Responsibility, Autonomy and Accountability: legal liability for machine learning

Chris Reed
Elizabeth Kennedy
Sara Nogueira Silva
Responsibility, Autonomy and Accountability: legal liability for machine learning†

Microsoft Cloud Computing Research Centre
Paper presented at the 3rd Annual Symposium
8-9 September 2016

Chris Reed, Elizabeth Kennedy, Sara Nogueira Silva*

This paper investigates the question of legal liability for the consequences of decisions made by machine learning technology rather than by humans, although we do not attempt a detailed analysis of the basis on which such liability might be imposed. This is a substantial task which would require far more space than is available here.

The initial focus is on private claims for personal injury, property damage and other losses caused by use of machine learning technologies. These claims will usually be made via the tort of negligence.

Equally importantly, we identify some of the threats to individual autonomy and fundamental rights which are created by the use of machine learning to make decisions. Breach of those fundamental rights is a second source of potential liability.

We conclude by suggesting a potential link between liability and the preservation of those fundamental rights which might achieve an interim solution to this issue, making use of the concept of accountability and its transparency attribute in particular.

1 Introduction

Machine learning has recently made the transition from a purely research activity, and has generated a range of technologies which are beginning to be implemented outside the laboratory. These technologies have the potential to replace human decision-makers, and where they do so they give rise to liability questions which the law has not so far needed to

† This paper has been produced by members of the Microsoft Cloud Computing Research Centre, a collaboration between the Cloud Legal Project, Centre for Commercial Law Studies, Queen Mary University of London and the University of Cambridge Computer Lab. The authors are grateful to Christopher Millard, John Cowcroft, Jean Bacon, Jatinder Singh and Guido Noto la Diega for their helpful comments and to Microsoft for the generous financial support that has made this project possible. Responsibility for views expressed, however, remains with the authors.

* Microsoft Cloud Computing Research Centre, CCLS, School of Law, Queen Mary University of London.
address. The most difficult issues of legal liability for machine learning decisions arise where the consequences of a decision result in loss or damage to a third party.

Some machine learning technologies clearly have the potential to cause physical injury or property damage. Self-driving vehicles\(^1\) are perhaps the most obvious example, and they will feature largely in this paper. Machine learning is also being used for medical diagnosis and treatment\(^2\); here too the potential for causing physical injury is clear, and the liability question is complicated by uncertainties as to the standard of performance which a machine learning technology should be expected to achieve. Our initial discussion of liability will focus mainly on physical injury and property damage because these are of the greatest concern to society, and also because we wish to focus on the high level principles involved without the distraction involved in dealing with special cases.\(^3\) Our examples will be drawn primarily from English law.

This does not mean that the potential for machine learning decisions to cause intangible damage is unimportant. As just one example, an inaccurate credit rating report produced by machine learning technology can damage a person’s financial reputation, and perhaps even their social reputation if the report were to, say, identify that person as a money-laundering suspect. This would, inter alia, engage the law of defamation. Intangible damage can also be caused if machine learning technology wrongfully discloses certain kinds of information. Such disclosures might give rise to liability under the law of confidentiality, or under data protection law.\(^4\)

In all these cases liability would arise because the machine made an incorrect decision. But it is also possible for liability to arise as a result of ‘correct’ decisions, at least for some values of correctness. Social policy decisions which are embedded in the law, such as the right not to be discriminated against on grounds of race or sexual orientation, constrain human decision makers to ignore what might otherwise be objectively relevant factors when making certain decisions. A well-known example is the law against sex discrimination, which has in recent years prevented motor insurers from granting insurance policies to women on more favourable terms than to men, even though the statistical evidence is clear that women present a lower risk.\(^5\) A machine learning decision which fails to observe these constraints may give what is objectively a correct decision, in terms of the available data, but because it fails to follow the law’s constraints that decision will be incorrect as a matter of law. Irrespective of whether there is liability under the law at present, the question whether

---

1 Self-driving, or autonomous, vehicles use technologies which have learnt how to operate the vehicle as an evolutionary process, and thus developed a model of the driving process, rather than being controlled in accordance with a model generated by the mind of a human programmer. Decisions about how the vehicle should be controlled are not longer made by a human driver, but by the technology acting autonomously.

2 ‘Google knows your ills’, New Scientist 7 May 2016, 22.

3 Such as liability for psychological damage and liability for purely financial losses.


5 Association Belge des Consommateurs Test-Achats ASBL v Conseil des Ministres (Case C-236/09, 1 March 2011). Employment decisions, such as shortlisting for interview, are increasingly assisted by machine learning and would give rise to liability on the same basis – see eg Chen-Fu Chien & Li-Fei Chien, ‘Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry’ (2008) 34 Expert Systems with Applications 280.
machine learning decisions should be required to comply with those fundamental rights merits investigation.

A complicating factor, however, is that machine learning decisions are hard for humans to understand. This is because the way the technology makes its decisions is not solely designed by a human developer – rather the technology learns patterns and relationships from data, and thereby builds a model of the process involved.  

Machine learning encompasses a range of techniques, and the level of comprehensibility varies across that range.  

At one end lie technologies such as decision trees, which embody chains of logic which lead to each possible decision, and thus allow the reasons for any particular decision to be explained.  

At the other end, neural networks identify patterns in large data sets and then make decisions based on pattern matching. Here it may not be possible for even the technology producer to explain a decision in terms of the logic and causation which the law looks for.

Technical systems whose workings are not understandable by humans are often described as ‘black box’ systems. For the purposes of the discussion below, a system might be a ‘black box’ to one person but not to another. For example, the producer of a machine learning technology might be able to explain how and why it reaches its decisions, whilst to the user of the technology these matters will be unknowable. This distinction is important because the liability questions discussed here are often dependent on what the responsible person knew, or ought to have known, at the time the liability arose. Thus in most cases, all that the user of the technology knows is that he is ignorant of its workings, and that it is de facto a ‘black box’. The law will also distinguish between what a technology producer knew or ought to have known in advance, for example through the process of testing and evaluating the technology, and what can be discovered after the event if the technology fails to make a correct decision.

Finally it is worth noting that most implementations of machine learning technology do not engage in further learning during use because, as in the example of self-driving vehicles, this would mean that each implementation might behave differently from every other. In some cases, data about the performance of some or all implementations will be collected and used to retrain the master version, and then all implementations updated together once the retraining has been evaluated. This characteristic is likely to affect the way the technology is supplied to users, and consequently its legal status for liability purposes.

2 Liability for loss or damage caused by machine learning decisions

2.1 Justifications for imposing liability

Liability is based on loss or damage which has been caused by some person, activity or property. The function of the law is to allocate responsibility for that causal element to some person, and then to assess whether liability arises based on the nature of that responsibility.

---

7 For more detail see ibid, parts 2 (learning and evaluation) and 3 (the role of data).
8 Though note that the logic embodied in a decision tree may be so comprehensive and complicated that it is in practical terms infeasible to comprehend it in advance, and thus only ex post facto explanations are likely to be achievable.
There are, broadly speaking, two possible bases on which responsibility can be assigned. The less common is known as strict liability, under which the responsible person is liable irrespective of any fault on that person’s part. In most cases the likely source of responsibility will be the second basis, the duty of care imposed by the law of negligence.

### 2.1.1 Strict liability

Strict liability under English law has developed piecemeal, essentially in response to the recognition of dangerous activities or states of affairs against the consequences of which the person responsible is required to indemnify the remainder of society. This means that liability is imposed no matter how careful the responsible person has been. Examples include:

- Keeping animals of a dangerous species, or keeping animals of a non-dangerous species which have characteristics which make them unusually dangerous compared to others of the same species.\(^9\)
- Owning an aircraft. The owner of an aircraft is liable for all personal injury or property damage caused by an aircraft in flight, or while taking off or landing.\(^{10}\)
- Accumulating on a person’s land something which is a non-natural use of the land, and which is likely to cause damage if it escapes from the land.\(^{11}\) This type of liability only extends to damage to the neighbouring land and not to personal injury.\(^{12}\)

Because machine learning is such a new technology, no laws have yet been enacted imposing strict liability on the users of such technology. However, in some fields of activity many of the Civil Law jurisdictions impose strict liability, eg on the drivers or keepers\(^{13}\) of motor vehicles\(^{14}\), and these laws would apply to loss or damage caused by machine learning technologies without difficulty because the cause of the loss or damage is irrelevant to liability.

In UK law there is one general category of strict liability which might be relevant to machine learning technology. This is the liability of producers or suppliers of defective products under the Consumer Protection Act 1987.\(^{15}\) The Act applies to producers or suppliers of products in the course of a business\(^{16}\), and makes them liable to anyone who suffers personal injury or property damage caused by a defect in that product, irrespective of any fault on the producer’s part. The UK Government considers that this might be an appropriate regime to deal with the liability issues of self-driving vehicles, though in our view the law will require more extensive modification than is currently proposed.\(^{17}\)

---

9 Animals Act 1971 s 2.
10 Civil Aviation Act 1982 s 76(2).
11 *Rylands v Fletcher* [1868] UKHL 1 (escape of water from a reservoir).
13 The term ‘keeper’ is commonly used in motor insurance law to deal with situations where the owner is not the person who has control over the use of the vehicle, eg hire purchase or contract leasing companies who are the legal owners.
16 Section 4(1)(c).
17 *Pathway to Driverless Cars: Proposals to support advanced driver assistance systems and automated vehicle technologies* (UK Department for Transport: Centre for Connected and Autonomous Vehicles, July 2016) paras 2.29-2.30.
The main requirements for liability are set out in section 2(1), which provides that subject to the remaining provisions of the Act:

...where any damage is caused wholly or partly by a defect in a product, every person to whom subsection (2) below applies\(^\text{18}\) shall be liable for the damage.

Products are, in essence, tangible or movable objects\(^\text{19}\), and so if machine learning technology is built into, say, a motor vehicle, it becomes part of that product. However, if the machine learning technology is hosted in the cloud, so that its users receive it as a service, the product liability regime will not apply.

Liability claims under the Act are limited to claims for death or personal injury\(^\text{20}\) or damage to property (including land) which is ordinarily intended for private use or consumption and was so intended to be used by the claimant.\(^\text{21}\) Product liability does not cover damage to the product itself or to any product containing the defective product and the damage must exceed £275.\(^\text{22}\)

Section 2(1) requires that the damage be caused by a ‘defect’ in the product, and this is defined in section 3. A product is defective if it does not provide the level of safety (in respect of property as well as the person) that persons generally are entitled to expect. This is where a product liability claim is likely to fail. Machine learning technologies which have the potential to cause physical injury or property damage are only likely to be adopted if their testing indicates that they provide a higher level of safety than does the human decision-making which they replace. There will almost always be available evidence to show this, and such evidence should be enough to demonstrate that the product is not defective under that definition.

It is conceivable though that a court might decline to assess the safety provided by the machine learning technology in aggregate, and instead concentrate on the particular circumstances in which the loss arose. An example might be the recent fatal crash of a Tesla motor car being operated by its Autopilot technology.\(^\text{23}\) At the time of writing, with the crash still to be fully investigated, it appears that the Autopilot technology failed to recognise an unusual configuration of goods vehicle in the particular light conditions.\(^\text{24}\) If the court interpreted the Act as requiring the car to provide an appropriate level of safety in those particular circumstances, it would be possible to find that the car was defective within the meaning of the definition, even despite evidence that it is safer overall than a car driven by a human.

Even if the product is unsafe, the product producer might be able to avail itself of the ‘state of the art’ defence provided by s 4(1)(e). This states that the producer can escape liability if he can show that the defect is such that a reasonable producer would not, in the current state of the art in that industry, have discovered the defect. In effect, this requires the producer to show that the combination of circumstances which led to the loss was not one which a reasonable

\(^{18}\) Ultimately, the producer of the product or its importer into the EEA, though intermediate suppliers are also liable until they identify some person higher up the supply chain who is liable.

\(^{19}\) See the definitions in the Consumer Protection Act 1987 of product (s 1(2)) and goods (s 45) and the Directive 85/374 EEC definition as any moveable (art 2).

\(^{20}\) Section 5(1).

\(^{21}\) Section 5(1) & (3).

\(^{22}\) Section 5(2) & (4).


producer would have tested for, and had the Tesla crash occurred in the UK this defence might well succeed if the reported circumstances are accurate.

All this tells us that although strict liability for products might sometimes apply to machine learning technologies, its application is likely to be limited to a small minority of those technologies. This is particularly so because one of the main aims of technology producers is that the technology should continue learning from the successes and failures of its implementations. Such continued learning is most effective if the experiences of all the instances of the technology in use are aggregated, and aggregation will require continual reporting back and updating. Products are fixed in nature, and so if the machine learning element of a technology is regularly changing via external updates it is likely to be marketed as a service, and treated by the law as such rather than a product.

In most cases of loss or damage caused by machine learning, we will therefore need to look to the law of negligence to identify responsibility and allocate liability.

2.1.2 Negligence

Negligence requires a defendant to be at fault. The defendant must have a responsibility to take care, have failed to do so, and to have caused damage as a result. The classic formulation is that negligence is “the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do.”

To establish liability, generally a claimant must establish all the elements of the tort of negligence on the balance of probabilities:

a. the defendant must have owed the claimant a duty of care;

b. the defendant’s conduct must have fallen below the standard of care (breach of duty); and

c. the claimant must have sustained damage which was caused by the defendant’s breach of duty (causation).

If the claimant succeeds in proving these elements the defendant may still be able to put forward one of a number of defences. The most important of these is contributory negligence, under which the defendant’s liability is reduced in proportion to the amount by which the claimant’s own negligence contributed to the loss or damage.

Duty of care

Where the parties have a pre-existing relationship, particularly where the defendant is providing some professional skill, it will not be difficult to establish that a duty of care is owed. Typical examples where machine learning might be involved are medical professionals, architects and professional advisers.

However, under general negligence the claimant must first show that the defendant owed him a duty of care. The judgment of Lord Atkin in the seminal case of Donoghue v

---

25 Blyth v Birmingham Waterworks Co (1856) 11 EX 781, 784 per Alderson B.

26 In many cases such a pre-existing relationship will be contractual, and a contract for services imposes an obligation that the services will be provided using reasonable care and skill – Supply of Goods and Services Act 1982 s 13. Here there will be no need to prove the existence of a duty of care, but the questions considered below about the scope of that duty, breach and causation will still arise.
Stevenson\textsuperscript{27} famously articulated the circumstances in which a general duty to take care can arise; this is known as the ‘neighbour’ principle:

You must take reasonable care to avoid acts or omissions which you can reasonably foresee would be likely to injure your neighbour. Who, then, in law is my neighbour? The answer seems to be – persons who are so closely and directly affected by my act that I ought reasonably to have them in contemplation as being so affected when I am directing my mind to the acts or omission which are called in question.\textsuperscript{28}

\textit{Caparo Industries v Dickman and Others}\textsuperscript{29} expanded upon the elements of establishing a duty of care. The court held that: the harm sustained must have been foreseeable; there must have been the requisite degree of proximity between the claimant and the defendant; and in the circumstances it must be ‘fair, just, and reasonable’ to impose a duty of care. Clearly, policy plays an active part in the process of allocating responsibility by imposing a duty of care.

\textbf{Breach of duty of care}

A defendant’s conduct is negligent where it falls below the objective standard of care required. In general, the standard of care is defined by reference to a ‘reasonable person’ acting in the same circumstances as the defendant.

In professional negligence though, the law recognises that individuals who hold themselves out to be professionals, skilled in a particular field, are to be judged by a different standard in the performance of their duties. This does not require the highest expert skill but that of an ordinarily competent person who is exercising that particular art.\textsuperscript{30}

There is also a temporal dimension to breach of duty of care. What was reasonably expected from a person in the position of the defendant should not be assessed in hindsight or in the light of subsequent technological or scientific developments.\textsuperscript{31}

\textbf{Causation}

Establishing that the defendant breached a duty of care is not sufficient for a finding of liability. There must be damage sustained because of the defective conduct. This requires a link between the breach of duty and the damage sustained. There are two types of causation; both must be present to establish this link. The defendant’s conduct must be the factual cause of the damage, and that damage must be attributable to the defendant’s breach of duty, ie the legal cause.

The starting point for determining factual causation is the ‘but for test’. This test asks whether ‘but for’ the defendant’s negligence the claimant’s injury would have occurred.\textsuperscript{32} The claimant must persuade the court that it is more likely than not that without the negligence of the defendant the injury would not have been sustained (\textit{causa sine qua non}). In \textit{Barnett v Chelsea and Kensington Hospital}\textsuperscript{33} three night-watchmen attended the casualty department of the defendant’s hospital complaining that they had been vomiting after

\begin{itemize}
    \item \textsuperscript{27} [1932] AC 562.
    \item \textsuperscript{28} [1932] AC 562, 580.
    \item \textsuperscript{29} [1990] 2 AC 605.
    \item \textsuperscript{30} \textit{Bolam v Friern Hospital Management Company} [1957] 1 WLR 582.
    \item \textsuperscript{31} \textit{Roe v Ministry of Health} [1954] 2 QB 66.
    \item \textsuperscript{32} \textit{Cork v Kirby MacLean} [1952] 2 All ER 402.
    \item \textsuperscript{33} [1969] 1 QB 428.
\end{itemize}
drinking tea. The night-watchmen were told to go home and call their own doctor in the 
morning. One of the men died that night. It transpired they had been poisoned by arsenic in 
the tea. The widow of one of the men brought an action against the hospital for negligence. 
The Court held that Mr Barnett would have died from arsenic poisoning regardless of 
whether the defendants had admitted and treated him. Mr Barnett would not have survived 
‘but for’ the defendant’s negligence, and therefore the defendants were not liable.

Where multiple events have caused the claimant’s injury it may be difficult for a court to 
apply the ‘but for’ test. For example, if a car was hit successively by two other vehicles, 
injuring the passengers, the ‘but for’ test may fail to hold either of the drivers liable for 
those injuries. This is because neither collision could be said to be the ‘but for’ cause of the 
injury. Had the first collision not occurred, the injury would still have been caused by the 
second collision and vice versa. Another example often quoted is of two hunters who 
carelessly fire shotguns one after the other and injure a third hunter. If it cannot be 
established which hunter’s shot hit the victim, it cannot be said that either was the ‘but for’ 
cause. In such situations, despite finding that both defendants acted carelessly neither could 
be found liable under the test.

Policy considerations weigh strongly against leaving a claimant uncompensated where the 
‘but for’ test fails to identify causation as between multiple careless defendants, and so the 
courts have developed a number of special tests to establish factual causation in these 
circumstances which are examined in part 2.2.2 below.

Once the claimant has proven a factual nexus between the conduct of the defendant and 
the damage he sustained, the claimant must demonstrate a legal connection. Establishing 
legal causation comprises two parts. Firstly, it requires consideration of whether the damage 
was within the foresight of a reasonable person in the position of the defendant at the time 
of the breach (the remoteness test). This requires that the type of damage sustained by 
the claimant was reasonably foreseeable. The extent of the actual injury does not have to be 
foreseeable for the remoteness test to be satisfied. Defendants are liable for the full extent 
of the harm they cause, even where that harm is more extreme than might ordinarily have 
been the case due to an existing vulnerability of the claimant.

Secondly, legal causation requires that there be an unbroken chain of causation between 
the defendant’s negligence and the claimant’s injury. Should any third party action break the 
foreseeable chain of causation (novus actus interveniens) the defendant will not be liable for

34 See Lord Nicholls’ discussion in Fairchild v Glenhaven [2003] 1 AC 32, 39 of the Supreme Court of 
Canada case Cook v Lewis [1951] SCR 830:

As between the plaintiff and the two hunters, the evidential difficulty arising from the 
impossibility of identifying the gun which fired the crucial pellet should redound upon the 
negligent hunters, not the blameless plaintiff. The unattractive consequence, that one of the 
hunters will be held liable for an injury he did not in fact inflict, is outweighed by the even 
less attractive alternative, that the innocent plaintiff should receive no recompense even 
though one of the negligent hunters injured him. It is this balance (‘... outweighed by ...’) which 
justifies a relaxation in the standard of causation required. Insistence on the normal 
standard of causation would work an injustice.

35 See Fairchild v Glenhaven, n 34, 67 per Lord Bingham: ‘I am of opinion that such injustice as may be 
involving in imposing liability on a duty-breaking employer in these circumstances is heavily 
outweighed by the injustice of denying redress to a victim.’

36 The Wagon Mound (No 1) [1967] 1 AC 617, Privy Council (Australia).


38 See for example Smith v Leech Brain & Co Ltd [1962] 2 QB 405.
the damage. Actions of the claimant are unlikely to be considered as novus actus interveniens, but instead will be considered at the stage of defences under the head of contributory negligence.

### 2.2 Machine learning and negligence liability

When examining how the law of negligence might apply to machine learning, it is important to remember that negligence is about the carelessness of persons. Thus negligence is not interested directly in the decision-making quality of a machine learning technology. The law asks other questions, such as whether the producer of the technology took sufficient care when designing, constructing or testing it, or whether the decision by the user of the technology to adopt it was a reasonable one, and whether it was operated with appropriate care and skill. In other words, it seeks human, not machine, failings.

Machine learning introduces new challenges for the law of negligence at each of its three stages: establishing a duty of care; assessing breach of duty; and determining if the breach of duty caused the loss or damage.

#### 2.2.1 Duty of care

Where claimant and defendant are in a relationship which automatically gives rise to a duty of care, such as that between doctor and patient, the use of machine learning raises no special duty questions. Machine learning will be relevant, as we will see below, to the question whether the duty of care has been discharged, but does not affect the duty itself. Thus in the doctor-patient relationship the duty remains the same; the doctor must use reasonable care and skill in diagnosing and treating the patient.

However, where there is no such pre-existing relationship the duty of care question becomes more difficult. Let us take the example of autonomous 39 driving technology.

In the case of a normal motor vehicle, its driver owes a duty of care to others in the vicinity who are potentially at risk if the car is not driven carefully. This group of persons includes other road users, pedestrians, owners of property adjoining the highway, etc. The duty is imposed under the principle in Donoghue v Stevenson 40 because it is foreseeable that those persons are at risk.

However, a motor vehicle under the control of autonomous driving technology does not, at that moment, have a human driver. What, then, will be the duty of care of the vehicle's occupant? Applying the Donoghue principle, we need to identify the foreseeable harms and how the occupant might take care to avoid those harms. The foreseeable harms created by the vehicle are the same as if there were a human driver, but they are not created by the occupant’s driving. Because a duty of care is placed on humans, not machines, in this case the occupant’s duty will be very different. We suggest that this duty will have two components:

- A duty to take reasonable care in making the decision to use the vehicle at all, in the circumstances. If the vehicle is approved for road use this component is unlikely to be relevant in normal conditions. However, if there are known limitations to the

---

39 The technologies currently under development exhibit a range of levels of autonomy: at one end the human occupant has no input at all into the driving, whilst at the other the human is always able to override the driving decisions made by the technology.

40 n 27.
machine learning technology, for example that it does not work reliably when there is snow on the road, then a decision to use the vehicle in those circumstances might well be a breach of the duty of care. This duty might also apply if the vehicle is likely to encounter unusual road or weather conditions, even if it is not known that the technology may not cope well with them, if the driver ought to have suspected that it might not cope or simply did not know the answer.

- A duty to operate the vehicle in accordance with any published procedures, such as maintaining continuous contact with the steering wheel, and to take over the driving from the machine technology whenever a reasonable driver would have decided to do so, assuming that the technology allows the occupant to take over.

From this we can see that an occupant of an autonomous vehicle would owe a far less extensive duty of care than a human driver in sole control, because the occupant is not fully responsible for controlling the vehicle’s actions and thus cannot take as much care for the safety of others. If the vehicle does not allow the occupant to take over, as might in future be the case for vehicles used only in special environments (perhaps shuttle vehicles from airport car parks to the terminal) then it is unlikely that the occupant would owe any duty of care at all. If the occupant shares control of the driving with the technology, the extent of duty might be limited to exercising reasonable supervision of the technology and taking over where necessary.

Might we, though, find a duty of care owed by the producer of the vehicle which would enable the law of negligence to impose liability for accidents? The problem we face here is that factual responsibility is shared between multiple parties.

At first sight it seems obvious that the manufacturer of the vehicle owes a duty of care to that segment of the public at large which is put at risk when the vehicle is operating. However, that duty is only to take reasonable care in the design and construction of the vehicle. If, as is likely, the self-driving technology is produced by a third party, then the manufacturer’s duty only extends to taking reasonable care when deciding to use that technology and when integrating it safely into the workings of the vehicle. There will probably also be a duty, on both the manufacturer and the producer of the technology, to alert users if defects in the technology become known.

The production of the technology itself may also involve split responsibilities. There are at a minimum four components to this technology: the element which controls the operation of the vehicle; that which makes decisions about the driving; the element which learns about driving; and the data on which the learning element is trained and improves itself. Any one or more of these might be developed by different entities. Because the technology will perform directly in a way which might cause loss or damage, without human intervention, we think it likely that the law of negligence will treat these producers in the same way as those who manufacture components for vehicles, and thus impose a duty of care on them. However, each will be entitled to rely on the others complying with their duty of care (in the absence of reasons to suspect that they have not done so), and so the likely scenario in any negligence claim is that each will assert that the loss or damage falls within the sphere of duty of one or more of the others, and not within its own duty of care. This is likely to make negligence litigation involving self-driving vehicles vastly more expensive than that which requires only determination of the level of care taken by a human driver.

If the machine learning technology does not have the potential to cause loss or damage directly, but instead assists humans in deciding how to act, then the duty of care question becomes more difficult. English law has taken a policy decision that those who give advice
will only owe a duty of care in negligence to that subset of society which is entitled to rely on the advice. This policy was first proposed by Lord Denning in the case of *Candler v Crane Christmas*41, where he pointed out that a maker of marine charts which contained a defect would otherwise have potentially unlimited responsibility to anyone in the world who used that chart, even if they have no stronger relationship with the producer than that of mere user. The policy has been adopted in a long line of cases since42 and is perhaps most clearly explained in *Caparo v Dickman*.43 In that case, the auditors of the company had approved its accounts, and in reliance on those accounts the claimants had invested money in the company. The accounts were defective and so the investors suffered financial losses. The House of Lords held that the auditors only owed a duty of care to those to whom they had undertaken a responsibility to act carefully in their audit activities. That group was restricted to the shareholders of the company, for whom the audit had been produced, and did not extend to external persons who were considering making an investment in the company.

Thus, if we turn to the example of a machine learning technology which assists in medical diagnosis and treatment by providing information to medical professionals, the producers of that technology will most likely only owe a duty of care to those professionals. To find that the producers of the technology owed a duty of care to patients, to whom no undertaking of responsibility has been given, a court would have to overturn this long-standing policy of the law. If the policy remains intact, we are faced with the situation where the medical professional has probably discharged her duty of care because the technology is normally so accurate in its advice that it is reasonable to rely on it, whilst the producer of the technology owes no duty of care at all. Thus, even if the patient’s loss was caused by carelessness on the part of the technology producer, there would still be no liability to compensate the patient.

### 2.2.2 Breach of duty and causation

The split of responsibilities between those involved in producing a machine learning technology is also likely to cause difficulties in assessing whether there has been a breach of duty, and if so whether the breach caused the loss or damage.

In the case of the human who owes a duty of care, and relies on the output of a machine learning technology to help in performing the activity in question, that human will only be in breach of their duty if it was not reasonable for them to rely on the output of the technology in the circumstances. Many kinds of machine learning technology will appear to their user to be a ‘black box’, which gives no clues about its workings and reasoning. Thus the law will focus its attention on the claims made by the producer of the technology, any independent assessments of its workings, how widely the technology has been adopted by others working in the same field, and the extent to which circumstances in which the technology fails to produce the correct answer have been identified and publicised. The test, as already explained, is whether a reasonable person working in the same field and professing the same expertise would have decided that it was safe to use the technology. These matters are comparatively easy to assess.

Deciding whether one of the multiple producers of the machine learning technology has failed to take reasonable care is a far less easy question. On the assumption that any machine learning technology will have been tested extensively before it is put into general

---

41 [1951] 2 KB 164.
42 For a review of the development of these cases see Jonathan Morgan, ‘The rise and fall of the general duty of care’ (2006) Professional Negligence 206.
43 [1990] 2 AC 605.
use, the likelihood is that the technology’s failure to produce the correct answer is because it is operating in circumstances for which it has not been tested. Thus to assess this matter, the court will need to understand why the technology failed. How far this is achievable will depend on how the technology works.

Where the technology in effect implements a decision tree it will be possible to explain the logical reasoning which led to the decision by the technology, at least if the technology has stored sufficient data in operation to enable the situation to be ‘replayed’ as a simulation. If that logic is defective, and this ought to have been noticed during analysis and testing, the court will be able to find that a breach of duty has occurred. However, there are two complicating factors here.

First, for all but the most simple systems the decision tree will not be a product of the human mind, but rather will have been induced by a machine learning algorithm from a mass of data. The logic of this decision tree may well be too detailed and complicated for the human mind to understand fully, what Burrell describes as:

> opacity that stems from the mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of semantic interpretation.44

This complexity will often mean that the only way of identifying defects is not through inspecting the logical reasoning of the algorithm, but rather by running the algorithm on test datasets and observing the decisions it makes. This then leads to the question whether those test datasets were sufficiently comprehensive to justify releasing that machine learning technology for widespread use. The evidence required by a court to decide these matters will be voluminous and technically complex. Even if it is possible to answer these difficult questions, the cost of doing so through litigation is likely to deter many claimants.

Second, where the machine learning technology’s decision-making element comprises a neural network, or some similar technology, it will be difficult and perhaps impossible to explain how the technology came to its decision, and thus how the loss or damage was caused.45 The producer of the technology will need to demonstrate that the technology learned how to make its decisions by analysing a sufficiently large and comprehensive training dataset, and also that the technology’s decisions were tested against a substantial real world dataset. This evidence will be sufficient to show that no negligence occurred unless the claimant can adduce evidence of deficiencies in either of these aspects of production. Given that the reasons for the incorrect decision are by definition unknown, it is hard even to imagine how a claimant might search for such evidence.

These problems of identifying how and why an incident occurred might be resolved using the principle of *res ipsa loquitur*, which is not a principle of substantive law46 but relates to rules of evidence. It is a label for common sense reasoning meaning ‘the thing speaks for itself’. In *Scott v London and St Katherine Docks*47 it was said:

> where the thing is shown to be under the management of the defendant or his servants, and the accident is such as in the ordinary course of things does not

---

46 Carroll v Fearon, Bent & Dunlop Ltd [1999] ECC 73, [15], per Lord Judge.
47 [1865] 3 H&C 596, 601.
happen if those who have the management used proper care, it affords reasonable evidence, in the absence of explanation by the defendants, that the accident arose from want of care.

*Res ipsa loquitur* applies where the incident would not be expected to have happened had the defendant not been negligent, and the cause of the damage was under either the sole management and control of the defendant or of someone the defendant had the right or responsibility to control. It follows that for the principle to apply there must be no evidence as to why or how the incident that caused the damage occurred, otherwise the claimant must use this evidence to establish on the balance of probabilities that the defendant was negligent.

The effect of successfully invoking *res ipsa loquitur* is to raise the inference of the defendant’s negligence. The precise procedural effects of this are unclear, and there is some uncertainty as to whether the burden shifts to the defendant to show on the balance of probabilities that they were not negligent, or whether it is sufficient for the defendant to provide a reasonable explanation for the incident which occurred. In a Privy Council case from 2009 which adopts the latter meaning, the Court held that ‘[t]his was the respondent’s aircraft, their flight and their pilot. Aircraft, even small aircraft, do not usually crash’, and thus it was ‘not unreasonable to place on them the burden of producing an explanation which is at least consistent with absence of fault on their part’. As the respondents failed to do so, they had not displaced the inference that their negligence caused the accident.

Applying the principle to machine learning technology is not so simple. The difficulty lies in showing that the only reasonable explanation for the incident is negligence on the part of one of more producers of the technology. It may sometimes be possible to apply the principle to technologies such as self-driving cars because there will be sufficient evidence from vehicle logs to assess whether the type of accident is one where it is most likely one of the technology producers had been negligent. Thus for example, if the car fails to stop at a red traffic light and its logs demonstrate that the machine learning control module gave no instruction to stop, it is reasonable to require the technology producers to explain how that failure occurred. Such a failure is one that any driver should have guarded against and thus should have been anticipated and tested by the technology producers. By contrast, the reported facts about the recent Tesla fatality are that the circumstances were highly unusual. Even assuming that a human driver would have recognised the presence of the other vehicle, there is likely to be evidence that the Tesla Autopilot drives better than humans in most other circumstances. What is the appropriate standard required here to displace the application of *res ipsa loquitur*? On an aggregate view, a technology which drives better than the average human can hardly have been produced negligently. But if the accident is viewed in isolation, here the technology performed less well than a human would have done. Is it necessary that machine learning performs better than humans in all possible circumstances, or is it sufficient that on average it performs better? We cannot predict what a court would decide here.

---

48 Clerk & Lindsell on Torts (21 Ed), 8-184, citing Scott v London and St Katherine Docks n 47.
49 Ibid.
53 See n 24.
If evidence as to causation is available, in many instances it will raise the problem of multiple causation, where either the loss or damage has multiple causes and it cannot clearly be said of any of them that they were the ‘but for’ cause of the loss, or where there is evidence that multiple defendants acted negligently, but it cannot be shown which act of negligence was the ‘but for’ cause. In *McGhee v National Coal Board* Lord Wilberforce explained the justification for departing from the ‘but for’ test in cases where it is impossible to prove that the breach of duty caused or contributed to the injury sustained because of a lack of knowledge about how these causes interacted:

> if one asks which of the parties, the workman or the employers, should suffer from this inherent evidential difficulty, the answer as a matter of policy or justice should be that it is the creator of the risk who, *ex hypothesi* must be taken to have foreseen the possibility of damage, who should bear its consequences.\(^55\)

One way of dealing with the difficulty is by applying a test of ‘material contribution to injury’ as an exception to the but for test.\(^56\) The courts employ this concept where, on the balance of probabilities, at least one defendant’s breach of duty affected the claimant’s outcome, and even if some of the other causal factors were not the result of negligence. *Bonnington Castings v Wardlaw* is the archetypical example of an injury which was potentially caused by multiple factors. The employee contracted pneumoconiosis during his employment with Wardlaw. The disease was found to be one of ‘gradual incidence’ which was caused by an accumulation of exposure to silica dust. There were two sources of the silica dust, from grinders and from pneumatic hammers, and both of these operated concurrently. The defendant had only breached its duty of care in relation to the second source of silica dust. The court held that Bonnington had proved that it was more likely than not that Wardlaw’s breach of duty materially contributed to his illness. This established sufficient factual causation between the negligent conduct of the defendant and the damage sustained by the claimant.

In *Bonnington* there was a clear understanding of how the illness could be caused by exposure to silica dust, and thus a causal link could be established. However, where the process of causation is not known with clarity, this test cannot be applied. Instead, the courts have developed the test of ‘materially increased risk’. The test was first developed in *McGhee* and was expanded upon three decades later by *Fairchild v Glenhaven*.\(^60\) The claimant in *Fairchild* developed mesothelioma and had worked successively for three different employers, all of whom had exposed him to asbestos in the course of his employment. All three employers had breached their duty of care to prevent the claimant inhaling asbestos dust or fibres because of the known risk of causing mesothelioma. The causation problem was that medical knowledge was unclear about how exactly mesothelioma evolved. The medico-scientific community did not know what level of exposure to asbestos triggered a risk of developing mesothelioma, and whether a single asbestos fibre, a few, or many, could cause the condition. Medical opinion held all of these explanations equally likely. All that was certain was that once the disease was caused it

---

\(^{54}\) [1973] 1 WLR 1.

\(^{55}\) Ibid, 6.


\(^{57}\) [1956] AC 613.

\(^{58}\) 625 per Lord Keith.

\(^{59}\) n 55.

\(^{60}\) [2003] 1 AC 32.
could not be aggravated by further exposure to asbestos. This meant that if Fairchild had contracted mesothelioma with his first employer any exposure to asbestos in the second or third employment would not have affected his condition. The Court described this problem as the ‘rock of uncertainty’ which made it impossible to determine which exposure contributed to the disease and which did not. But because each employer had materially increased the claimant’s risk of contracting the disease, all were found to have met the test for causation and were thus liable. This principle is also applicable to non-medical cases where a similar lack of knowledge prevents use of the ‘but for’ test.

In the case of a machine learning technology which has multiple producers the court would be likely to apply one or both of these tests if the producers were held to owe a duty of care. The material contribution to injury test would be used if, for example, the training dataset had been compiled negligently and that contributed to the loss or damage, even though it could not be shown how (if at all) the other non-negligently produced elements of the technology also contributed to the loss or damage. The materially increased risk test would be used if there was evidence that two or more elements of the technology had been produced negligently but it was not possible to explain which one of them had been the cause of the final decision.

2.2.3 Conclusion

The advent of machine learning technologies is likely to present three very difficult challenges for laws imposing liability, particular where liability is based on negligence.

First, the technology changes the allocation of responsibility, and thus duties of care. The primary responsibility for the safe driving of road vehicles switches from drivers to technology producers, resulting in proof of breach becoming substantially more difficult and expensive. Professionals who use machine learning technology may well have discharged their duty of care by relying on the technology to make decisions, and in the current state of the law technology producers may owe no duty of care at all to those who are victims of technology failures, irrespective of how careless they were in their production. Doubtless the law of negligence would evolve over time to cope with these problems, as it has evolved before to cope with new technologies, but this evolution will take decades. Machine learning technology is likely to be adopted rapidly if it is effective. Once this happens the population as a whole will be at risk of loss or damage, not merely early adopters. The time lag for evolution of the law of negligence may be too long.

Second, machine learning challenges the law’s current approach to failure of responsibility, in particular breach of duty of care. This problem is partly an evidential one, because of the difficulties in explaining after the event why a machine learning technology made its decision, and also because of the complex interaction between the different elements of a machine learning technology. It will require the courts to take a new approach to the standard of care required.

Currently the standard of care required is of a reasonably careful human undertaking the activity in question. But how is that standard to be applied, for example, to machine learning technologies in the field of medicine? Under the current law the relevant standard expected

---

62 Indeed, this was explicitly recognised by Lord Bingham in Fairchild, [34]: ‘It would be unrealistic to suppose that the principle here affirmed will not over time be the subject of incremental and analogical development. Cases seeking to develop the principle must be decided when and as they arise.’
from a technology producer would be that of a reasonably competent software engineer. This feels wrong intuitively and may not be acceptable to society at large. Should the standard of care required by that of a reasonably competent doctor? Or should the two standards be combined, so that the court posits a reasonably competent doctor who is also a software engineer? The answer to these questions is likely to be based on the policy considerations underlying the imposition of a duty of care, fairness, justice and reasonableness, and can only be given in response to specific facts and not yet as a general proposition.

Perhaps more importantly, is comparison to a reasonably careful human the correct standard for machine learning technologies? Machine learning aims to make decisions to a higher standard than the average human. The same policy considerations will be used by a court to decide whether it should be held to that higher standard when addressing the question whether reasonable care was taken in its production, or in the decision to implement its use outside the research lab.

Finally, the law will need to align itself with society's concepts of fairness and justice. At present, the law requires those who have a responsibility to avoid causing loss or damage to others to be able to explain themselves in terms of human competence. Machine learning competence is not expressed this way — instead, its use is justified in largely statistical terms. Is this justification good enough to satisfy society’s demands, and if not what explanations should the law require from machine learning technology?

3 Responsibility and the wider interests of society

3.1 Individual autonomy

Autonomy is a fundamental principle under which individuals should be entitled to make their own choices as an act of will rather than having those choices made for them and forced upon them. This principle links to the doctrine of natural rights, which recognises the individual’s actions in expressing his or her will as an aspect of that individual’s natural freedoms. As a consequence, the law recognises such autonomous actions as being particularly valuable for society, given that they represent the subject’s will and freedom.

Initially the law concentrated on conferring individual autonomies and resolving conflicts between them. Thus in the 18th and 19th centuries the principle of party autonomy contributed to the increase of commerce and manufacturing, primarily through contract law which enabled individuals to choose with whom they wanted to undertake commercial activities and freely agree on the terms of their dealings.

But as society evolved towards its modern form, driven by social, political and economic changes, the law’s conceptualisation of autonomy has needed to change with it. The transition from liberalism to an interventionist state, for example, moved the legal focus from managing the clash of individual autonomies to questions about how far the public interest should be permitted to limit an individual’s autonomy. This reflects Rousseau’s social contract theory, which acknowledges the idea that ‘the man is naturally free, however the life in society reduces such freedom, which has to be consented within the

---

63 See text to n 29.
64 Alex Weil, François Terré, Droit Civil, Les Obligations (Paris: Dalloz 1975) 51.
limits and conditions that [social] contract determined. According to this theory, public authority is based on the agreement of individuals to form a society, a social contract, and thereby those individuals agree to restrain some of their rights in favour of the society. Society could not coexist if all subjects exercised the maximum extent of their autonomy.

How far individual autonomy should be restricted within a society is inevitably a matter of debate. Rousseau’s social contract theory attributes a utilitarian character to the subject within the society, so that society’s needs always override an individual’s autonomy, whereas Kant and Hegel take the opposite view, that society ought to guarantee the fulfilment of individual human freedom. Kant asserts that individuals are the ‘end in itself’, they are not merely tools or devices by means of which we can satisfy our own goals or purposes, individually or collectively; instead, they are valuable in themselves. Hegel recognises Rousseau’s fundamental contribution, that a union of individuals within the state thus becomes a contract, which is accordingly based on their arbitrary will and opinions, and on their express consent given at their own discretion, but denies, as Kant does, that the subject is just an instrument of society. Each society needs to find its own balance between these viewpoints and express it in the law.

As a consequence, the law plays two distinct roles in the preservation of individual autonomy. On the one hand it places limits on individual autonomy in the interests of wider society as a whole. This includes limits which are designed to protect weaker parties such as consumers, because without those protections the weaker parties’ autonomy would be limited to choosing between the terms offered by stronger parties. Individual autonomy is also limited to preserve wider societal values; for example, freedom of agreement between commercial organisations is regulated by competition law to ensure that free and fair competition is maintained.

On the other hand, the law protects individual autonomy, even against the interests of wider society, by granting fundamental rights to individuals. All of these provide substantial protection for individual autonomy. The clearest example is the right not to be held in slavery or servitude under Article 4 of the United Nations Declaration of Human Rights 1948, because slavery is the complete deprivation of an individual’s autonomy. Similarly the rights to privacy in Article 12 and freedom of expression in Article 18 prevent particular restrictions on individual choice and freedom of action, such as preventing them from communicating with others, or constraining their activities in private through fear of disclosure. Even those rights which are not obviously designed to protect autonomy have the practical effect of doing so; thus the right to social security under Article 22 provides individuals with the resources without which they could not in practice act autonomously.

We can see that autonomy is a core value, both at an individual and a societal level, which is embedded deeply in the fabric of the law. But decisions taken by machine learning

---

66 Ibid, 220.
69 Neil Cooper (1988), The Formula of the End in Itself, Philosophy, 63, pp 401-402.
70 See note 57 above, 244.
72 See Westphal, n 68, 243.
technologies are, by definition, designed to replace human decisions. This removes autonomy from the former decision-maker, and potentially results in a reduction of individual autonomy for those affected by the decision.\textsuperscript{74} So we need next to examine the nature of such reductions, and to consider how the law should react to them.

3.2 Machine learning and the preservation of autonomy

3.2.1 Risks to autonomy

Machine learning presents a particular problem for the preservation of autonomy because the choices it makes are based on what it has learnt from its training data, and not on any particular principles. These choices will rarely be explicit, and are usually completely invisible to the technology user – the technology just works, apparently, and precisely what it is doing and how it is doing it is unknown. Sometimes the workings of machine learning may even beopaque to the technology producer, as in the case of black box learning systems like neural networks.

Control of information sharing

Digital technology has greatly enhanced some aspects of individual autonomy. It is now possible for an individual to communicate globally to an unlimited number of recipients, and similarly to receive information from anyone, anywhere in the world. For most, though, this freedom is perhaps more theoretical than real, because the volume of information communicated digitally is so great that most communications reach only a small audience.

To cope with this problem, the providers of communication platforms have increasingly adopted machine learning technologies so as to filter out those communications which they believe an individual will not be interested in receiving. For example, Facebook uses a machine learning algorithm to curate each subscriber’s News Feed.\textsuperscript{75} Initially this algorithm worked mainly on the level of connection between a post and the individual (for example, assuming that you are interested in what your Facebook Friends are interested in) and, once the ‘Like’ button was introduced, how popular a post is (again assuming that popularity among Friends is most important). New measures of relevance to the individual are constantly added – for example, the algorithm now identifies the time spent viewing a post and not merely whether someone ‘Liked’ it – and research into how well these measures match subscribers’ actual level of interest is continuous.

[The developers] at Facebook have taken pains to avoid putting their own editorial stamp on the news feed. Instead, their working definition of what matters to any given Facebook user is just this: what he or she would rank at the top of their feeds given the choice.\textsuperscript{76}

The consequence is that most of the content on Facebook is hidden from each individual. ‘Facebook says the average user has access to about 1,500 posts per day but only looks at 300.’\textsuperscript{77} This is a substantial reduction in individual autonomy, but something like it is

\textsuperscript{74} Thus a smart refrigerator, discussed below, might limit one’s choice of diet without this being easily apparent.

\textsuperscript{75} For an accessible description of the development of this algorithm see Victor Luckerson, ‘Here’s How Facebook’s News Feed Actually Works’ \textit{Time Magazine} 9 July 2015.

\textsuperscript{76} Will Oremus, ‘Who Controls Your Facebook Feed?’ \textit{Slate} 3 January 2016,\url{http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html}.

\textsuperscript{77} Luckerson, n 75.
necessary because the autonomy to access all Facebook posts would be largely meaningless given the huge volume of communication on Facebook. It has potentially serious consequences for individuals and for society as a whole, in particular through the creation of ideological silos in which individuals receive only the kinds of information which reinforce what they already believe rather than challenging those beliefs. Facebook is attempting to restore some autonomy to its subscribers by enabling them to make choices about their News Feed, eg by allowing them to set priorities among their Friends and to identify posts which contain the kinds of material they don’t wish to see again, but inevitably these tools can produce only a limited enhancement. And any changes which Facebook deliberately makes to its algorithm can produce unintentional, or even intentional reductions in autonomy.

Control of individual behaviour
Control of information influences human decisions, and therefore affects their behaviour indirectly. But where machine learning controls physical devices it can control human behaviour directly.

One example of machine learning which is commonly discussed is the smart refrigerator. This technology would recognise the refrigerator’s contents from their RFID embedded chips and identify food consumption patterns as items are removed and replaced. One of the aims is to enable the refrigerator to re-order items which it has learnt will be wanted in the near future. But it is only a small step, and an obvious one to a nutritional scientist, to use the information about food consumption patterns to identify health risks arising from the individual’s diet, and then to modify the ordering of food so as to reduce those risks. Should the law permit this kind of behavioural control in the absence of express consent by the individual?

Another machine learning problem which has generated substantial interest is the choice which has to be made when it is inevitable that a self-driving vehicle will crash. Let us suppose that the crash can be managed in two ways; one way will kill the occupant, whereas the other way will kill one or more pedestrians on the pavement. How will this choice be made? Should the technology grant the occupant autonomy to take over and decide who lives and dies, or should the technology itself make that decision? And if it does so, is society content that it should do so on the basis of a utilitarian calculus, for example by estimating the economic value of the lives which will be extinguished? Dietary choices might be

---

79 Oremus, n 76.
80 For example the recent experiment to study the emotional effect of biasing individual News Feeds towards positive or negative content – Adam D I Kramer, Jamie E Guillory & Jeffrey T Hancock, ‘Experimental evidence of massive-scale emotional contagion through social networks’ (2014) 111 Proceedings of the National Academy of Sciences 8788, http://www.pnas.org/content/111/24/8788.full
something which should lie outside the law, but life and death is always a matter of legal interest.

3.2.2 Regulation to preserve autonomy?

We saw in part 3.1 that the law currently regulates those actions by humans, organisations and governments which are likely to have the effect of reducing individual autonomy. Should we consider regulating machine learning technology on a similar basis?

Lessig has argued forcefully that computer code can be used to control human behaviour. From this he concludes that code often acts in the same way that law acts, and so it should meet the same minimum standards as law:

If code is a lawmaker, then it should embrace the values of a particular kind of lawmaking. The core of these values is transparency. What a code regulation does should be at least as apparent as what a legal regulation does. Open code would provide that transparency – not for everyone (not everyone reads code), and not perfectly (badly written code hides its functions well), but more completely than closed code would.

Some closed code could provide this transparency. If code were more modular – if a code writer simply pulled parts off the shelf to plug into her system, as if she were buying spark plugs for a car – then even if the code for these components was closed, the functions and regulation of the end product would be open. Componentized architecture could be as transparent as an open code architecture, and transparency could thus be achieved without opening the code.84

Once the effects of code are understood, Lessig says that the law can respond through legislation and judicial decisions to restore individual autonomy so far as seems necessary.

Lessig’s argument here is persuasive, but there are clearly substantial difficulties in regulating machine learning technologies as a general class to require such transparency. To mention briefly only the most obvious:

- Machine learning as a discipline is young, and the technologies it utilises are constantly changing. Without some years of real-world experience of the use of those technologies it is debatable what transparency might mean, let alone how it could be achieved. The history of legislating prospectively for the digital technologies is one of almost complete failure.85
- Any regulation in one country would affect the producers of the technology, who are likely to be scattered around the world. A patchwork of differing national laws, each requiring different matters to be disclosed, is a recipe for confusion. Ideally a closely approximated set of national laws would be agreed and enacted, but such a project would require decades to achieve.
- Much machine learning technology is likely to be hosted in the cloud, whilst its outputs are consumed globally. This would in practice make foreign users reliant on the national regulators of favoured hosting jurisdictions to monitor and enforce

---

83 Though even this is not obvious given the costs to society of, eg, diet-induced Type II diabetes.
Some form of regulation at the national or regional level might be necessary to permit the use of machine learning for activities which are particularly risky, such as self-driving vehicles, but such regulation would still face the difficulty of identifying which aspects of a fast-developing technology are appropriately regulated.

And of course, machine learning technology is already subject to law and regulation which preserves autonomy, even though it was not produced with machine learning in mind. Most countries have a range of laws which protect fundamental rights – the most obvious are those which prohibit discrimination on the grounds of race, sex, sexual orientation, religion, etc. These laws will generally not apply to machine learning technologies directly, but rather to those persons who are using the technology to assist them in carrying out another activity.

A recent decision of the Wisconsin Supreme Court suggests that, so far as fundamental rights are concerned, it should not be permissible to make decisions solely on the basis of a machine learning technology which cannot give a satisfactory account of its actions. In *State of Wisconsin v Loomis* it was alleged that the use by a criminal court of a machine learning tool which predicted the likely risk of a defendant reoffending invalidated the court’s sentencing decision. The reasoning used by the tool was not disclosed because it was claimed to be proprietary confidential information, and Loomis argued that as a consequence his constitutional right to challenge evidence had been breached, and also that the tool used gender in making its predictions which thus contravened his right against sexual discrimination.

The Supreme Court found that the trial court had made its sentencing decision on other evidence, and had merely used the tool’s predictions to check that there was no reason to impose a lesser sentence, and so confirmed that decision. However, the Supreme Court was sufficiently concerned about the risks to fundamental rights in future cases that it issued detailed guidelines about how courts should use the tool. Three points in particular are worth noting.

First, the producer of the tool had made it clear that the tool was intended to identify whether a defendant was part of that group for which a supervisory, rather than a custodial sentence, had proved effective in the past. The Supreme Court held that it was permissible to use the tool’s output as the basis for imposing a supervisory sentence because that was the purpose for which it had been designed, and there was sufficient evidence that the tool’s predictions were accurate enough to justify such a decision.

---

86 Even within harmonized systems such as that of the EU’s data protection laws, problems arise from the split of regulatory control across national borders. See eg the French CNIL Decision no 2016-007 of 26 January 2016 issuing formal notice to Facebook Inc and Facebook Ireland, [https://www.cnil.fr/sites/default/files/atoms/files/d2016-007_med_facebook-inc-facebook-ireland-en.pdf](https://www.cnil.fr/sites/default/files/atoms/files/d2016-007_med_facebook-inc-facebook-ireland-en.pdf), which requires both corporations to comply with French law even though, for Facebook Ireland, the primary data protection regulator is the Irish Data Protection Commissioner.

87 For an early regulatory proposal see Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL), European Parliament Committee on Legal Affairs 31 May 2016).

88 (2016) WI 68.

89 Ibid paras 87-92.
Second, if the tool were used as the main basis to decide the quantum of sentence, eg the amount of a fine or how long a prison term should be, that use would be improper. This was because the tool would be being used outside its intended purpose, and also because the workings of the tool were not disclosed so that a sentencer who relied on it would not be giving proper consideration to the fundamental rights of the defendant.\(^9^0\)

Finally, the Supreme Court held that the guidelines it had laid down should be kept under review. In particular it noted that research into the use of the tool had already identified some limitations, for example that it favoured white defendants over black defendants, and thus required those responsible for the guidelines to keep further research under review and warn sentencers of the implications of new limitations as they were identified.\(^9^1\)

We think that the principles set out in \textit{Loomis} are likely to prove attractive to a court which is faced with a liability claim based on a breach of fundamental rights. Thus, for example, if an employer made its shortlisting decisions on the basis of a machine learning profiling tool, and was alleged to have discriminated unlawfully against an applicant as a consequence, the court would ask whether the tool was designed to make such decisions, whether there was evidence that it made them without unlawful discrimination, and whether its limitations had been properly investigated and taken into account when making the decision. Failure to provide satisfactory answers would make it very likely that the liability claim would succeed, and if so the person who was liable would want to seek compensation from the provider of the machine learning technology which led to the breach.

These difficulties in preserving fundamental rights arise because of the difference in perspective between laws which protect human rights and machine learning technologies. The former take the perspective of the individual as their starting point – was that individual treated fairly and reasonably in the light of the fundamental rights they possess under the law? By contrast, machine learning technologies examine past decisions, and from those decisions considered in aggregate they develop the model for making their future decisions. This latter approach can easily lead to a violation of fundamental rights.\(^9^2\) We saw earlier\(^9^3\) that there is a statistical correlation between the sex of a driver and the risk of that driver making an insurance claim. A machine learning technology which was deciding insurance premiums based, in part, on past claims, would inevitably charge a lower premium to a woman than to a man with otherwise identical characteristics. This would render the insurer in breach of the law, and liable to compensate the man discriminated against. Without some explanation of how the technology is reaching its decisions, it is not possible for that insurer to be certain that it is complying with the law. Quite how the producers of machine learning technologies could integrate an understanding of fundamental rights into their systems is a mystery, at least to lawyers, but there are many areas of activity which might usefully employ machine learning where such an understanding will be necessary.

The risk of infringing fundamental rights is not limited to decisions which affect only individuals, but has the potential to arise in the development of public policy if machine learning is used in the process. As an example, immigration policy might be developed on the basis of the potential economic benefit of accepting an immigrant into the country, using machine learning to analyse the economic benefits from previous immigrants and

\(^{9^0}\) Ibid paras 93-97.

\(^{9^1}\) Ibid paras 98-101.

\(^{9^2}\) See UK House of Commons Science and Technology Committee, \textit{Robotics and artificial intelligence} (HC 145 12 October 2016) paras 47-49.

\(^{9^3}\) See n 5.
thereby derive a model for future immigration admissions. Such a utilitarian approach might well end up favouring women over men, or one race over another, without this being apparent to the policymaker.

Brownsword identifies the dangers of a utilitarian approach, where any minority’s claims are at risk of being overridden in the interests of the greater good, and proposes that the better approach is that of a ‘community of rights’, where ‘the substantive moral approach embedded is rights-led, being committed to the protection and promotion of individual rights’. Mere proceduralism in legal policymaking is not enough on its own – those whose fundamental rights are adversely affected:

\[
\ldots \text{will need to be satisfied that the regulators have made a conscientious and good faith attempt to set a standard that is in line with their best understanding of the community’s rights commitments. Regulators do not have to claim that the standard set is right; but, before a procedural justification is accepted, regulators must be demonstrably trying to do the right thing relative to the community’s particular moral commitments.}
\]

In order to demonstrate this where machine learning has been used, the policymaker will require an explanation of why and how the machine learning technology makes its decisions. This, again, demonstrates a need for transparency of reasoning in machine learning technologies.

### 3.3 Accountability and autonomy

Transparency is an aspect of the wider concept of accountability. The Accountability For Cloud (A4Cloud) project has identified five core Accountability Attributes:

**Transparency**: the property of a system, organisation or individual of providing visibility of its governing norms, behaviour and compliance of behaviour to the norms.

In the context of this paper, transparency requires that the reasons how and why a machine learning technology made its decisions can be understood, and also that the accuracy and comprehensiveness of its training data can be assessed. Transparency is clearly the most important accountability attribute for liability questions, and the question about how increased transparency might be achieved is explored in part 4 below.

**Responsiveness**: the property of a system, organisation or individual to take into account input from external stakeholders and respond to queries of these stakeholders.

We saw in part 2 that there are circumstances in which the user of a machine learning technology will be liable for loss or damage caused as a result of its use, and for this reason alone there is an incentive for technology producers to be responsive to those liability concerns and provide assistance in coping with liability issues (which might well be via increased transparency).

---

95 Ibid, 24.
96 Ibid, 127.
**Responsibility:** the property of an organisation or individual in relation to an object, process or system of being assigned to take action to be in compliance with the norms.

Responsibility, in the A4Cloud taxonomy, means the recognition that there are actions which an organisation ought to undertake in response to problems arising. Responsibility might be imposed by law or might be voluntary, for example through a publicised agreement between the producers of different elements of a machine learning technology as to which of them will take the lead in attempting to resolve problems.

**Remediability:** the property of a system, organisation or individual to take corrective action and/or provide a remedy for any party harmed in case of failure to comply with its governing norms.

Remediability might include accepting legal responsibility, usually in the form of liability to compensate for loss, but accountability equally puts emphasis not only on whom to blame but how to repair the damage. This latter point is important, because lawyers have a natural tendency to focus on liability (in terms of compensation or punishment) and thus ignore other ways in which a satisfactory remedy might be offered. It is noteworthy that successful businesses in the digital arena tend to offer a high level of non-financial remediation, and this helps to reduce liability claims as well as enhancing reputation.

**Verifiability:** the extent to which it is possible to assess compliance with accountability norms.

Verifiability is important because accountability requires both the defining and displaying of relevant norms, and behaviour which demonstrates compliance with those norms. An example of verifiability which would be highly relevant to liability questions might be documenting the ways in which a training dataset was designed, structured and collected, as this would help demonstrate the degree of care used in producing the technology and also how far fundamental rights have been taken into account.

The transparency attribute of accountability is particularly important to the liability and individual autonomy issues discussed in this paper because, for both, many of the legal difficulties are caused by the apparent ‘black box’ nature of machine learning technology. Transparency, if it could be achieved, would help to explain the options available to the technology and the choices it made between them. This alone would be of great assistance in helping to resolve liability questions. Ideally transparency would go further and explain how and why the technology made these choices, ie its reasoning. The current legal liability schemes discussed in part 2, particularly the law of negligence, are very dependent on evidence of causation, and require explanations (accounts) to be given of how and why loss and damage were caused.

Accountability, with transparency playing the main part, can also help to assure users of machine learning technology and those affected by machine learning decisions that the fundamental rights of individuals are being respected in the decision-making process. As for liability, accounts of how and why the decision was made can help provide the justifications which the law requires.

From a legal perspective it is unfortunate that machine learning development has, naturally, focused on the quality of learning and decisions rather than on providing accounts of how decisions are justified. But there is a growing body of research into how explanation

---

98 UK House of Commons Science and Technology Committee, n 92, paras 52-58.
facilities can be incorporated into machine learning,\textsuperscript{100} and also research into how explanations for the reasoning of ‘black box’ technologies can be reverse engineered retrospectively.\textsuperscript{101} Additionally, if the data used as input to a decision is captured and logged, the technology should produce the same decision if that data is used as input again, which might allow some understanding of a particular decision to be developed by, eg, an expert witness in litigation. Developments in all these areas may help resolve this difficulty over time.

4 Some tentative conclusions

It is only possible to draw two firm conclusions from the discussion above. The first is that it is far too early to devise a liability regime for machine learning generally because in the current state of development of the technology the law would rapidly fail to accord with technological change. The second is that, both in respect of liability and the fundamental rights which underpin individual autonomy, society will need to make some difficult choices, because if machine learning is to be adopted widely then the existing legal and regulatory settlement will not cope adequately. Whilst not suggesting what the final legal and regulatory settlement should look like, our analysis above suggests that it would most effectively focus on the training and testing of machine learning technologies. Attempts to regulate the decision-making algorithms which emerge from these processes will be aiming at a moving target, because the algorithms will change constantly as a result of continuous learning.

More tentatively, we can make a number of further suggestions.

It seems clear to us that a liability regime under which there was potentially no person who had responsibility (and thus liability) for the consequences of a machine learning decision would be incompatible with the way society is currently organised. The current system of liability laws focuses mainly on human failings, assigning liability on the basis of and in proportion to those failings. Machine learning failings are different in kind to human failings, and assigning responsibility to humans for those failings on the current legal basis would be faced with problems of shared responsibility and lack of knowledge about how the machine learning was trained and the basis on which it makes its decisions. Resolving these under the current law would be likely to require imputations of knowledge to the persons using or owning the technology which are obviously untrue.\textsuperscript{102}

Many of these problems would be much more easily soluble if machine learning technology could explain itself, by giving an account of how the loss or damage occurred, the reasoning

\textsuperscript{100} For a useful review see Keith Darlington, ‘Aspects of Intelligent Systems Explanation’ (2013) 1 Universal Journal of Control and Automation, 40.


\textsuperscript{102} For example, in order to hold a human doctor liable in negligence for loss or damage caused by failures in a machine learning medical diagnosis system which had, apparently, been tested robustly, the court would need to impute to the doctor knowledge about how the system came to its decisions, so that the doctor’s failure to recognize that the decision was incorrect would amount to a breach of duty.
which led to the machine acting as it did, and the processes through which the machine learnt to act in that way. It is thus tempting to propose some regulatory solution which requires such accountability mechanisms to be built in to all machine learning technology.

But we need to recognise that even the lightest regulation has a potential chilling effect on the development of new technologies. It is surprisingly easy to regulate a developing technology out of existence. And requiring accountability to be built in to all machine learning technology would not be a light regulatory burden – some technologies are inherently resistant to providing accounts, and where accountability would be possible it will often reveal commercially confidential information about the working of the technology, leading technology producers either to refuse to exploit their work or to introduce unnecessary and expensive restructuring to enable them to comply with regulation without disclosing trade secrets. Even where neither of these applies, providing an account is not cost-free.

We suggest that there is a continuum of machine learning technologies, and that the proper approach of the law changes along this spectrum. At one end accountability might be not merely desirable, but essential to ensure that the fundamental rights of individuals and society’s wider interests are not damaged. At the other, the focus should instead be on ensuring that loss or damage is compensated even if no accountability is practicably achievable.

Self-driving vehicles clearly reside at the latter end of the spectrum. It is apparent, even at this early stage of development, that they are (viewed in aggregate) safer than vehicles driven by humans. Their introduction to the roads would thus result in a real social benefit. Vehicles on roads create inherent risks of death, injury and property damage, whoever or whatever is driving, but providing an explanation of how loss or damage was caused does not assist the victim to any great degree. So, if adequate compensation is available to victims, even in the absence of accountability the law should in theory be satisfied. Even so, we are already seeing demands for accountability in this area. The recently-published US Federal policy document sets out guidance for the automobile industry which, in addition to requiring good design and implementation practices, introduces obligations for accountability to be designed into the technology:

For all [autonomous vehicle] systems, the manufacturer or other entity should address the cross-cutting items as a vehicle or equipment is designed and developed to ensure that the vehicle has data recording and sharing capabilities; that it has applied appropriate functional safety and cybersecurity best practices; that HMI design best practices have been followed; that appropriate crashworthiness/occupant protection has been designed into the vehicle; and that consumer education and training have been addressed.

---

104 Federal Automated Vehicles Policy – Accelerating the Next Revolution in Road Safety (US Department of Transportation/NHTSA, September 2016).
105 The guidance has no legally binding status, and in theory only sets out the expectations of the US authorities. However, these are de facto binding because the NHTSA has legal authority to take action in respect of vehicle defects, to demand recalls and to enforce the law. The document makes it clear that the guidance set out in the policy document will be relevant in exercising that authority – ibid, 11.
These obligations fall into four main areas:

- Systems must record and share data in sufficient detail to allow explanation of why the vehicle behaved as it did if the consequence of that behaviour was death or non-minor damage to property.\(^{107}\)
- The design, development and training of systems ‘should be fully documented and all changes, design choices, analyses, associated testing and data should be fully traceable.’\(^{108}\)
- The vehicle while in use must explain its operating status to the human operator or occupant.\(^{109}\)
- The ethical decisions which the system makes, such as how to decide between inevitably injuring occupants or those outside the vehicle, must be articulated explicitly and developed with appropriate consultation.\(^{110}\)

It seems to us that these accountability requirements are not aimed at resolving the legal liability questions, but rather at reassuring the public that self-driving technology has been developed with public safety in mind, and in a way which allows problems to be identified and rectified. These social considerations may not apply to other uses of machine learning where the potential losses are perceived as less serious. In those cases, as we argued above, providing a liability remedy might be sufficient even if the technology cannot account for its decision-making failures.

By contrast, at the other end of the spectrum the risks are not of injury or damage but of breaching fundamental rights. The risk is that the machine learning technology unexpectedly discriminates against women, or prevents certain kinds of speech, or interferes in an individual’s private life. Here damages are not an adequate remedy, because allowing fundamental rights to be contravened at will so long as compensation is paid is, or ought to be, legally unacceptable. Of course breaches should be compensated, but more importantly

---

\(^{107}\) Ibid, section E 17 ff. See also Pathways to Driverless Cars, n 17, paras 2.23-2.24.

\(^{108}\) Ibid, 21.

\(^{109}\) Ibid, 22:

At a minimum, indicators should be capable of informing the human operator or occupant that the [autonomous vehicle] system is:

1. Functioning properly;
2. Currently engaged in automated driving mode;
3. Currently ‘unavailable’ for automated driving;
4. Experiencing a malfunction with the HAV system; and
5. Requesting control transition from the HAV system to the operator.

\(^{110}\) Ibid, 26:

Even in instances in which no explicit ethical rule or preference is intended, the programming of an [autonomous vehicle] may establish an implicit or inherent decision rule with significant ethical consequences. Manufacturers and other entities, working cooperatively with regulators and other stakeholders (e.g., drivers, passengers and vulnerable road users), should address these situations to ensure that such ethical judgments and decisions are made consciously and intentionally …

Since these decisions potentially impact not only the automated vehicle and its occupants but also surrounding road users, the resolution to these conflicts should be broadly acceptable. Thus, it is important to consider whether HAVs are required to apply particular decision rules in instances of conflicts between safety, mobility, and legality objectives. Algorithms for resolving these conflict situations should be developed transparently using input from Federal and State regulators, drivers, passengers and vulnerable road users, and taking into account the consequences of an HAV’s actions on others.
those whose use of machine learning technologies might lead to such breaches should be required to take care that breaches do not occur. If breaches do happen, steps should be taken to ensure they do not recur. None of this is possible if the machine learning technology cannot provide an account. And there are already legal incentives for technology producers to build accountability mechanisms into the technology, most notably Article 22 of the EU GDPR which grants data subjects a limited right not to be subject to ‘a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her’ and Articles 13 and 14 which proved that when profiling takes place using personal data the data subject has a right to ‘meaningful information about the logic used’.\footnote{See further Bryce Goodman and Seth Flaxman, ‘European Union regulations on algorithmic decision-making and a “right to explanation”’, paper presented at 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY, https://arxiv.org/abs/1606.08813; Dimitra Karaminou and Christopher Millard, n 4.}

It might be possible, once a machine learning technology has been in use for sufficiently long, to decide where it resides on the spectrum and to regulate it appropriately. But that day is some distance in the future.

Strict liability might be a workable solution where the use of machine learning creates only a risk of physical injury and property damage. A strict liability regime would be comparatively easy to design for self-driving vehicles because it is already compulsory for vehicles used on the public highway to be insured against loss or damage for which the driver is liable.\footnote{See eg the UK proposals in Pathways to Driverless Cars, n 17, part 2.} Thus a strict liability regime for self-driving vehicles would mainly have the effect of reallocating responsibility for insurance between those who might have liability under the current system. The overall insurance burden would probably rise at first because \textit{all} injuries and damage would be compensated, not merely those caused by negligence. But this might well be compensated for by the expected higher degree of safety which self-driving vehicles produce compared to human drivers. Deciding where to allocate the responsibility to insure might at first sight appear to raise difficult political and societal questions, but we take the pragmatic view that the only clearly identifiable person to whom that responsibility might be allocated is the keeper of the vehicle.\footnote{This approach is supported by the Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL), European Parliament Committee on Legal Affairs 31 May 2016) paras 29-31, though we take the view that the Report’s suggestion to introduce such a scheme to cover \textit{all} loss and damage caused by autonomous robots is somewhat unrealistic.} This follows the precedent for aircraft owners,\footnote{See n 10.} which appears to cause no insoluble problems in that sector. We express no view on whether the existing negligence liability scheme should continue to apply to claims by insurers against the producers of machine learning technologies, other than to say that maintaining it in place would secure a comfortable income to the legal profession because of the likely complexity of litigating such claims.

Whether there are other candidates for strict liability regimes is more difficult to assess. We have referred above to the use of machine learning in medical diagnosis and treatment, and the adoption of a no-fault system of compensation for personal injuries as in New Zealand\footnote{The scheme is run by the Accident Compensation Corporation – for details see http://www.acc.co.nz.} would certainly remove this from the list of problem areas. However, such a system would need to apply to all medical failures, not merely those resulting from machine learning technology, and might fail to accord with society’s apparent need to find a person who is at
fault and who can be held responsible; it is noteworthy that New Zealand’s scheme has not been followed fully elsewhere, though no-fault compensation schemes for medical injuries and road accidents appear to be acceptable in some countries.\textsuperscript{116}

How to protect the wider interests of society in preserving the fundamental rights of individuals against the risk that machine learning decisions are made on impermissible grounds is more problematic. Clearly the risk is reduced if each machine learning technology can account for its decisions. But attempting a general regulation requiring machine learning technology producers to design accountability into their systems is likely to be practically infeasible because of the risk that it would chill technology development by requiring the disclosure of trade secrets, and also because of the cross-border issues discussed in part 3.2.2. Self-regulation by technology producers, eg through codes of practice,\textsuperscript{117} would help to identify ways of overcoming the barriers to producing an account, and making liability more stringent where a technology cannot explain itself would encourage compliance with such codes. Additionally, the range of potential machine learning applications is so wide that it is not clear that any element of human life would remain untouched by machine learning. Thus any regulatory project would be of uncertain, but undoubtedly enormous, scope.

Before regulation is attempted, we need to allow machine learning technologies a period of use so that we can discover what legal problems they in fact create and how best to address them. This interim period could be dealt with by very minor adjustments to liability law, making liability for unexplained machine learning decisions more stringent as a means to persuade technology producers to build accountability into their technologies where appropriate. In negligence this might be achieved by creating a legal presumption that, where a person owes a duty of care and uses machine learning technology to undertake that person’s activities, any loss or damage caused resulted from the negligence of the user of that technology unless the user could provide an explanation of the technology’s workings sufficient to displace that presumption. In effect this would be an extension of the doctrine of \textit{res ipsa loquitur}, and as is the case under that doctrine the obligation on the person owing the duty would be to put forward an alternative explanation of how the loss or damage was caused. The likely consequence would be that users would seek out technology producers who had devised accountability mechanisms, and also have the social benefit that machine learning technologies which did not have those mechanisms, and thus presented unquantifiable risks, would be not be widely adopted. If a technology’s benefits were so clear that it was worth adopting even without accountability, users would insure against what would now \textit{de facto} be strict liability. This would at least mitigate the litigation problems which we noted earlier. In particular there would be no need for a claimant to produce evidence that the production or training of the technology had been faulty, but rather the law would put the onus on those who produce and use the technology to provide evidence\textsuperscript{118} that no negligence had occurred in its production.


\textsuperscript{117} There has already been a call for a code of practice for technology producers under which they would need to consider these, among other ethical issues - Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL), European Parliament Committee on Legal Affairs 31 May 2016) paras 6 & 7, Annex.

\textsuperscript{118} And thus there would also be an incentive for technology producers to preserve their training data and keep records about the machine learning process which might, through examination after the
Protecting fundamental rights during this interim period is more difficult. Even so, introducing a presumption that a machine learning technology failed to protect those rights appropriately, where there is at least a plausible allegation that a breach might have occurred, would make it easier to claim compensation for victims and thus incentivise technology producers to build in the appropriate protections. Clearly a long-term solution needs to do more than merely compensate for breaches of those rights, but experience suggests that in the early days of any new technology, using the law to encourage an interim market solution is likely to produce better (and quicker) results than regulation of the technology, and can help identify the problems which later regulation should most appropriately address.