conceivable combination of circumstances. Avionic and rail control systems are typically designed with wrong-side failure rates of less than 1 event in 10^9 hours (more than 100,000 years). Designing safety-critical systems to take risks is a novel area of engineering, which will require a new approach both to design and to regulation.

7. Road safety and public acceptance

The number of fatal road accidents in the UK has dropped significantly over the past decade from more than 3,000 in 2006 to around 1,800 in 2016. Government statistics[8] show that more than 80% of fatal and serious accidents are attributable to the driver, not to the vehicle or the infrastructure (Figure 3).

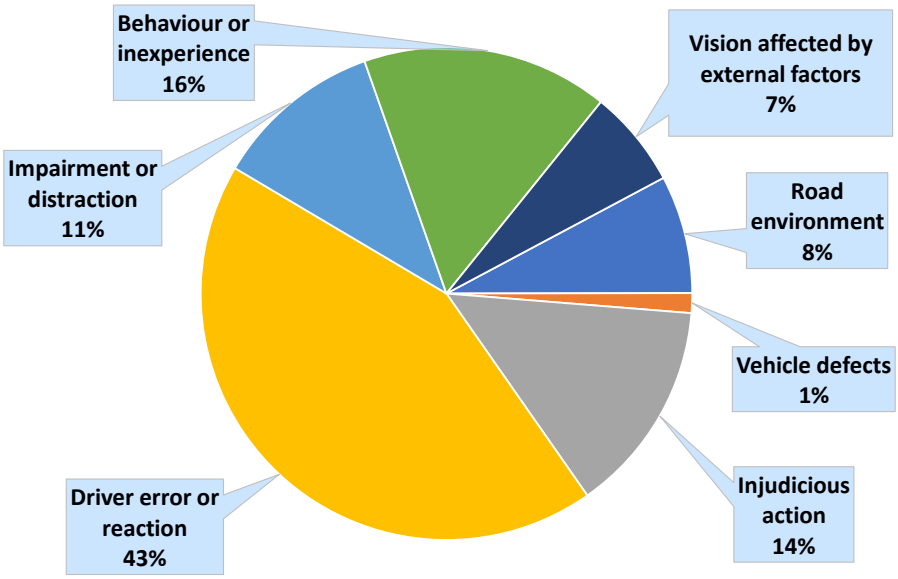

Figure 3: Causes of fatal and serious road accidents in the UK [9]

Breaking down the data, some of the more common causes are:

- Driver careless, reckless or in a hurry
- Driver failed to judge movements of other vehicles
- Driver failed to look properly
- Following too close
- Loss of control
- Poor turn or manoeuvre
- Sudden braking
- Travelling too fast for conditions

Data on manually driven cars shows a safety performance of just over 1 fatality per billion km.[10] Research published by the Rand Corporation concludes:[11]

> 'The results show that autonomous vehicles would have to be driven hundreds of millions of miles and sometimes hundreds of *billions* of miles [under realistic traffic conditions] to demonstrate their reliability in terms of fatalities and injuries. Under even aggressive testing assumptions, existing fleets would take tens and sometimes hundreds of years to drive these

[8] https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-annual-report-2016

[9] The diagram shows contributory causes, as recorded by a police investigation. It covers only those accidents in which the police were involved and the recorded causes may be subjective, rather than the result of a detailed analysis.

[10] Professor Martyn Thomas, Is Society Ready for Driverless cars?, Lecture at Gresham College, 24 October 2017.

[11] Nidhi Kalra and Susan M. Paddock, How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability? The Rand Corporation, 2016.

> miles – an impossible proposition if the aim is to demonstrate their performance prior to releasing them on the roads.' [Italics in the original.]

If it is not practicable to demonstrate the safety of AV systems by testing, can safety be proved by independent analysis of the code? This is discussed in the following section.

8. Validation of safety-critical software using artificial intelligence (AI)

Regulating the safety of autonomous vehicles will be particularly challenging as most use artificial intelligence (AI). Most validation of safety-critical software-based systems is based on the assumption that the software validated is identical to that which will be put into service. The process used to validate avionics software requires that tests are generated from a formal requirement specification and are shown to exercise every decision point in the software in all logically possible outcomes. Hayhurst and Veerhusen[12] describe the process:

> *"For systems that are safety and mission critical, extensive testing is required. However, the size and complexity of today's avionics products prohibit exhaustive[13] testing. … …*
>
> *"For level A software (that is, software whose anomalous behavior could have catastrophic consequences), DO-178B [14] requires that testing achieve modified condition/decision coverage (MC/DC) of the software structure. MC/DC is a structural coverage measure consisting of four criteria mostly concerned with exercising Boolean logic. The MC/DC criteria were developed to provide many of the benefits of exhaustive testing of Boolean expressions without requiring exhaustive testing."*

Chilenski and Miller[15] provide a more detailed summary of the tests normally undertaken on high-integrity avionics software systems:

> ***Statement coverage (SC)*** *– every statement in the program has been executed at least once.*
>
> ***Decision coverage (DC)*** *– every point of entry and exit in the program has been invoked at least once, and every decision in the program has taken all possible outcomes at least once.*
>
> ***Condition / decision coverage (C/DC)*** *– every point of entry and exit in the program has been invoked at least once, every condition in a decision in the program has taken all possible outcomes at least once, and every decision in the program has taken all possible outcomes at least once.*
>
> ***Modified condition / decision coverage (MC/DC)*** *– every point of entry and exit in the program has been invoked at least once, every condition in a decision in the program has taken on all possible outcomes at least once, and each condition has been shown to independently affect the decision's outcome. A condition is shown to independently affect a decision's outcome by varying just that condition while holding static all other possible conditions.*

---

[12] *A practical approach to Modified Condition/Decision Coverage*, Kelly J. Hayhurst, NASA Langley Research Center, Hampton, Virginia and Dan S. Veerhusen, Rockwell Collins, Inc., Cedar Rapids, Iowa, NASA 2001

[13] Exhaustive, in this context, can be taken to include checking every path through the software in all possible combinations.

[14] RTCA/DO-178B*, Software Considerations in Airborne Systems and Equipment Certification*, RTCA, Inc., Washington, D. C. December 1992. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20040086014.pdf In January 2012, DO-178B was replaced by the updated DO-178C. The text of the latter is identical to document ED-12C *Software considerations in airborne systems and equipment certification*, issued by the European Organisation for Civil Aviation Equipment (EUROCAE) which is used by the European Aviation Safety Agency (EASA) as means of compliance to European Technical Standard Orders (ETSOs).

[15] Chilenski, John Joseph, Steven. P. Miller, *Applicability of modified condition/decision coverage to software testing*, Software Engineering Journal, Vol. 7, No. 5, pp. 193-200. September 1994.

> ***Multiple-condition coverage (M-CC)*** *– every point of entry and exit in the program has been invoked at least once, and all possible combinations of the outcomes of the conditions within each decision haw been taken at least once.*
>
> ***Reach coverage*** *– at least one definition clear subpath from each definition to each reference for each object has been executed.*
>
> ***All-uses coverage*** *– at least one definition clear subpath from each definition to each reference and each successor of the reference for each object has been executed.*
>
> ***All-sub-paths coverage*** *– all definition clear subpaths from each definition to each reference and each successor of the reference for each object haw been executed.*

For control systems that can be defined in terms of Boolean logic, such as railway signalling interlocking or the conditions that have to be fulfilled before control rods are raised in a nuclear reactor, the above process is straightforward, if time-consuming. It also requires a comprehensive and validated functional specification that can be translated unambiguously into Boolean algebra.

However, many of the Artificial Intelligence (AI) functions on an AV are not so clear-cut. For example, a combination of video, lidar and radar inputs are fed into software that is required to identify that an object approaching a the vehicle is a cyclist, a dog-walker, a group of children, a kangaroo[16] or a newspaper blowing in the wind. That type of decision-making is not amenable to validation by rigorous analysis of logical expressions, such as:

> "**IF** *track-1* is occupied **AND** *track-2* is occupied, **THEN** *signal-6* shall be red."

The control systems of an AV will contain many functions that cannot be subject to strict logical analysis. Apart from the identification of people or things on or near the route of the vehicle, the AV will have to predict how each is likely to move, whether they are likely to conflict with the intended route of the AV and, if so what avoiding action should be taken. Decision-making becomes more complex when the AV is simultaneously tracking a dozen or more pedestrians, cars, trucks, cyclists or animals, all of which might alter their own behaviour in the light of actions of the AV or of the other road users.

Many of the prototype AVs currently being tested, include adaptive self-learning software so, for example, the vehicle learns what a whole variety of pedestrians look like to its sensors and how they move and respond. If a learning-mode continues into commercial operation, the vehicles in service could behave very differently to the samples validated at the pre-production stage, and it is likely that vehicles would develop different regional characteristics (an AV learning what pedestrians look like in Sweden in winter could develop a rather different picture to a similar vehicle learning in Italy in summer). A solution to this problem could be to allow prototypes to learn and then to freeze the software prior to the validation and service operation, but this has implications on the extent of supervised prototype operation in real environments, and it also trades-off future learning, which might improve safety, against difficulty of validation.

To date, in no industry has software as complex as that in an AV been validated for use in safety-critical applications. The usual processes, such as used in the avionics industry, outlined above, appear to be impracticable. A recent session of the House of Commons Science and Technology Committee[17] took evidence on the safety validation of AI software:

> *The EPSRC UK-RAS Network noted that the verification and validation of autonomous systems was "extremely challenging" since they were increasingly designed to learn, adapt and self-improve during their deployment. Innovate UK highlighted that "no clear*

[16] *Volvo admits its self-driving cars are confused by kangaroos.* Naaman Zhou, The Guardian 1 July 2017.

[17] House of Commons Science and Technology Committee. Robotics and artificial intelligence. Fifth Report of Session 2016-17 paragraph 40

paths exist for the verification and validation of autonomous systems whose behaviour changes with time" while Professor David Lane from Heriot-Watt University emphasised that "traditional methods of software verification cannot extend to these situations".

Safety validation to allow AI systems to be used in safety-critical systems, such as autonomous vehicles, is an active research field and it would be optimistic to make the assumption that an agreed methodology and organisational infrastructure will exist to allow an AV regulatory body to validate software systems using AI much before 2030, unless it is prepared to accept very different levels of risk to other transport systems regulators.

9. Functional system safety and what is acceptable

It is because neither exhaustive testing of complicated software systems nor independent analysis of the code is practicable, that the international standard for the functional system safety, IEC 61508,[18] relies heavily on a rigorous structured development process. However, this has the disadvantage that this has to be applied during the whole of the specification, design and development phases; it cannot be used retrospectively on an existing product, which would be the situation in the UK and most other European countries.

If a safety validation process could be established and automation eliminates all the accidents caused by human drivers, the number of fatalities on UK roads might be reduced from 1,800 to fewer than 400 p.a. This would be a great improvement. However, it is likely that, from time-to-time, the computer systems will malfunction or misinterpret data and thus will themselves cause accidents. If computer systems cause 500 fatal accidents p.a. the overall total will be half what it is today. But would the public accept 500 fatalities p.a. caused by computers? Or 50? Or even 5?

When a human driver makes an error, the attitude of the press, police, magistrates and opinion formers (most of whom drive cars) is likely to be: "He made a mistake – anyone could have done that." The same level of empathy is not extended to engineering systems. Searching the Driver and Vehicle Standards Agency (DVSA) database on vehicle recalls [19] produced 31 for the Ford Focus, one of most popular cars in the UK. Many of these had fairly minor consequences – "oil filler cap may become loose, wiper motor may fail and overheat" and so on.

As an example of the seriousness with which manufacturers, regulators and the press treat car safety concerns, Toyota is reported to be recalling around 1.3 million hybrid cars due to a wiring issue that could potentially cause a fire.[20] The fault has led to one incident in Japan which produced smoke from the vehicle. No injuries have been reported as a result but it was considered sufficiently serious by the British press to have been treated as a news item.

Against that background, it is inconceivable that accidents potentially caused by computer systems on autonomous vehicles would not be investigated. The question is how that would be done.

10. Accident investigation

All fatal road accidents are investigated. Often there is not much evidence to go on – some skid marks on the road, physical damage to vehicles and the infrastructure, breathalyser readings and

[18] IEC 61508 is the international standard for electrical, electronic and programmable electronic safety related systems https://www.iec.ch/functionalsafety/

[19] https://www.dft.gov.uk/vosa/apps/recalls/default.asp accessed 2/11/18

[20] https://www.express.co.uk/life-style/cars/1013377/Toyota-recall-2018-hybrid-Prius-CHR accessed 2/11/18. The report states "The wiring harness could come into contact with a cover and, if dust accumulates on the harness or the cover, the insulation on the wire could wear down over time. If this occurs, it could cause an electrical short circuit."

statements from the people involved. In some cases, reviewing the evidence takes only a few days as there is so little hard information.

The situation with computer-controlled vehicles could be very different. AVs are likely to have stored video feeds, radar and lidar information, speed and acceleration data, brake actuation and many other parameters. Given a few weeks of analysis and full cooperation from the manufacturers, a team of technically-literate accident investigators could form a well-documented picture of what happened.

It is straightforward, if time-consuming, to determine ***what*** happened, it is much more difficult to ascertain ***why***. The computer systems that control an AV take inputs from its stored maps, radar, video and other sensors to form a view of the world. They use this information to recognise the infrastructure and other road users, including pedestrians, cyclists, animals, cars and trucks, roadworks, children playing near the road, debris, litter and other obstructions. Having identified the other actors in that space, the computers have to predict what each is going to do, basically the same process as a human establishing situational awareness (Figure 2). Having understood the environment, the control systems then have to create a route that avoids other actors, does not subject the occupants to excessive accelerations and goes in the direction required by the journey objectives.

For an accident investigation team to understand ***why*** an AV carried out a certain manoeuvre requires a detailed knowledge of all aspects of the control systems, including software that may run to hundreds of millions of lines of code. This is a task that would take the designers many months of full-time effort. For an independent accident investigation team, such as those deployed by the UK's Air Accident Investigation Branch (AAIB)[21] or Rail Accident Investigation Branch (RAIB)[22], determining the cause of a single accident to an AV could take a large team of software specialists several years. It would also require full cooperation of the suppliers of the software systems and full access to design data – at present considered as confidential proprietary information.

It is likely that many accidents involving AVs will not be caused by the computer-driven car but by other road users. However, in many of these cases, accident investigators would need to analyse the computer systems on the AV to enable them to come to that conclusion. Thus the overall accident investigation load is likely to be significantly greater than might be expected from the number of accidents for which AV's computer systems are responsible.

In a recent interview with The Guardian,[23] Alison Saunders, the retiring head of the Crown Prosecution Service (CPS), said that Britain's criminal justice system is "creaking" and unable to cope with the huge amounts of data being generated by technology. She said the CPS and police were failing to investigate thousands of cases efficiently – from rape to fraud to modern slavery – and were critically short of the skills and resources required to combat crime.

Searching dating sites and email servers for evidence relating to an allegation of rape is time consuming but is not particularly complicated. Analysing the AI software in an autonomous vehicle is several orders of magnitude more challenging and requires a highly specialised competence that is in short supply (and is expensive to employ). Realistically, there is no possibility of thorough and independent investigation of all accidents involving AVs. Society would then be left with conclusions such as:

> *"The fatality appears to have been caused by faults in the AV software. We have not identified these faults and we have no plans to find a solution or to instigate a recall."*

[21] https://www.gov.uk/government/organisations/air-accidents-investigation-branch

[22] https://www.gov.uk/government/organisations/rail-accident-investigation-branch

[23] https://www.theguardian.com/law/2018/oct/27/cps-alison-saunders-justice-system-cannot-cope-resources accessed 2/11/18

Bearing in mind the sensitivity of the press and public towards technical hazards, it is unlikely such an approach would be readily accepted, but there appears to be no alternative, if AVs are to be introduced as politicians are planning.[24]

11. Conclusions

Superficially, computer-driven cars offer many advantages, particularly in eliminating many of the human errors that cause accidents, reducing the amount of city-centre space taken by parking and providing mobility for an ageing population. Despite these important benefits, there are also reasons why there is opposition to the widespread introduction of this technology, including further inroads into personal privacy by technology companies,[25] fragmentation of society, damage to the economics of public transport, increased CO_2 emissions and many new opportunities for cyberterrorism and criminality.

This paper has not covered the wider social implications of computer-driven vehicles. Instead, it has concentrated on how their introduction could be regulated. In the UK and most Member States of the EU, all potentially hazardous technologies are regulated by government agencies on behalf of society – transport systems are no exception. It is inconceivable that autonomous vehicles should be exempt.

In March 2018, the UK government said it had not established an agency to act as the regulator for autonomous vehicles and had no immediate plans to do so. The establishment of agencies and standards to regulate the safety of railway control systems and aircraft control systems took many years, and a few unsuccessful attempts.[26] The control systems on autonomous vehicles are far more complicated than either railway signalling or aircraft autopilot and flight control systems. Setting-up a regulatory structure and acceptance standards, particularly with the disruption of Brexit to the harmonisation of regulations with other countries, will need a large team with expert knowledge of safety critical systems and will take several years.

Unlike manually driven cars, it is impracticable to demonstrate the safety of autonomous vehicles by testing prototypes. To establish that a particular set of control systems achieve an equivalent safety performance to human drivers would involve a test fleet of scores of vehicles and could take decades – by which time the technology would be obsolete. And independent validation of safety by examining millions of lines of code, would be equally challenging. A rigorous design and development process, using validated tools and overseen by a professionally competent regulatory agency is the only practicable route but, because the UK government has no plans to establish such an agency in an appropriate timeframe, it is difficult to see how this could be implemented.

Most AVs are expected to use Artificial Intelligence (AI) which adapts system response in the light of experience. To-date, the regulation of safety-critical systems requires that any software offered for validation is identical to that which is installed on the system; self-learning adaptive systems, such AI, are thus formally excluded. The adoption of computer-driven vehicles will thus require a new approach to regulation of safety-critical software systems, requiring a change in philosophy that has, so far, been debated only in very general terms and is unlikely to result in a consensus for several years. AVs will also have to take risks; this will be novel for regulated systems using safety-critical software and will require significant philosophical debate between manufacturers, regulators and politicians.

---

[24] https://www.bbc.co.uk/news/business-42040856 Chancellor Philip Hammond told the BBC the objective was to have "fully driverless cars" without a safety attendant on board in use by 2021. "Some would say that's a bold move, but we have to embrace these technologies if we want the UK to lead the next industrial revolution". accessed 2/11/18

[25] Maria Cristina Gaeta; *GDPR and automotive: The issue of data protection in self-driving cars.* Paper delivered to conference *Transforming Cities with Artificial Intelligence: Law, Policy, and Ethics*, London, November 2018

[26] See, for example, http://www.railwaysarchive.co.uk/docsummary.php?docID=39

An important part of the safety regulation of transport systems is a thorough analysis of accidents and "near-misses", to learn from experience and correct any systemic errors found. Because of the complexity of their control systems, it is unrealistic to expect accident investigation to determine the root cause of all accidents in which the "thinking process" of AVs could be implicated.

Against this background, the timescales discussed by UK ministers for the introduction of AVs by 2021 seem unrealistic, unless they are prepared to adopt a radically different philosophy to the risk-based approach that has underpinned UK regulation in all areas since the 1974 Act.

Roger Kemp