Liability Issues in Autonomous and Semi-Autonomous systems

John Buyers, Partner, Osborne Clarke LLP

Abstract

Autonomous (or intelligent) machines present new challenges for our existing models of liability which are largely causative based. Where it is unclear why a machine has behaved in a particular manner due to its innate complexity or learned behaviours, how are we to determine attribution of "fault" or "defect" for liability purposes?
This paper examines those issues, together with a discussion of how such problems may be potentially resolved in the future.

Liability Issues in Autonomous and Semi-Autonomous systems

Introduction – AI – Expectations vs. Reality

The idea of the intelligent machine plays heavily on our collective psyche – after all the Domesday scenario of machines out of control or with killer instincts are a favourite Hollywood staple and make for great entertainment.

We shouldn't also forget that this fire has been stoked somewhat of late by the likes of Stephen Hawking who has professed that the development of full artificial intelligence could spell the end of the human race.[1]

Our yearning for thinking machines is peculiarly quixotic. On the one hand we expect such machines to work to infallible logic and surpass us in our reasoning and abilities. On the other we still expect them to be imbued with the essential characteristics of fallibility and humanity.

The reality is that we are still some considerable way off developing true artificial sentience and truly self-aware machines. There are however some serious liability consequences to be learned from a potential HAL 9000.[2]

AI Liability: A Sliding Scale

I present consequences, responsibility and hence liability as essentially scalable. At the one end (which is where the current state of the art sits), with AI systems as an assemblage of complex components using existing contract, tort and consumer protection principles to trace liability and at the other end as self-aware sentient thinking machines that accrue artificial personhood (in some form) and assume self-responsibility and hence liability for their actions.

It is at the latter end that we begin to skirt more speculatively from fact into speculation and science fiction but it is important to understand where our current liability models weaken and how potentially they could be extended to cover such systems. Ultimately, as it always does, the law will need to flex, to adapt and accommodate this new evolution in technology.

In legal terms we need to be clear what we mean by "*intelligent*". At its most basic level, we have machines that can respond and make limited pre-defined decisions on a limited basis in response to external stimuli and in accordance with programmed software parameters. This is where the current "*state of the art*" is at the moment.

This progresses to machines that have the capacity to learn and to make autonomous decisions that are not directly traceable to their programming.

So although we have machines that can exhibit some elements of autonomy, we are still some way off from the holy grail of creating a truly self-aware, sentient machine that is capable of spontaneous creative output and the ability to independently and freely communicate with us irrespective of our direct instructions to it.

To analogise on a rather more colloquial level, you can currently get into a Google Car which will (quite effectively) drive you from A to B and will avoid traffic collisions based on its programming, and inputs from its satellite navigation systems and radar, but you still cannot have a sensible conversation or argument with it.

Liability is essentially a sliding scale which is based factually on the degree of consequential legal responsibility society places on a ***person***. As we will see later on, historically responsibility and hence liability levels are not static – the able minded and children and mentally incapable adults have different levels of liability – the latter having little or no responsibility for their actions and therefore a commensurate degree of low accountability and liability.

Until relatively recently, the question of whether or not a machine should be accountable (and hence liable) for its actions was a relatively trite one – a machine was merely a tool of the person using or operating it. There was absolutely no question of a machine assuming a level of personal accountability or even "*personhood*" as they were incapable of autonomous or semi-autonomous action. This is essentially the way in which the law has evolved to deal with machine generated consequences.

We are now however poised to deal with how to creatively twist our liability frameworks to account for the consequences created by autonomous machine action.

Existing Liability Frameworks

So with this in mind, let's review the existing liability frameworks which could conceivably apply to what are termed "*machine generated consequences*".

Current liability frameworks are focused on product liability generally – broken down into three distinct

[1] See e.g. The Washington Post, 20th January 2016 "*Why Stephen Hawking believes the next 100 years may be humanity's toughest test*"

[2] The sentient supercomputer character from Stanley Kubrick's *2001: A Space Odyssey*

Liability Issues in Autonomous and Semi-Autonomous systems

categories: Contract law, Negligence (tort) and, so far as the UK (and EU) is concerned strict liability under consumer protection legislation (in the UK the Consumer Protection Act 1987).

We'll review the scope of each briefly in a moment and then consider potential issues in relation to their application to AI systems. Clearly the most conventional analysis we can apply to intelligent or semi-intelligent machines is as complex products.

Taking the driverless car as an example, what we can see is that in fact it is an assemblage of many and varied integrated systems that are produced by multiple manufacturers. For a driverless car to work effectively, it needs sensors to navigate road obstructions, such as radar and laser detection. It must have a computer to direct its actions and that computer needs to have a logic framework within which to operate – internally by use of its own operating software and also externally by reference to map data. All of these systems need to work together effectively, and this is without consideration of all the usual mechanical components which form a standard car, which must also be present and functioning.

This complexity gives rise to a potential plethora of liability targets, ranging from the vehicle manufacturer itself, all the way down to the designer of an individual component, depending upon where the actual defect, fault or breach occurs.

Existing causative liability models work well when machine functions (and hence responses) can by and large be traced back to human design, programming and knowledge. They begin to break down when this cannot be done.

Contract

Contract clearly has a role in determining product based liability. Contracts ensure that manufacturers and retailers sell products that meet contractually determined standards. Contract liability is obviously aimed at the recovery of financial (or pure economic) loss as a result of breach of these contractual standards, however obviously, as we are all aware, contract liability can in some circumstances also lead to the recovery of damages for consequential loss and/or damage.

Contract terms may be either express – as to defects and warranties or implied. In the UK there are implied terms as to quality, fitness for purpose, title and description in the Sale of Consumer Rights Act 2015 (for B2C contracts) and the Goods Act 1979 (for C2C and B2B contracts). Although there is not a focus on "*defects*" per se under the Sale of Goods legislation, there is clearly an emphasis on conformity with description. Arguably that could amount to nearly the same thing: a failure to conform to a description or specification is very close to a "*defect*" in practical terms.

We'll take a look at the relevant strengths and weaknesses of contract liability in this context in a moment but it is worth remembering that contract is a ***causative*** based liability framework. In order to found liability, the Claimant must prove that there was a breach of either an express or an implied term and that that breach caused the loss. Contract causation is often overlooked as the poor cousin of tort causation but ever since the time of ***Hadley v. Baxendale***[3], it is still a relevant factor in determining liability. This is an important factor – and again one we'll return to later.

Finally, it is worth remembering that the primary remedy for breach of contract is damages (as assessed to put the innocent party in the position they would have been had the contract been correctly performed).

The primary advantage of contract liability is of course that it is open to the contract counterparties to determine the scope of the contract responsibilities and obligations as between them (and hence the liability) if things do go wrong. This means that it is quite open to tailor the agreement to the functions and performance of the AI system involved.

So far as the implied terms are concerned, Section 9 of the Consumer Rights Act provides that where goods are sold "in the course of a business" there is an implied term that the goods are of *satisfactory quality*[4] and *fit for a purpose*[5] that the buyer has made known to the seller. Products are therefore of satisfactory quality if they meet the standard that a reasonable person would regard as satisfactory, taking into account their description, price and all other relevant circumstances. In other words, it could be argued that contract implied terms create a consumer expectation test.

The major disadvantage of contract liability of course is that it is not a liability that applies generally to the "whole world" but rather is one which is constrained by *contract privity* (albeit with limited exceptions in some jurisdictions – such as the Contract (Rights of Third Parties) Act 1999 in the UK).

This means that obligations can only be enforced by contract counterparties. Of course – in some situations

[3] [1854] EWHC J70
[4] s9
[5] s10

Liability Issues in Autonomous and Semi-Autonomous systems

it is possible to conceive of a contract relationship subsisting in your use of an intelligent systems – but equally speaking in many others, there will not be.

Aside from this, there is a lack of consistency in contractual standards in relation to contracts for the sale of goods which make the application of the framework complex. So for example, for public policy reasons, there are higher standards which apply to contracts made with consumers – the Consumer Rights Act 2015 requires that in assessing whether products are of satisfactory quality that account is taken of any "*public statements on the specific characteristics of the goods made about them by the Seller or the producer*"[6].

Tort

Product liability in tort refers to a breach of a duty of care in negligence. Since the seminal case of ***Donoghue v. Stevenson***[7], tortious duties can run concurrently with contractual liabilities.

The essence of the case was that if a consumer purchases products in a form intended to reach him or her without the possibility of reasonable intermediate examination and with the knowledge on the part of the producer that the absence of reasonable care in the preparation of the product will result in personal injury or property damage, which is reasonably foreseeable, then that producer owes a duty to take reasonable care in their production**.** ***Donoghue v. Stevenson*** concerned decomposed snails in ginger beer bottles but it does not take much to extrapolate that analysis to a driverless car or a surgical robot.

Again, it is worth pointing out the causative nature of tort as a liability framework – it is essentially fault based. The claimant must prove that the defendant owed him or her a duty of care, they failed in that standard and damage was caused as a result.

In contrast to contractual damages, tort based damages are awarded on the basis of putting the injured party in the position they would have been had the tort not occurred.

The scope of potential liability in tort is wide. It could equally apply to manufacturers, producers and anyone directly involved in the manufacture and distribution of a product with a defect. You do however need to establish that a duty of care subsists and was breached – and irrespective of this that the relevant chain of causation is not broken by the damage being too remote.

There are a number of disadvantages with this liability framework – and it is worth pointing them out briefly. There are very real difficulties in claiming damages for pure economic loss in tort – certainly so far as the UK is concerned, and since the high water mark of ***Junior Books v. Veitchi***[8], there are only a limited number of circumstances where this is possible, including for example, negligent advice from surveyors.

Contributory negligence can also act as a defence to liability, if it is shown that the claimant should have known of the defect but negligently failed to recognise it or negligently used the product or failed to take account of its operating instructions. In such cases damages are reduced to a degree which is commensurate with the claimant's negligence.

Volenti non fit injuria – or voluntary assumption of risk is less common in product liability cases – on the basis that if a claimant knows of the defect they are less likely to use it and if they do, that usually breaks the causative chain between defect and damage.

Finally, it is worth pointing out that factually, proving liability in tort can be very difficult – especially in product liability cases, as very often the details that are required to show liability are held by the defendant.

Obviously there are mechanisms in litigation to get hold of that information – but a defendant will not go out of the way to disclose it and might use any number of established disclosure tricks to bury the key facts or at least make them very difficult to ascertain.

Strict Liability: Consumer Protection

Finally, in relation to product liability, there is the Consumer Protection Act 1987 ("CPA") which implements the EU directive on Liability for Defective Products[9]. This Act introduces a strict liability regime which does not affect the general availability of Contract and Tort based remedies. What the Act provides is that a person who is injured or whose personal property is damaged by a product will be able to claim against a manufacturer or supplier of that product (and certain other third parties) if it can be shown that the product was defective.

There is no requirement to prove fault on the part of the manufacturer but obviously there is a requirement on the claimant to show that the defect existed on the preponderance of the evidence.

[6] s9(5)
[7] [1932] AC 562
[8] [1983] 1 AC 520
[9] 85/374/EC

Liability Issues in Autonomous and Semi-Autonomous systems

The Act introduces a consumer expectations test in that a defect exists where "*the safety of the product is not such as persons generally are entitled to expect*"[10]. Consumer expectations themselves are subject to a reasonableness test.

In terms of the advantages of the CPA regime – as previously discussed – there is no requirement to show fault; neither is there a privity requirement – the regime itself allows for a wide variety of potential liability targets, including suppliers and manufacturers.

There are still some problems with consumer protection product liability however – obviously causation still exists – although it is limited to the finding of defects and moderated by a consumer expectation test. Furthermore, the Act is designed to cover claims for real damage, so does not encompass claims for pure economic loss.

In the context of AI, there are also problems with the definition of "*Product*" under the Act. Product is defined as "*any goods or electricity and includes products aggregated into other products, whether as component parts, raw materials or otherwise*"[11]. Quite ambiguously for our purposes the Act is not clear as to whether software and/or other products of an intellectual type are included in the definition of its scope. Disembodied Software per se is not treated as a "good" under English law although there is an argument which might encompass software embedded into functional hardware.

Finally there is also the "*developmental risks defence*", which provides a defence to the manufacturer "*if the scientific and technical knowledge at the time the product was manufactured was not such that a producer of a similar product might have been able to discover the defect*"[12].

This is obviously highly relevant to our current discussion which inevitably involves the "state of the art" in relation to machine development.

AI: A failure in causation?

So to recap all of the previously discussed liability frameworks require some element of causation to a greater or lesser degree. In general terms, the scenarios that I have listed below are very comfortably dealt with by them; traceable defects; machine decisions that can be traced back to defective programming; failures to provide correct operating instructions – and incorrect operation of machines.
But what happens when the "defect" is inexplicable, or an event cannot in fact be traced back to a defect or a fault or a directly related human error?

As intelligent machines and AI systems "learn" for themselves, their behaviours are increasingly less and less directly attributable to human programming. These machines are not acting on a prescriptive instruction set, but a system of rules that may not have anticipated the precise circumstances under which the machine should act.

To take the example of our driverless car, what if our vehicle has been programmed to look after and preserve the safety of its occupants and also to avoid pedestrians at all costs and is placed in an unavoidable situation where it has to make a decision as to whether to avoid a pedestrian crossing into its path (and thereby run into a brick wall, injuring or even killing its occupants) or running over the pedestrian (and thereby saving its occupants). Can any outcome of that decision be said to be a failure or a defect – even if people are injured or possibly killed as a result?

It is at this relatively new interface where existing product liability frameworks begin to weaken and break down.[13]

There are some partial fixes in the existing liability frameworks. Tort in particular provides for the principle of ***res ipsa loquitur*** – or the thing speaks for itself. The doctrine is equally applicable in the US and the UK.

Res ipsa loquitur is useful in dealing with cases where there are multiple successive inexplicable failures which cannot in themselves be readily explained. A classic example of the application of this was in the US case of ***Toyota Motor Corporation***[14], where Toyota found that for no particular reason, many of its high end Lexus model cars simply accelerated – despite the intervention of their drivers. Despite much investigation, the cause of these failures could not be pinpointed. Toyota took the step of settling 400 pending case against it after an Oklahoma jury applied the doctrine *of res ipsa loquitur* and awarded the plaintiffs in that case $3m in damages.

Of course, as we have already seen the Consumer Protection Act in the UK also provides a partial fix as the

[10] s3(1)
[11] s1(2)(c)
[12] s4(1)e

[13] See for example "*Machines Without Principals*", by David C. Vladek, Washington Law Review Vol 89:117
[14] (2013) WL 5763178 (Texas)

requirement to identify a defect in a "product" is moderated by a consumer expectations test.

What this does not fix of course is the inexplicable isolated incident. That is where existing frameworks fail and we move increasingly into more and more speculative territory.

Future Liability Frameworks

Insurance

There are clear public policy arguments now for the introduction of a strict liability insurance based model. The 1973 Accident Compensation Act in New Zealand is a classic example of such a system working in practice – of course not directly in relation to AI systems, but rather in connection with motor vehicle accidents. In New Zealand, road traffic accidents are not litigated, but rather victim compensation is automatically paid at Government set tariffs and funded by motor insurance premiums.

Negligence and breach of contract actions are becoming more and more complex to litigate – as resources are spent identifying what has (or indeed might) have gone wrong. In particular, the argument runs that it is better spending the money compensating the victims of accidents and incidents involving autonomous systems, than it is on expensive lawyers and expert witnesses.

So far as research and development is concerned, a strict liability insurance based model will also incentivise research on new intelligent AI based systems, rather than forcing R&D divisions of corporations to consider what defensive steps they should be taking to avoid a class action.

Turing Registries

Some commentators[15] have argued specifically that for intelligent machines we need to go a step further and set up what have been termed "Turing Registries" after the great computer pioneer Alan Turing.

This would work by submitting intelligent machines to a testing and certification process to quantify the certification based on a spectrum as follows: the higher the intelligence and autonomy and hence greater consequences for failure, the higher the premium payable to "release" that machine into the working environment.

The premium payable for certification would be paid by the developer or manufacturer of the AI entity wanting to deploy the AI into the market. Much as in the manner of the New Zealand strict liability insurance model discussed earlier, the premiums would fund a "common pool" under which risks would be paid out. The system could become self-fulfilling if AIs were prohibited from use without this certification.

As has been pointed out, this model is similar, but *not identical* to insurance – it does remove causation and proximate cause but also allows for the wilful acts of AIs – normally something that is excluded by insurance.

Individuation

Now firmly in the realms of science-fiction, we need to consider briefly the inevitable question of how our liability rules will cope with machines when they *individuate* – that is to say develop distinct legal personalities and individual identities of their own.

Clearly if machines possess a sentient personality of their own then there is no reason why they cannot directly accrue liability in the same manner in which living breathing humans accrue it.

There are no answers here obviously, but rather questions. We can however consider some current and historical legal analogues. How are we going to treat self-aware machines and are we going to fall into the unfortunate historical bear traps of the past? The historical treatment of slaves bears particular scrutiny in this regard and the questions raised by that treatment are no less relevant now, although not for discussion in this paper.[16]

If we walk further down the path of individuation, and give them personal liability, how are we going to apply legal duties of care to intelligent machines?

Again, I don't propose to dwell too deeply on this issue, but this quote (ironically from a 1957 US case) is particularly apposite:

"a human being, no matter how efficient, is not a mechanical robot and does not possess the ability of a

[15] See for example "*Liability for Distributed Artificial Intelligences*", Curtis E.A. Karnow, Berkeley Technology Law Journal 1996, Vol 11.1, page 147

[16] See e.g. the description of "*Slave*" in the 1825 version of the Louisiana Civil Code at Article 35 – "*One who is under the power of his master, and who belongs to him; so that the master may sell and dispose of his person, of his industry, and of his labor, without his being able to do anything, have anything, acquire anything, but what must belong to his master.*"

Liability Issues in Autonomous and Semi-Autonomous systems

radar machine to discover danger before it becomes manifest. Some allowances… must be made for human frailties and reaction…"[17]

Ironically, this takes us full square back to our quixotic desire to imbue intelligent machines with infallible characteristics – the point being that a super intelligent machine may well exhibit the same frailties as a human being, precisely because it is simulating (or emulating) one.

Criminal Liability

Criminal liability is the final area of discussion in this paper. Obviously intelligent machines committing heinous crimes play to our darkest fears and make for sensationalist headlines.

Without going into too much detail, there are as we are all aware, two elements of criminal liability which need to coincide in the guilty actor for liability to be made out – ***Actus Reus*** – the criminal act and ***Mens Rea***, the criminal intent or criminal mind.

The ***Actus Reas*** element as we know can be expressed by acts or omissions and the ***Mens Rea*** element has various levels ranging from knowledge and specific intent all the way to criminal negligence or "recklessness".

Commentators[18] have proposed that there are three possible models for machine based criminal liability:

- The "perpetuation by another" model
- The Natural Probable consequences model; and
- The Direct Liability model

The "perpetuation by another" model is perhaps the most "conventional" criminal liability analysis and does not confer any personal human attributes to the machine. It is merely an "innocent agent". The machine on this basis is essentially treated as a tool or instrumental extension of the perpetrator, which is where the real criminal liability is focused.

To take a specific example, imagine a sophisticated aircraft that ejects its pilot out of the cockpit, thereby killing him.
In the case of AI systems, who is likely to be the "*perpetrator by another*"?

One candidate might be the programmer of the AI software – so for example if the programmer wrote the AI software with the specific mens rea of killing the pilot, then they are the perpetrator.

Another candidate might be the user of an AI system, where specifically the user orders the AI to take a particular course of conduct which would lead to a crime being committed – such as a person who orders his dog to attack a burglar. The dog commits the assault but the user is deemed to be the perpetrator.

The next model which has been proposed in the context of machine liability is what is termed the "*natural probable consequences*" model. I should say at the outset that this is a highly problematic doctrine in its own right and has been largely discredited in many US states (as well as comparable jurisdictions, such as the UK).[19]

The doctrine evolved as a way to embrace accomplice liability (i.e. where there been offence committed by an accomplice which is not part of a conspiracy but which would otherwise be a "natural probable consequence" of a criminal scheme).

In this model, the AI takes action which is a "*natural and probable*" consequence of its programming. So going back to our earlier example of the airplane ejecting the pilot, in this model, the programmer does not need to have a specific intent (or ***mens rea***) to kill the pilot, but rather a state of criminal negligence – that is to say reckless disregard as to whether the programming supplied to the AI could lead to the ejection of the pilot – ***if*** a reasonable person in the place of the programmer could have forseen the offence as a natural probable consequence of the AI's programming.

Finally, we have the "*direct liability*" model which assumes some level of personal responsibility on the part of the machine for its actions.

To a certain degree this plays back to our science-fiction themed debate on individuation earlier. We are not currently at a technological state of the art where this is relevant, but it is useful to consider how the criminal law might evolve in this circumstance.

In order for this model to work, self-awareness and an understanding of act and consequence on the part of the AI perpetrator are essential.

Obviously the law already recognises circumstances where there is a reduced level of legal liability owing to a commensurately reduced ability to distinguish right from

[17] *Arnold v. Reuther* 92 So. 2d 593 (La. Ct. App. 1957)
[18] See "*The Criminal Liability of Artificial Intelligence Entities*", Gabriel Hallevy SSRN 1564096
[19] See "*The Natural and Probable Consequences Doctrine: A Case Study in Failed Law Reform*", Michael G. Heyman, Berkeley Journal of Criminal Law, Vol 15, Issue 2

Liability Issues in Autonomous and Semi-Autonomous systems

wrong – the principle of ***Doli Incapax*** in relation to infants and the mentally incapable spring to mind in this regard. Might some machines that are not fully capable of understanding the consequences of their actions fall within this principle in the future?

Finally, and even more speculatively, commentators have tried to understand how individually accountable machines might be punished on a criminal basis.

Skirting into the realms of philosophy as well as science fiction, the death penalty might mean deletion for a self-aware machine. Imprisonment might mean removing the AI from its intended purpose and a criminal fine might be translated into re-using the machine for a community purpose away from its original role.

Conclusion

We are now effectively at a "*tipping point*" in how we manage the machine generated consequences of our new AI creations. As we have seen, Contract, Tort and Strict Liability consumer protection laws are effective to a degree in relation to managing these consequences but effectively break down where cause and effect cannot be made out. As has been pointed out by non-legal commentators:

"If we want to avoid the injustice of holding men responsible for actions of machines over which they could not have sufficient control, we must find a way to address the responsibility gap in moral practice and legislation"[20]

John Buyers
Partner, UK Head of Commercial
T +44 20 7105 7105
john.buyers@osborneclarke.com

[20] "*The Responsibility Gap – Ascribing Responsibility for the actions of learning automata*", Andreas Mathias, University of Kessel Computing Centre