Law
Commission
Reforming the law

AI and the Law

A Discussion Paper

This publication is available at https://lawcom.gov.uk/project/friendly-societies/

Contents

Introduction

The term artificial intelligence or "AI" was coined in the 1950s. Since then, AI has undergone periods of expansion and active development, and periods of middling progress. The 21st century has seen a surge of AI development, culminating in significant achievements, including:

1) The AI system "AlphaFold 3" predicts protein structures 50% more accurately than the best alternate traditional methods.[1] In 2024, the Nobel Prize in Chemistry was awarded (in part) to Demis Hassabis and John Jumper for their work on AlphaFold. In the same year, the Nobel Prize in Physics was awarded to John Hopfield and Geoffrey Hinton for their work on AI.

2) "AlphaGo" became the first AI system to defeat a professional player of the boardgame "Go".[2] This was a significant accomplishment, because the complexity of Go meant that creating an AI system to defeat professional Go players was orders of magnitude more difficult than the same for chess (which accomplishment was achieved in 1997 when Deep Blue defeated then world chess champion, Garry Kasparov).

3) Developers of self-driving vehicles achieved "Level 4" capability. Level 4 vehicles can operate autonomously in certain conditions, without any human intervention required. The company Waymo now provides autonomous vehicle rides for passengers in the United States—in Phoenix, San Francisco, and Los Angeles.

4) The development and release of large language models (LLMs), such as OpenAI's GPT models or Google's Gemini models (among others). Applications built with certain LLMs can produce text, image, and audio outputs, often with striking similarity to actual human outputs. ChatGPT, an LLM accessible on the internet, is now the 5th most visited website in the world.[3]

With the rapid development and improved performance of AI has come increased investment and wider and more frequent applications of it. AI is expected to deliver social and economic benefits, leading to increased productivity, boosting economic growth and output, and may lead to innovations that can save and improve lives, such as the development of new cancer drugs or new medical treatments. Taking advantage of those opportunities is a focus for Government, as set out in its AI Opportunities Action Plan, published in January 2025.[4] In 2025, Government also reached agreements with leading AI developers Anthropic, Google, and OpenAI to take advantage of opportunities offered by AI and explore increased investment in and use of AI.[5]

However, as with other technological developments, AI's potential to deliver benefits comes with risks that it will cause harm. AI has been used to perpetuate fraud, cause harassment, assist in cyber hacks, spread disinformation that harms democratic processes, and can create "deepfake" images of people as a form of abuse or to enable identity theft, among other examples.[6] There are also concerns that increased use of AI could cause harm by way of social upheaval, that AI will replace existing workforces, at scale, in a wide range of industries, from manual to highly-skilled.[7] Further concerns exist about the environmental impact of technology that is using an increasingly large quantity of energy and water.[8]

AI's potential to cause harm has led to increasing governmental and regulatory attention worldwide,[9] a prominent example being the European Union's Artificial

Intelligence Act (AI Act),[10] which provides a regulatory framework for AI. The AI Act came into force on 1 August 2024 but applies from 2 August 2026, subject to Article 113, pursuant to which certain provisions apply from 2 February 2025 and 2 August 2025, and the remainder apply from 2 August 2027.[11] In preparation for provisions regarding general purpose AI models applying from 2 August 2025, the EU has also published a General-Purpose AI Code of Practice. Among other things the Code of Practice provides guidance to help industry comply with certain AI Act requirements and to manage system risks of general-purpose AI.[12]

AI's potential to cause harm has also led to a growing volume of work aimed at meeting the challenges posed by AI, including legal challenges.[13] Society's collective legal systems have developed over thousands of years to guide and manage the behaviour of highly intelligent human persons, and subsequently, also non-natural legal persons. It is not yet clear that those same systems will apply equally well to new technology that is also intelligent to varying degrees.

While the expanded use of AI may have a range of social and economic consequences, this paper is focused on potential legal challenges. Some of those legal challenges are well-known. Numerous creative groups and creative workers have raised serious concerns about leading AI models being trained, in part, on works protected by copyright, potentially in breach thereof, without compensating creators.[14] The concerns of discrimination by way of AI bias are also well publicised. Among other things, there are concerns that biased AI systems might lead to discrimination, for example if an AI system is used in making a hiring decision or to support predictive policing.[15]

However, there are many other, lesser known, legal challenges raised by AI. As Lord Sales JSC stated in 2019, the topic of AI and the law is "a huge one".[16] In addition to the specific issues those challenges raise, legal uncertainty regarding AI may also delay the safe development and use of AI. For example, legal uncertainty can be an impediment to obtaining appropriate insurance, the lack of which can obstruct projects commencing and thereby stunt innovation. Further, if insurance cover is not in place, and harm occurs, people may be left without assistance or require Government assistance at public expense.

At the Law Commission, we anticipate that AI will increasingly impact the substance of our law reform work. It may be that AI will itself be the focus of a particular project in future, for example, considering specific questions about civil or criminal liability for acts or omissions of AI. In other cases, the prevalence of AI in a particular context may require analysis of issues raised by AI, for example, any work on intellectual property is likely to require analysis of intellectual property protections regarding AI and its outputs. In fact, the Law Commission has already completed work relating to or involving AI, with our project on automated vehicles[17] and with respect to deepfakes in our project on intimate image abuse.[18] We also have an ongoing project on aviation autonomy, as well as a pending project on product liability, which will consider, in part, AI.

Given the rapidly expanding use of AI, and the fact that the *rate* of AI technology development is increasing,[19] we consider it important that a wide audience be able to understand and engage with the legal questions and issues raised by AI. Therefore, the purpose of this paper is to raise awareness of AI and the law and to foster further discussion on such important issues. This paper is intentionally high-level and less detailed than our typical law reform publications and therefore does not contain within it proposals for reform. It is a step towards clarifying the large and complex field of AI and the law, to identify those areas most in need of law reform.

With that purpose in mind, this paper is structured as follows:

(1) We define what we mean by "AI" and briefly discuss how AI works.

(2) We discuss the following themes as a means for discussing how legal issues may arise with AI:

(a) AI autonomy and adaptiveness.

(b) Interaction with and reliance on AI.

(c) AI training and data.

Finally, while we do not propose options for reform in this paper, we note throughout that many of the legal issues raised by AI arise, partly, because AI does not have legal personality. Accordingly, we conclude by considering a potentially radical option for AI law reform: granting some form of legal personality to AI systems. Current AI systems may not be sufficiently advanced to warrant this reform option. But given the rapid pace of AI development, and the potentially increasing rate of pace of development, it is pertinent to consider whether AI legal personality requires further discussion now, in the event that such highly advanced AI arrives in the near future.

Various staff have contributed to this Discussion Paper. The central team was Laura Burgoyne (team manager), Michael Workman (lead lawyer, on secondment from Steptoe International (UK) LLP), and Saiba Ahuja (research assistant). The central team would also like to thank Colin Oakley, Connor Johnston, Laura Jones, Dr Nicholas Hoggard, Rob Kaye, Tusmo Ismail, and Christopher Long.

What is AI and how does it work?

There is considerable debate over how to define AI. The issue has proven intractable partly because there is "little agreement about what intelligence is".[20] Further, as AI technology improves and carries out more complicated tasks, there is a tendency to redefine those tasks as *not* requiring intelligence. John McCarthy, one of the founders of the field of AI, remarked: "As soon as it works, no one calls it AI anymore".[21]

Given the high-level nature of this paper, it is not necessary to enter into that debate. We adopt here a definition that enables discussion of certain elements of AI that raise potential legal challenges. A leading definition that is helpful for present purposes is that published by an expert working group of the Organisation for Economic Co-operation and Development:[22]

> An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

In particular, it is their autonomy and adaptiveness in making inferences that distinguish modern AI systems from other systems that have historically been labelled AI. For example, for much of the 20th century, the leading approach to AI development was "symbol-based" or "rule-based" AI (sometimes described as "Good Old-Fashioned AI").[23] Such systems store or encode specific human expertise in a system of symbols and rules (such as numbers and logical operators). Those symbols and rules determine how an output is generated from an input to the system. Such systems are not autonomous or adaptive in the same way as the AI models that are now pervasive. Their rules (or symbols) are provided to the system, not learned by the model in training. They are therefore more determinative—in that the same input produces the same output—and more predictable.

The modern success of AI is largely the story of the development of "machine learning". Machine learning is a form of AI technology that enables models to adapt or "learn", pursuant to which the model's performance improves over time on a given task. Such models learn themselves by way of processing data. For example, the most well-known AI models today are LLMs, such as OpenAI's GPT models or Anthropic's Claude models (among others), which are machine learning models trained on large volumes of natural language text (though, as discussed further below, AI models can be trained on other data, including images and audio).

The success of machine learning in the 21st century is due, in part, to the sharp increase in the amount of data available to train AI, as a result of the expansion of the internet and other digital technologies. Generally speaking, more training data has meant improved AI performance. To put the sheer amount of data used to train modern AI models in context, if a human person were to read the datasets used to train leading LLMs word-by-word, 24 hours per day, 7 days per week, it would take them thousands of years just to finish reading.[24]

Further, AI models can be trained on a wide range of data, not just natural language. A "mode" refers to the type of data a model computes. Natural language text is one mode. Audio data, such as speech, is another. Images and videos can also be used to train AI models (some early machine learning models were trained to improve image recognition).[25] A specific AI model is "multi-modal" if its inputs or outputs are of different modes. OpenAI's GPT-4o model "…accepts as input any combination of text, audio, image, and video and generates any combination of text, audio, and image outputs."[26]

An important feature of machine learning AI is that it is not tied strictly to one task or function. It has general application. Leading AI models are capable of being trained for, and completing, an array of tasks or solving an array of problems, based on training with multi-modal data. There are now AI metrics for measuring AI performance in areas as diverse as language, speech, coding, mathematics, imagery, reasoning and robotics.[27] This is what enables machine learning models to be used in many different areas or fields of application, referred to as "domains". An AI model used as part of an automated vehicle is operating in a different domain to an LLM used to help draft text and video for a company's marketing campaign.

One of the ways that AI is prepared for use in different domains is by using a leading AI model as a foundation which is further trained for more specific tasks. One method for doing this is called "fine-tuning". For example, a pre-trained AI model (such as an LLM) could be fine-tuned to emulate a specific writing style through additional training on sample texts representing that style. For this reason, the largest and most advanced AI models are sometimes referred to as "Foundation Models".[28] While most Foundation Models are developed by only a small set of organisations, they are made available to third parties who can (and do) further train those models as well as incorporate them into other systems for specific uses. As we will discuss below, the many different entities involved in the development and application of AI can raise questions as to who is "responsible" for AI outputs.

Machine learning AI models are becoming increasingly sophisticated and capable.[29] Whereas early versions could complete simple tasks such as recognising hand-drawn numbers,[30] advanced AI models today can complete tasks that require multiple steps with minimal human oversight.[31] For example, they are increasingly successful in completing technical tasks, such as reviewing medical images for the purpose of subsequent diagnosis. AI models are now used in a very wide variety of domains and their scope of application is only widening.[32] From a legal perspective, this means many different types of law might be impacted by AI.

However, machine learning AI models have limits. It is commonly quoted that for AI the hard things are easy, and the easy things are hard.[33] That is, it may be (relatively) easy to design systems capable of solving complex mathematical problems in superhuman timeframes, but incredibly difficult to build a robot that has the perception and mobility skills of a one-year-old human child. Further, while AI systems can generate text and images with remarkable detail and accuracy, this does not mean that those systems are capable of the kind of reasoning that humans do. AI models cannot "check" their answers based on an understanding of the world.[34] They routinely "hallucinate", being where an AI system provides an answer or information that is false, including producing outputs that have been "created" by the system, such as non-existent academic publications or works of fiction. Such hallucinations often appear highly credible, leading people into error. There are now numerous examples of lawyers using AI systems to conduct legal research, only to cite non-existent authorities in court.[35]

Another concern is "reward hacking". This occurs where an AI system seeks to optimise an outcome but not in a manner its developers intended, and potentially did not expect or want. This raises the concern that AI systems may develop without aligning with society's values and thereby cause harm.

For example, in recent research, a leading AI model developer tested 16 advanced AI models by allowing them to send e-mails and access sensitive information in a test business environment,[36] as a means of completing simple business tasks allocated to them by their fictitious governing company. The models were then tested for how they would respond when told by their governing company they would be replaced or where the company's objectives changed and conflicted with the AI model's assigned goals. In some cases, the models resorted to malicious behaviour, such as blackmailing fictional executives and leaking sensitive information to fictional competitors, even when expressly directed not to do this. Further, some models showed worse behaviour when told the test environment was real (as opposed to where they were told they were in a test environment).

Part of the reason for the difficulties in controlling and understanding machine learning models is that they are highly complicated mathematical functions, based on complex statistics. It is generally accepted that these AI models do not have conceptual models of the world in the way natural persons do. While they make inferences from inputs to generate outputs they do not "reason" in the same way as humans do.[37] The models themselves do not "understand" the functions to which they are put in the same manner that human persons understand their tasks.[38] This is significant in understanding why it is difficult, even for technical AI experts and developers of AI models, to know why machine learning produces the outputs it does. This difficulty in discerning why machine learning models produce the outputs they do is referred to as the "black box" nature of AI, or its "opacity".[39]

The above is a very short summary of machine learning and AI, but it is sufficient for our purpose of discussing certain features of machine learning AI that give rise to potential legal issues, to which we now turn.[40]

How might AI legal issues arise?

The impressive capabilities of machine learning technology represent opportunities for potential benefits and pose risks of potential harm. The general capabilities of machine learning also mean that those opportunities and risks are not limited to a narrow range of applications. They may arise wherever machine learning is deployed, which, increasingly, is almost everywhere. The potential legal issues raised by AI are equally broad. In some cases, it may be that the existing laws of England and Wales can manage the issues without reform, whereas other issues may require law reform. Our intention is that raising awareness of these issues will encourage future work to resolve those questions, both by the Law Commission and others.

In this section we consider certain key features of AI that potentially give rise to legal issues. Some of those features are features *of* AI (for example, autonomy) and some are not strictly features of AI itself, but are key parts of AI development, training or practical application (for example, the training of AI and use of data). The discussion below is very much the starting point concerning AI and law reform. The features raised, and the potential legal issues they raise, are not exhaustive. Much work has already been completed pertaining to AI and the law,[41] and much more work is underway. This section is intended to raise awareness of some of the issues raised in those works and to encourage further discussion of this important topic.

Autonomy and adaptiveness

Autonomy refers to the ability of an AI system to complete objectives with limited or no human input, control or oversight.[42] For example, when AlphaGo defeated professional human Go players in 2015, the AI system itself determined what moves to make in the game. It did not rely on human direction (although it could not move the physical Go pieces on the board, so it did require human assistance to make its moves in that setting).[43]

Adaptiveness is the ability of AI to learn and evolve and thereby change its outputs over time. AlphaGo adapted over time to play Go at increasingly higher levels and ultimately defeat professional players. The strategies it deployed were not expressly taught to it. In fact, in one match, AlphaGo deployed a strategy that baffled expert Go observers, who considered it an unwise move. It became apparent later in that game that the move was a prudent, and ultimately successful, strategy.[44] Further, while the original AlphaGo model was trained playing humans and playing against different versions of itself, subsequent, more advanced, versions were given only the rules of Go and were then trained playing themselves. Contrast this with the rule-based systems discussed above, where the rules of the system are provided by its developers.

AI systems are becoming increasingly autonomous and adaptive. Leading AI developers are researching and developing “AI Agents”.[45] The goal of AI Agents is to execute complex multi-step tasks with no or minimal human input. Consider a person giving a direction to an AI Agent to plan and book a holiday. The AI Agent would break down that goal into discrete steps (choose a holiday location, search for accommodation, book travel, factor in a limited budget and so forth) and then take the steps required to achieve that goal (book flights and accommodation online via a web browser, download boarding passes to digital wallets, prepare an itinerary, and so forth). This is a simple example, but the goal of increased research and development of AI Agents is that they undertake increasingly complex and difficult tasks. Versions of AI Agents already exist and are available to use.[46] As this technology develops, it is also likely that AI Agents will begin interacting with other AI Agents.

In the holiday example, the AI Agent searching for hotel accommodation might interact with a hotel's AI Agent to find a preferred room, negotiate a price, and book, all with minimal or no human oversight or input. This simple example may seem innocuous, but it is not difficult to hypothesise troubling examples: two AI Agents from separate competitors in the same market interact and, unbeknownst to the entities who deployed them, collude on price.[47]

There are also growing concerns regarding AI adaptiveness. There are various examples of AI systems "hacking" games to achieve a reward or other desired outcome by flouting the rules of a game.[48] As noted above, recent studies show that advanced AI systems are developing the ability to "scheme", such as strategically introducing mistakes into responses and disabling oversight mechanisms.[49] As a result of these concerns, an active area of technical AI research is how to solve the "alignment problem": ensuring that AI systems develop with society's values (and in compliance with law) to avoid undesirable, even intentionally harmful, adaptations.[50]

Of course, natural persons, and legal persons such as corporations, are also highly autonomous and adaptive. They are accountable for their decisions and their conduct, pursuant to applicable laws. If a natural person commits fraud, it may be possible to prosecute them. If a company negligently causes harm, it can be sued. If a public authority makes a decision without the necessary power to do so, it may be possible to have its decision reviewed and potentially overturned.

However, autonomous and adaptive AI systems do not currently have separate legal personality and therefore can neither be sued or prosecuted. Therefore, it is necessary to find a person (or persons), whether natural or artificial (such as a company) who can be held accountable for the AI system. In this context, a central concern regarding increasingly autonomous and adaptive AI systems is that they could lead to "liability gaps", where no natural or legal person is liable for the harms caused by, or the other conduct of, an AI system.[51] While the autonomy and adaptiveness of AI raises the possibility of liability gaps, it does not guarantee they will crystallise. With the many varied potential uses for AI in future, it is difficult to assess where such liability gaps may in fact arise, though they are more likely to occur in connection with highly autonomous and adaptive systems, given that other systems are likely to be more predictable and easier to control.

One important issue may be identifying the entity at risk of liability, as well as which entity ought to be liable. There are potentially many different persons involved in the development and use of an AI system. There is the developer of the AI model, the entity that provides and prepares the data for developing a model or for fine-tuning it, the software developer who incorporates an AI model into another piece of software or another product, the company that packages and sells the AI system or otherwise makes it available to end-users (without themselves designing or developing the software or other product), or the end user of the AI system, whether a person, a company or a public authority. This highlights the complicated supply chains of AI systems, which we return to below. In any given case, some, or all, of those entities might be liable for an AI system, for example by way of owing a duty of care to a third party in relation to that system. Resolving that question for each entity in the AI supply chain may not always be straightforward.

Even once a person is identified, the next question is whether existing laws can satisfactorily determine liability for AI systems. In this section, we discuss how some such issues might arise, with particular reference to the laws pertaining to causation in both private and criminal law, and in relation to knowledge and recklessness in certain criminal offences.

Establishing causation

In the context of private law, some claims have a required element of causation. Typically, it must be proven on the balance of probabilities that a harm was caused by some relevant conduct (or an omission). For example, to establish a claim in the tort of negligence, it is necessary to prove (among other things) that the defendant's breach of duty caused the claimant's harm.

In the context of criminal law, for certain offences an element of the offence is that a specific consequence occurs, and for some of those offences it is also required that the defendant caused the specific consequence. For example, to prosecute a person for causing death by careless, or inconsiderate, driving, it is necessary to prove beyond a reasonable doubt that the defendant's careless, or inconsiderate, driving caused the victim's death.[52]

Causation typically requires establishing, first, that "but for" the relevant conduct (or omission) the resultant harm or other consequence would not have occurred ("factual causation"). One method for challenging causation is to argue that an "intervening cause" broke the "chain of causation" between conduct and event. An intervening cause is when something unconnected with the conduct or omission occurs such that it would be unjustifiable to hold the original person, A, responsible for the ultimate consequences that befell B. However, even if factual causation is established, it is necessary to also show that the harm caused was "reasonably foreseeable" (sometimes described as satisfying "legal causation"). This requires establishing that the resulting harm was a reasonably foreseeable consequence of the defendant's relevant conduct (or omission).

Returning to AI systems, given that they cannot themselves be liable, the focus of the causation enquiry is on a person potentially liable *for* the AI system. The question is whether it is possible to establish factual and legal causation regarding a person in relation to the outputs of an AI system.

The autonomy and adaptiveness of AI potentially raises issues with respect to both factual and legal causation. Those potential issues arise due to the autonomy and adaptiveness of AI, such that they are less predictable and controllable than other algorithmic systems. If it is not known how or why an AI system produced the output it did, it may be difficult to prove that but for the defendant's conduct (for example, in training or developing or deploying the AI system) the harm would not have occurred. It might also be possible for defendants to argue that an AI system's unpredicted output was an intervening cause or otherwise simply was not reasonably foreseeable.

To elucidate how those issues might arise, we will consider one example in relation to the requirement that for a defendant to be liable in negligence, the harm caused by the defendant's breach must be reasonably foreseeable.

Suppose an AI system is incorporated into a domestic robot, which is subsequently sold to consumers. The AI robot has a high level of autonomy and carries out a range of domestic tasks, such as cleaning and minor maintenance, without human direction or oversight. The robot and accompanying AI system have been vigorously tested prior to deployment, following all applicable standards and regulations. After being purchased by a consumer, one of the AI robots deployed in a home injures an infant in the course of cleaning. It is not known why the AI system harmed the infant. For example, it is not known if the AI system misidentified the infant as rubbish that needed to be cleaned or if it determined that the home would be cleaner if the infant was prevented from dirtying it in the first place.

The question is whether the harm caused by the domestic robot was reasonably foreseeable in the position of the potential defendants in relation to the AI robot. Accordingly, the enquiry is in relation to those potential defendants and their involvement with the development and deployment of the AI robot. Assume the AI system in the robot was expressly designed to identify human persons and avoid causing them harm, and had never caused any such harm in an enormous number of tests. The defendants could argue that the harm caused by the AI robot was not reasonably foreseeable. If they were successful, there may be no person liable for the harm caused.

The difficulty for the defendants in this example is that under the law of negligence, even very small risks of harm can be reasonably foreseeable (even if unlikely). The exact causal mechanism of the harm does not need to be reasonably foreseeable. It is sufficient that the harm caused is describable as a foreseeable risk of the activity. In this case, it is foreseeable that an AI robot, that navigates the same physical environment as an infant, might cause harm to the infant (for whatever reason). This is true even if such harm occurring was reasonably considered to be highly unlikely, because the AI robot was robustly trained to avoid causing such harm.

This does not necessarily mean the defendants in this example would be liable; the other elements of the tort of negligence would still have to be established. To the extent that the defendants took all reasonable precautions to avoid or mitigate that risk of harm, they may simply not be in breach of any duty of care owed. If not in breach, the defendants would not be liable (even if the harm caused was reasonably foreseeable).[53] The potential problem that follows is whether all harms caused by unpredictable, autonomous, and adaptive AI systems are reasonably foreseeable, even if any particular instance of harm is highly unlikely and not predicted.[54] In other words, with AI, unexpected outcomes are expected.

This is a simple example, involving a relatively simple form of harm, but it shows the complexity of resolving the potential difficulties AI raises with respect to just one aspect of causation in just one area of law (negligence).

Finally, even if instances of actual liability gaps are identified, it does not follow that they are *inherently* a problem of AI, or unique to AI. It may be that, in certain circumstances at least, no person is liable for the harm caused by an AI system, or no person is criminally responsible for its outputs. That is already the case for natural and legal persons. Not every instance of harm results in civil liability. Only certain conduct results in criminal liability. If liability gaps do arise, the crucial question is whether those gaps are unsatisfactory, having regard to the purposes for imposing liability in private law and in criminal law, such that it is necessary to plug those gaps and, if so, how.

Mental element

Some types of private claims and most criminal offences require a particular state of mind, such that the requisite wrong or crime can only be established if the defendant can be shown not only to have committed the act or omission, but also to have had the requisite state of mind, such as knowledge or intent.[55] For a private law example, to bring a claim in knowing receipt of trust property, it is necessary to establish that a person received trust property *knowing* that such property came by way of a breach of trust or fiduciary duty. For a criminal law example, to be guilty of perjury the prosecution must prove that the defendant knew the statement was false or believed it not to be true.[56]

These requirements potentially also pose issues with AI systems, because of the disconnect between the autonomous and adaptive AI system, and the persons using the systems. The question is whether the relevant person had the requisite knowledge, not whether the AI system had the requisite knowledge. The position is still more complicated where the criminal liability of a company or other corporate body is concerned, as the body will generally only be liable if the actions and state of mind of a relevant natural person can be “attributed” to the body[57] or there is some other provision by which a relevant body can be criminally liable where an associated natural person commits the offence.[58]

For example, it is an offence under section 89 of the Financial Services Act 2012 knowingly or recklessly to make a false statement with the intention of inducing another person to enter into a relevant agreement or being reckless as to whether making the statement may do so (among other things and subject to other requirements). Suppose a company uses an advanced, autonomous, adaptive AI system to manage all communications with investors and potential investors. The AI system made false statements to investors and potential investors, after it learned that it was more likely to convince investors to enter agreements if it made false statements as to the company’s investment products. Given the high level of automation of the AI system, it wrote and sent the communications without any human input; no humans checked the outputs before they were sent, and the company employees did not therefore know that the false statements were being made.

As discussed, the AI system cannot be criminally liable as it does not have legal personality. Criminal liability would therefore have to apply either to the company using the AI system or to a natural person (such as an employee or officer of the company). Assuming we are considering the company’s liability (and assuming the company “made” the statements issued by the AI system), for the company to be liable it would have to be established that: (i) the company “knew” the false statement was made; or (ii) was reckless in relation to the truth or falsity of the statements being made, by being aware of the risk, and unreasonably taking that risk.[59]

With respect to the former, it may be difficult to show the company actually *knew* a false statement was made in circumstances where it did not check the statements prior to being sent. With respect to the latter, it would have to be established that the company was aware there was a risk that the autonomous and adaptive AI system would send statements that were false, and took that risk unreasonably.

This may be difficult to establish. Although the company was aware the AI system was making statements, it was not aware it was making any particular statement.

Further, what is reasonable in all the circumstances would turn on factors relating to the deployment of the AI system.[60] Recklessness may be difficult to establish, where, because of the autonomous and adaptive nature of the system, the particular risk might not have been foreseen, or may have been considered highly unlikely, such that taking that risk did not appear unreasonable in the circumstances, particularly taking into account how the AI system was developed, trained and deployed. If knowledge or recklessness were not established there would be no criminal liability, despite the AI system having made false statements to investors, inducing them to enter into relevant agreements.

As above with respect to the private law example regarding causation, this relatively simple example shows the potential complexities that AI raises in relation to mental elements, in this example with respect to criminal law.[61] As above, if any liability gaps exist in this context, the question is whether that is acceptable or if law reform is required, and, if so, what that reform should be.

Who might be liable?

Above we have discussed the possibility of liability gaps without discussing *who* might be liable. The process for developing and deploying AI systems can involve many different (potentially repeating) steps, and many different entities. AI supply chains can therefore be intricate and complicated.[62] For example, the development and deployment of AI can include the following:[63]

(1) Collection and preparation of data to train an AI model.

(2) Development of an AI model, including training (or “pre-training”), and making the trained model available to other entities.

(3) Designing a software package or a product that will use an AI model as a component. For example, a customer service application might use an AI model to help produce responses to customer queries, but it also needs other software to do so, such as the operating system in which it operates.

(4) Fine-tuning an AI model (such as a Foundation Model) for use for a specific task in a specific domain. This may require obtaining and preparing further data.

(5) Once a fine-tuned AI system is prepared, testing the system for errors or unwanted features.

(6) Making the AI system available for use by end-users (for example, the use of LLMs by the public) or the direct deployment of an AI system for a specific use (for example, the use of an AI diagnostic system by a healthcare company).

(7) Once in use, monitoring of the AI system to optimise its operation and to detect any errors or unwanted features in the system after it is deployed.

Some of these steps may be repeated. Once an AI system has been deployed, it might be re-trained using new data to improve performance or to undertake additional tasks. Each step may also involve a separate entity. The body obtaining and preparing training data can be different to the Foundation Model developer. The entity fine-tuning the model can be different from the body designing the software package of which the AI model is a component. But each step need not be undertaken by separate entities. For example, a Foundation Model developer can be involved in obtaining and preparing data, fine-tuning AI models, deploying AI systems, and so forth. Finally, the entities in these supply chains can be based in many different jurisdictions.

As an example, consider the development of an AI medical diagnostics system. A healthcare provider contracts with another company to have them develop and deliver the system. That company obtains access to a Foundation Model from a Foundation Model developer, which they then fine-tune for a specific diagnostic function. Additional data is compiled and prepared for fine-tuning, completed by the healthcare provider with the assistance of a specialist data collection and preparation services provider. Alternatively, that developer contracts with another entity to build and train a new model for the AI diagnostics system from scratch. Once fine-tuning is complete or the new model is trained, the same contractor may contract with a software developer to develop the surrounding software that enables the AI system to work. Once the system is fully developed, it is tested before deployment to identify any errors. Once testing is complete, the new medical diagnostics AI system is made available to the healthcare provider, which it then uses to deliver services to patients via healthcare professionals employed by the provider.

The challenge raised by these complicated supply chains is that it may be difficult to determine who should putatively be responsible for the outputs of the AI system.[64] For example, in the context of a claim for negligence, for a party to be liable, they must owe the victim a duty of care. In the context of these complicated AI supply chains, it may be unclear which parties owe the victim a duty of care in relation to particular harms caused by an AI system.

In the above example, the healthcare provider will be likely to owe a duty of care to patients, being the body that has engaged others to have the AI diagnostics system developed and is using it, via healthcare professionals, to deliver healthcare services to patients. It might however be difficult to show that the healthcare provider had been negligent if they have used the AI system as it was intended to be used and otherwise acted reasonably in the circumstances. The harm may have been caused by conduct of another party in the supply chain, but it is less clear the extent of any duty of care that may be owed by other entities in the supply chain to the patient at the end of the chain. What about the company that obtained and prepared the data for fine-tuning the model? Or the software provider that developed the software surrounding the AI model? Does the Foundation Model developer whose pre-trained model is the foundation of the AI diagnostics system owe the victim a duty of care? This is a challenge given that "upstream" decisions regarding model design can have "downstream" consequences for people interacting with the system.[65]

The issues caused by complicated supply chains are not strictly unique to AI. In private law, for example, in the law of negligence the classic case of *Donoghue v Stevenson*[66] involved harm caused to the purchaser of a bottle of ginger beer that contained a snail, for which the producer of the beer, not the retailer, was ultimately liable in negligence. This led to the establishment of the "neighbour principle".

The product liability regime implemented in European countries (including the UK) from the 1980s, some 50 years after that seminal case, partially reflects the perceived need to make producers account for harms caused by defective products in circumstances where producers were far removed from users of those products. Where it is not possible or practical to sue the producer of a defective product, the injured party may have a claim against the importer of the product or, in some cases, the supplier. One of the purposes behind that regime is to allocate responsibility to those in product supply chains with the greatest ability to reduce harm and insure against such harm occurring. Accordingly, the law in these areas has developed to account for complicated supply chains.

While these existing laws might be able to identify the persons potentially liable for AI,[67] the issue remains that the duties owed by each entity in the supply chain are presently unclear.

Further, factors relevant to determining those duties, such as the proximity of potential defendants to a victim and their ability to control or prevent the resultant harm occurring, may be difficult to apply in relation to AI systems. By reference to the healthcare provider example, while the healthcare provider has immediate control over the AI system, in that it is using the tool and could cease using it to prevent harm,[68] it may have had little control over the design and development of the system. By contrast, the entity that collected and prepared the data for fine-tuning the Foundation Model is far removed from the patient, and the use of the AI diagnostics system, but it may have had significant influence over the system.

As above regarding causation, the fact that it may be difficult to identify the appropriate parties to be liable for the outputs of AI systems does not mean that existing laws are necessarily unable to deal with these issues. The relevant outcomes depend upon the facts and the specific context, including the type of harm that occurs.

There are also potential practical (as opposed to strictly legal) issues regarding the allocation of liability for the outputs of AI systems. For example, in the context of private law and the complicated AI supply chains above, a victim may need to prove why an AI system produced the output it did, to determine the cause of harm. As we discuss in the next section, the opacity of AI means that it can be exceedingly difficult to determine why an AI system produced the output it did, if possible at all. Given the highly technical and complicated nature of AI systems, determining why an AI system acted as it did will be most likely to require expensive expert evidence. Further, even if that expert assistance is available to, and affordable for, a victim, there may also be barriers to obtaining the evidence necessary for the expert to provide a reasoned analysis of the system. Such evidence may be difficult to obtain because the relevant data could be protected as trade secrets, particularly the underlying technical detail of an AI model.

Opacity

By "opacity" we refer to the fact that it can be difficult to explain how or why AI systems make the outputs they do, if not impossible in some circumstances. AI systems may be opaque because of a lack of transparency about their design, development and training, but, even where that information is available, they may be opaque because of their nature as complicated mathematical functions.

As to the former, there is a general lack of transparency regarding the technical details of AI systems. Foundation Models are developed at great cost, and are highly valuable, such that the full technical details of such models are carefully protected.[69] This can also be true of other, simpler, algorithms, that are used for specific purposes.

For example, in the Wisconsin case *State v Loomis*,[70] an algorithmic risk assessment tool called COMPAS provided an assessment of the offender's likelihood of reoffending, which was relied on, in part, to sentence the offender. The offender sought post-conviction relief on the basis that the use of the COMPAS risk assessment violated his due process rights, because he was entitled to an individualised sentence and a sentence based on accurate information. Part of the difficulty was that the COMPAS methodology to produce risk assessment reports was a trade secret and therefore not made available. The Wisconsin Supreme Court ultimately dismissed the offender's application, finding that his due process rights had not been violated, in part because the COMPAS risk assessment was based on publicly available data and the offender's own responses, and that the COMPAS risk assessment had not been the sole factor in determining the sentence. However, Justice Abrahamson noted that, despite the developers of COMPAS being routinely asked how it worked, "[f]ew answers were available".[71]

Second, even where the technical details of such systems are transparent, AI systems are difficult to understand and explain even for experts in AI, including the very designers and developers of AI systems. Modern AI systems are complicated mathematical functions of enormous scale. Their algorithmic nature, coupled with their sheer computational size, makes it extremely difficult to know why an AI system has generated a particular output or how it will adapt. Even when experts look "under the hood" they may not be able to identify the reasons and causes behind the system's outputs, in the way we explain people's decisions by way of their knowledge, beliefs, intentions, and reasons.[72] As leading AI developer Anthropic has stated "a surprising fact about modern large language models is that nobody really knows how they work internally".[73]

Current AI models do not have a model of the world by which they "understand" it. This may be partly why LLMs are known to hallucinate and provide false answers, even obviously false answers, to simple questions. Accordingly, even where it is possible to "ask" leading AI models how they arrived at a certain output, such answers are not necessarily reliable. The model's answers remain a product of these advanced statistical and mathematical methods, not an "understanding" of the task they have been given.

As briefly raised in the prior section, the opacity of AI raises (or exacerbates) legal issues in relation to AI. The prior section discussed private law and criminal law, and the opacity of AI raises issues there as well (in relation to determining causation for example), but opacity may be a particular problem in relation to public law.

Opacity and AI-influenced decision-making

Public authorities have lawful authority to make certain decisions and take certain action in relation thereto, including obligations to act in certain circumstances. Their authority is constrained by applicable legislation, but also by certain common law rules.[74] In relation to the latter, assuming the decision-maker has the power to make the decision in question, those rules typically impose obligations on the way decisions are made. For example, decisions must be made following the right procedure (such as hearing from the right people); not be biased; be made taking into account all relevant considerations and no irrelevant considerations; and final decisions must be reasonable or proportionate.

The opacity of AI poses potential problems for public law in relation to ensuring accountability for public decision-making by way of these rules.[75] For example, in relation to the requirement to take into account all and only the right considerations, with natural and legal persons, this involves identifying the sets of relevant and irrelevant information for making a decision and then determining what information the decision-maker did (or did not) take into account. In determining a person's application for a spousal visa, a relevant factor would be the relationship status of the applicant with their partner; an irrelevant factor would be the applicant's hair colour. The problem with decisions made by AI systems is that it may be difficult to determine if the system has taken into account the relevant factor (relationship status) and not taken into account the irrelevant factor (hair colour). Where a natural person has made a decision, they can be asked what factors they took into account when making the decision. Along with other evidence regarding how the decision was made, that can be used to determine whether the decision was made lawfully, taking into account all and only relevant considerations.

The same is not necessarily true where an autonomous and adaptive AI system makes a decision. If it is not known how an AI system produced its output, it is difficult to determine whether those requirements have been met. It may be particularly difficult to determine that an AI system has not taken an irrelevant factor into account. This could be because the technical data relating to the model is unavailable (due to, for example, trade secrets protecting a commercial provider of an AI system) or it simply may not be possible to discern from the AI system's inner workings how it made its decision or produced its output.

This is not only a problem where decisions are made by an AI system with minimal or no human oversight. It is also potentially a problem where a human decision-maker makes a decision partly informed by an output of an AI system. The risk is that an error in the AI system's original output could infect the natural person's decision with error, such that it was unlawfully made. In that case, the same problem regarding opacity arises. This risk may also be exacerbated by the issue of automation complacency and bias. Automation complacency refers to the problem that as automation systems (such as AI) improve and less is required of a natural person operator, the operator begins to assume the system is infallible and less actively monitors the system.[76] Automation bias is where the operator of the automated system trusts it so much that they over-rely on it compared with other sources of information.[77]

A similar issue arises in cases where there is a requirement for a public authority to provide reasons for a decision made.[78] Where reasons are required to be given, they must be adequate, where adequacy is determined on a case-by-case basis. In general, however, reasons should be sufficient to understand why a decision was made, what factors were taken into account, how any issues of law or fact were resolved, and enable a person affected to assess the validity of a decision and whether it is open to challenge. The opacity of AI again raises potential issues with this requirement. If an AI system has autonomously made a decision without human oversight, it may not be possible to provide adequate reasons. Similarly, where a decision is made by a person, but reliant on an output of an AI system, it may not be possible to provide adequate reasons to explain how the AI system produced the output that was relevant to the decision made.

The above issues arise in relation to decisions that have been made, and reasons given, where a public authority has the power to make a decision. However, the use of opaque AI systems might also raise issues with respect to determining whether a public authority even has the power or jurisdiction to make a decision in the first place (or has an *obligation* to act). For example, under section 20 of the Children Act 1989, a local authority shall provide accommodation for any child in need within their area who requires accommodation (among other things). Determining the age of a child is required to determine the local authority's obligations. In some circumstances it may be difficult to determine the age of some children. A human agent of a local authority may exercise their judgement to determine the age of a child and will be afforded a certain degree of deference in doing so. However, what if an opaque AI system were to be used to determine the age of a child for the purpose of section 20 of the Children Act 1989? What deference would be given to such a system when making that determination? Given the opacity of the AI system, it might be difficult to understand how the AI system arrived at its determination. Therefore, what is the right approach for determining whether the public authority had the jurisdiction to take a certain action? Or, as in the above example, had an obligation to take certain action?

Though we have discussed public law decision-making generally here, this issue also arises in the criminal justice system. Algorithm-based risk assessment tools and predictive policing methods are already used in the United Kingdom.[79] Their use is likely to expand, including, for example, in the preparation and analysis of evidence.[80] In future, they may be used more extensively, including in preparing or analysing evidence for prosecutions. Accordingly, the opacity of autonomous and advanced AI systems, due both to a lack of transparency (such as in the Wisconsin decision of *Loomis*) and the "black box" nature of such systems, raises concerns regarding a defendant's right to a fair trial, and to procedural fairness in relation to decisions on sentencing and parole.[81]

Finally, above we referred to a concern in the context of private law that victims might face obstacles in bringing claims because they lack the expertise, or resources to obtain expertise, to analyse and understand AI system outputs. This issue also arises in relation to opacity and decision-making by public authorities and within the criminal justice system. It may not be possible for a person to challenge a decision made by an AI system, or with the assistance of an AI system, without the assistance of expert evidence and without access to potentially commercially protected information.[82] Accordingly, the potential issues regarding the opacity of AI are not just doctrinal; they also may raise practical issues regarding access to justice.

Oversight and reliance on AI

As noted above, because AI has no legal personality, the legal issues regarding AI at present relate to natural and legal persons interacting with, and relying on, AI. How they do so is the counterpoint to the discussion above regarding the difficulty in allocating liability for AI systems, given the autonomy and adaptiveness of those systems. Even where the appropriate party to be responsible for the outputs of an AI system is identified, it is still necessary to determine the scope and content of their duties in relation to those systems (as the EU's AI Act seeks to do by reference to different kinds of AI systems).[83]

As noted above, AI systems can produce false or otherwise undesirable outputs. A common example being AI "hallucinations". Some hallucinations are benign. Some are not. There are now examples from many jurisdictions of lawyers citing case authorities in written submissions that do not exist, let alone stand as authority for the point for which the cases were relied on.[84] The problem has arisen where lawyers have used an AI system to assist in the preparation of their written case. They may have asked an LLM to "research" relevant legal authorities or to draft a court document, such as a written argument, for the proceedings. The AI system produces the research or court document and includes authorities that look real and convincing. They have plausible-sounding case names, case citations, and they may even have fictional quotes from the fictitious case with purported pinpoint references. Those lawyers who have submitted such authorities in real-life examples have been too reliant on an AI system, insufficiently sceptical of its outputs. It is well-known that AI systems hallucinate, so any case authorities referred to by an LLM must not be relied upon as accurate. All such references must be checked and, if necessary, corrected.

In some cases, these questions of reasonable reliance on AI may be straightforward. Lawyers have a duty not to mislead the court and must therefore ensure that any authorities put forward are real and stand for the proposition being supported (and are not otherwise misleading). It is not difficult for lawyers to check the accuracy of authorities. It should go without saying that lawyers must not rely on non-existent legal authorities, nor make false statements as to the content of those cases (or at all), and that the lawyer is responsible for the content of their own statements no matter how they are produced. Producing misleading documents due to over-reliance on AI will almost certainly be a breach of their regulatory obligations, may give rise to liability for professional negligence, and may even risk the lawyer being liable for contempt of court.[85]

In some cases, determining the content and scope of a person's obligations when using and relying on AI may be more difficult. For example, where a medical professional uses an AI system to help analyse medical images, at what point does the human professional decide not to follow the AI system's suggestion or recommendation? If an AI system diagnoses a person with a tumour and recommends chemotherapy, when would it be reasonable for the professional to decide otherwise? This might be especially difficult in those circumstances where there is evidence that AI systems are superior to natural persons in respect of the relevant analysis (as there is for some diagnostic tests).[86] If the medical professional decides not to follow the system, and harm results, could they be in breach of duty for failing to follow the AI system's recommendation?

The issue of reliance may also be heightened where the persons using AI systems do not know the technical details underlying them, including the data on which they were trained. There might be limitations with the use of an AI system of which they are not aware (and reasonably so), but which should change how they use and rely on the AI system. Perhaps the distributor or developer of that AI system has, or should have, a duty to ensure that anyone using the system is adequately informed of its uses and limitations to ensure it can be used safely and without causing harm. However, those developers may be also unaware of issues within the data on which the AI system was trained, despite having fully tested the AI system prior to deployment, such that, for example, undetected bias in the data leads to harm that was not foreseen. In such a scenario, it might seem that all natural and legal persons interacting with the AI system acted reasonably, despite the fact that by following the AI system and allowing it to function unimpeded, the system caused harm or otherwise produced a negative output.

This issue may be one that is particularly acute in public law. For example, public authorities cannot fetter their own authority. Decision-makers are accorded a degree of flexibility in the exercise of their authority. They can, and should, take individual circumstances into account when exercising their discretion, rather than inflexibly applying a procedure or policy.

A concern with AI systems is that human decision-makers will habitually rely on the outputs of those systems (akin to the lawyer example above), and, in effect, fetter their discretion. The difficult question is where to draw the line in terms of reliance.

While this discussion has focused largely on an individual's reliance on an AI system, the point also applies in relation to *oversight* of AI systems. Depending on the AI system, natural and legal persons may have more or less oversight of its outputs. Referring again to the medical diagnosis example, at present, a medical professional, trained in the relevant area may be responsible for checking an AI system's diagnostic outputs prior to any diagnosis being provided to a patient. The AI system reviews the scans, produces a diagnosis and potentially a treatment plan, which is reviewed by a medical professional (and amended as considered necessary), before being provided to the patient. The system's outputs are closely controlled. Its final output is reviewed by an expert in the field, prior to it being provided to the patient, with any necessary adjustments. At the other end of the spectrum is an example where there is effectively no oversight of the AI system's outputs. An AI system that is used as a customer service chat bot is likely to have minimal human oversight or intervention, except insofar as customer enquiries or complaints are elevated to a natural person.

Some AI systems require more oversight and monitoring than others. For example, the use of an AI system to process welfare applications and oversee compliance with welfare conditions. Such a system may provide benefits to both the public authority and applicants by way of increased productivity. More decisions may be processed in less time, and potentially at lower cost. However, without oversight, biases or other errors in the system may present a risk of serious harm being caused by way of applications being unlawfully denied or claims of welfare fraud being incorrectly made.[87] The question is what level of human oversight is required to balance these competing considerations.

These issues are also pertinent in the use of AI in the justice system, both civil and criminal. AI systems could lead to increased productivity for courts in criminal trials and progressing civil disputes, including in assisting judges to review and analyse evidence, and to improve the speed of delivery of judgments and reasons, where applicable. All of which could lead to a more efficient justice system delivering fairer, more timely outcomes. But the risks regarding AI systems, from poor oversight and control and overreliance on them, need to be managed to obtain the greatest benefits with the least harm. Determining the duties for use of AI systems in the justice system in the future will be one of the many important legal issues regarding AI.

Given the issues raised above regarding the duties of persons relying on and overseeing AI, a further practical concern is how to implement those duties legally. Should they be determined iteratively by the common law? Should they be determined by way of guidance from regulators or industry bodies in particular sectors (sometimes referred to as "soft law")?[88] Or is comprehensive AI legislation required to provide overarching clarity and certainty?

Training and data

As noted, a unique feature of modern, machine learning AI is that it is *trained*. It is not merely given the rules or instructed how to complete its tasks; it *learns* how to complete its tasks, from training. The legal issues discussed above concerned difficulties arising from the deployment and operation of AI systems, due to their autonomy and adaptiveness and exacerbated by their opacity, and related to how natural and legal persons interact with AI. However, some legal issues regarding AI arise from the role of data in the training and operation of AI systems.

As noted in the explanation of AI systems above, leading Foundation Models are trained on enormous datasets. Given the size of those datasets, it is unsurprising they include many copyrighted works and people's personal data. Accordingly, there are well known concerns regarding modern AI systems and copyright and data protection.

The issues with copyright and AI systems are perhaps the best known and discussed legal issues regarding AI. Given the scale of AI training and AI use, the threat of copyright infringement (in both training and application) is of particular concern to creative groups and creative workers, as evidenced by the recent campaign in the UK on copyright and AI. A Government consultation paper published in December 2024 on copyright and AI has been controversial because it proposes granting a broad data mining exception to copyright (allowing data mining on copyright protected works without rights holders' permission), unless a copyright holder were expressly to reserve their rights.[89] More recently, the Parliamentary passage of the Data (Use and Access) Act 2025 was delayed by proposed amendments in relation to AI and copyright.[90] The issue is also in the spotlight internationally with recent high-profile decisions in the United States, regarding the use of copyrighted materials in training by leading AI developers Anthropic and Meta.[91]

Another significant legal issue is in relation to data protection, in particular the operation of the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018.

The UK GDPR imposes obligations on data controllers and processors, when processing personal data, including:

(1) Data controllers must be transparent about how personal data will be processed.

(2) Personal data must be processed lawfully, fairly and in a transparent manner, and can only be processed for a limited number of grounds.

(3) In terms of processing personal data fairly, such data must be processed as people would reasonably expect and not in a way that would have unjustified adverse effects on them.

AI poses obstacles to compliance with these obligations. Reliance on the ground of consent may be tenuous for AI processing personal data because of the opacity of AI. If it is difficult or impossible to explain to individuals how their personal data will be used (or even whether it will be used), it may not be possible to obtain their "informed" consent for any such processing. There are also challenges with the "legitimate interest" ground, which requires proof of the necessity and proportionality of the processing.[92] Due, again, to the opacity of AI, it may be difficult to establish that any such processing was necessary and proportionate, particularly if data subjects do not reasonably expect their data to be processed by the AI system for the purpose in question. These issues also may make it difficult to "fairly" process personal data, because it may be hard to explain to data subjects how their data has been processed in a manner they can understand. While a data subject has the right to object to a decision being made solely on automated processing pursuant to Article 22 of the UK GDPR (as recently amended by section 80 of the Data (Use and Access) Act 2025), that provision is subject to certain exceptions, including that the automated processing is required or authorised by domestic law.

Another central issue with training and data of AI systems is bias. Bias in the data underlying AI systems can be re-produced in the outputs of those systems. While there are some mechanisms for seeking to manage bias, the complicated and opaque nature of advanced AI systems can make doing so challenging. Bias in these systems creates the risk that they will produce discriminatory outcomes.

A widely publicised example of such discrimination occurred in the United States with the use of an algorithm to predict healthcare risks.[93] Specifically, the algorithm assisted with identifying patients that might need additional care management, for example due to chronic illness. The purpose of the assessment was ultimately pre-emptively to identify those who needed additional care and provide it to them to minimise the risk of further healthcare complications and thereby also reduce costs. The algorithm used prior healthcare spending as a proxy for medical needs, which was a common proxy for healthcare needs. However, using that measure as a proxy led to a racial bias, because Black and White patients who spent the same amount on healthcare did not necessarily have the same underlying care needs. Ultimately, this meant that proportionately fewer very sick Black persons were assessed as high need for medical care, revealing an indirect and implicit discrimination in the model.

This algorithm was not an autonomous and adaptive AI system as we have described above. In this case, the measures the algorithm used were known, so it was possible to identify how the bias in the algorithm led to discrimination, and to therefore correct it. However, the same risk of bias and discrimination arises with respect to the use of autonomous and adaptive AI systems, only with more limited ability to determine how and why such systems have produced their outputs, due to the opacity of AI.

By way of example in the context of public law, it may be difficult for a public authority to determine what, if any, bias exists in training data. The public authority may not be able to access the AI system's training data, due to commercial and contractual confidentiality[94] and, even if it can access that data, the opacity of the models may mean it is not apparent whether bias exists. Public authorities therefore face the difficult challenge of determining how to select, use and monitor AI systems in compliance with the Public Sector Equality Duty (for example).[95] Further, given the potential scale of AI use by public authorities in future, biased AI systems in public decision-making have the potential to impact a significantly larger set of people than biased human decision-makers.

This issue is not limited to a particular area of public or private law. It applies wherever AI systems are used to generate outputs that affect people, such as in job recruitment decisions, including where humans are "in the loop". If a human person relies upon information or an output from a biased AI system, their decision can still unlawfully discriminate (even if unbeknownst to the person using the AI system).

Separate legal personality for AI?

Above, we have discussed a range of potential legal issues raised by AI. As noted in the introduction, the purpose of this paper is to raise awareness and discussion of issues raised by AI and the law, so we have not set out potential options for reform for the many issues raised above.[96] However, given that a key difficulty raised above was the challenge in identifying a natural or legal person to be responsible for AI systems, in circumstances where AI systems do not have their own legal personality, we consider it useful to discuss the, perhaps radical, option of granting AI systems some form of legal personality. While this may seem futuristic, it has already been considered in academic discourse,[97] and as AI systems advance, it may become an increasingly salient option. The consequences of doing so may not be entirely clear but it is one potential option to be considered.

Legal personality is a creation of legal systems and has been granted to a range of entities. Therefore, what entities have legal personality can change over time. It is often theorised as a bundle of rights and obligations,[98] commonly including: the ability to own property, to acquire rights and owe obligations in relation to others' rights, to enter contracts, and to sue and be sued in the legal person's own name. However, it is possible for different categories of legal persons to have different bundles of rights and obligations. Corporations do not have the same rights and obligations as natural persons.

In fact, a form of legal personality has also been granted to temples in India[99] and a river in New Zealand.[100] In October 2017, Saudi Arabia granted "citizenship" to a robot called Sophia[101] and in the same year Tokyo's Shibuya district granted an AI system "residency".[102] Some commentators have suggested those technology-related announcements were for publicity, but in February 2017, the European Parliament stated that the autonomy of "robots" raised the question of whether a new category of legal personality needed to be created for them,[103] though no further steps have been taken in that regard. A group of experts in various fields, including AI, have responded to the proposal with an open letter stating that granting robots legal personality is ethically and legally inappropriate.[104]

Determining whether certain AI systems should be granted legal personality is a complex issue, and what follows is only the briefest summary of some of the relevant considerations in response to three broad questions:

(1) What are the reasons for and against granting AI a form of legal personality?

(2) If some AI systems were to be granted a form of legal personality, what features would AI systems need to possess to warrant being granted legal personality?

(3) What type of legal personality should be given to relevant AI systems?

There are many reasons for and against granting AI legal personality.[105]

Reasons in favour include:

(1) Filling the gaps regarding liability and responsibility discussed above.

(2) Potentially encouraging AI innovation and research (by granting AI developers separation in terms of liability).

(3) Even encouraging AI systems themselves to develop safely (as if the systems can themselves be liable they can themselves be incentivised to avoid liability).

Reasons against include that:

(1) It may lead to AI systems being used as "liability shields" protecting developers from reasonable accountability.

(2) The complexity of granting AI the ability to hold funds and assets such that they can be held meaningfully accountable, for example by way of claims being brought against them.

If the decision were made to grant (some) AI systems legal personality, the next question is to which AI systems, or types of system, this should apply. Legal personality does not seem appropriate for all AI systems. It may seem intuitive that legal personality should not be granted to an AI system which is used as a tool for a single task, such as filtering spam e-mails. But what about AI Agents, discussed above? Theorists have posited various features as a threshold for granting AI legal personality, including their degree of (i) autonomy; (ii) awareness; and (iii) intentionality.[106] Whatever criteria are used to determine to what systems to grant legal personality, the difficult question is where to draw that line, and how that point should be defined.

Assuming that issue is resolved, the next question is what type of legal personality should be granted to AI legal systems—what bundle of rights and obligations should such a system be granted? Should AI systems be granted a form of legal personality where they are required to be owned by natural or legal persons, similar to corporations having shareholders? If so, should there be a form of limited liability for the owners of those systems? Limited liability has been described as a "privilege",[107] to be exercised subject to creditor safeguards. In English law, to obtain limited liability status, a company must be registered with the state. It must disclose the names of its directors and "people with significant control",[108] and must file annual accounts. Something similar would likely be required should AI systems be granted separate legal personality with underlying owners or people with control. If they were entirely separate legal persons they would still require means for identification, just as there are forms of identification for natural persons, such as names, birth dates and government identification numbers (for example, national insurance numbers). Further, some mechanism would need to be put in place such that an AI system could be subject to sanction were it to commit a criminal offence.

Finally, even if some AI systems were to be granted a form of legal personality, that would not automatically resolve the legal issues discussed above. The next consideration would be determining what rights and obligations the AI systems should have. For example, if it had duties to take reasonable care, by what standard should this be assessed?[109] In the context of professional negligence, for example, would it be compared to the behaviour of a reasonable professional in the same circumstances? Is that the correct comparison for an AI system that may have superior skills to a human in some respects, but inferior skills in others? Even if AI systems were afforded legal personality, legal questions would remain.

Even from this short discussion, we can see that determining how to implement legal personality for AI systems is a large and complicated topic. It is also not clear presently that any AI systems are sufficiently advanced to warrant being granted legal personality. However, AI technology continues to progress. Depending on the pace of that development, the option of granting some AI systems legal personality is likely increasingly to be considered. It may be a streamlined solution to the immediate problem of liability, yet the numerous issues that would arise were this to be adopted must also be addressed.

[1] Predicting protein structures is important towards drug design and enzyme design, among other uses. Google DeepMind AlphaFold team, *AlphaFold 3 predicts the structure and interactions of all of life's molecules* (8 May 2024), https://blog.google/technology/ai/google-deepmind-isomorphic-alphafold-3-ai-model/.

[2] Go is a strategy board game that originated in China more than 2,500 years ago. It is usually played on a grid of 19x19 squares, although beginners use a smaller grid of 9x9 or 13x13. It has simple rules but is extremely complex and has a staggeringly high number of potential legal board positions.

[3] *Most visited websites in the world*, https://www.semrush.com/website/top/ (last updated June 2025).

[4] AI Opportunities Action Plan (2025) Cm 1242. In addition, on 6 February 2024, the previous Government published its consultation response with respect to the regulation of AI: A Pro-Innovation Approach to AI Regulation (2024) Cm 815.

[5] "Memorandum of Understanding between UK and Anthropic on AI opportunities" (13 February 2025),

https://www.gov.uk/government/publications/memorandum-of-understanding-between-the-uk-and-anthropic-on-ai-opportunities/memorandum-of-understanding-between-uk-and-anthropic-on-ai-opportunities; "New Google partnership will help rid taxpayer of 'ball and chain' legacy tech and aim to upskill 100,000 civil servants in tech and AI" (9 July 2025), https://www.gov.uk/government/news/new-google-partnership-will-help-rid-taxpayer-of-ball-and-chain-legacy-tech-and-aim-to-upskill-100000-civil-servants-in-tech-and-ai; "Memorandum of Understanding between UK and OpenAI on AI opportunities" (21 July 2025), https://www.gov.uk/government/publications/memorandum-of-understanding-between-the-uk-and-openai-on-ai-opportunities/memorandum-of-understanding-between-uk-and-openai-on-ai-opportunities.

[6] *AI Incident Database*, https://incidentdatabase.ai/apps/incidents/; and AI Action Summit, *International AI Safety Report: The International Scientific Report on the Safety of Advanced AI* (January 2025), https://www.gov.uk/government/publications/international-ai-safety-report-2025.

[7] McKinsey, *The Economic Potential of Generative AI: The Next Productivity Frontier* (14 June 2023), https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction.

[8] Stanford Institute for Human-Centered AI, *The AI Index 2025 Annual Report* (2025), pp 72 to 74, https://hai.stanford.edu/ai-index/2025-ai-index-report; and P Li, J Yang, M A Islam, and S Ren, "Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models" (26 March 2025) *arXiv preprint*, https://arxiv.org/pdf/2304.03271.

[9] Florence G'Sell, *Regulating Under Uncertainty: Governance Options for Generative AI* (September 2024) Chapter 5, https://cyber.fsi.stanford.edu/content/regulating-under-uncertainty-governance-options-generative-ai. With the support of the Government, a working group of 96 international AI experts released a final report on AI safety in February 2025. The report comprehensively sets out AI risks, including risk management. See: AI Action Summit, *International AI Safety Report: The International Scientific Report on the Safety of Advanced AI* (January 2025), https://www.gov.uk/government/publications/international-ai-safety-report-2025.

[10] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, Official Journal L 1689 of 12.7.2024.

[11] For a useful summary of the implementation of the AI Act, see: https://artificialintelligenceact.eu/implementation-timeline/.

[12] European Commission, *General-Purpose AI Code of Practice* (2025), https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai. The Code of Practice is split into three chapters: Transparency, Copyright, and Safety and Security.

[13] For example: R Susskind; *How to Think About AI: A Guide for the Perplexed* (2025); M Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (2019); J Turner, *Robot Rules: Regulating Artificial Intelligence* (2019); R Abbott, *The Reasonable Robot* (2020); and S Russell, *Human Compatible: AI and the Problem of Control* (2020) (among many others).

[14] The Government launched a consultation on AI and copyright in December 2024, which closed in February 2025, and, under the Data (Use and Access) Act 2025, the Secretary of State is required to publish an impact assessment and policy report on its options for managing AI and copyright, within nine months of the act coming into force on 19 June 2025.

[15] Louise Hooper, "Artificial Intelligence and Human Rights" in M Hervey and M Lavy KC (eds) *The Law of Artificial Intelligence* (2nd ed 2024) paras 4-130 to 4-145.

[16] Lord Sales, *The Sir Henry Brooke Lecture: Algorithms, Artificial Intelligence and the Law* (November 2019), https://www.bailii.org/bailii/lecture/06.pdf.

[17] Automated Vehicles (2022) Law Com No 404; Scot Law Com No 258.

[18] Intimate Image Abuse (2022) Law Com No 326.

[19] Stanford Institute for Human-Centered AI, *The AI Index 2025 Annual Report* (2025), Chapter 1.3, https://hai.stanford.edu/ai-index/2025-ai-index-report.

[20] The quote is from: Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (2016) p 1. See also: Organisation for Economic Cooperation and Development, *What is AI? Can*

you make a clear distinction between AI and non-AI systems? (2024), https://oecd.ai/en/wonk/definition; and J Zerilli and A Weller, "The Technology" in M Hervey and M Lavy KC (eds) *The Law of Artificial Intelligence* (2nd ed 2024).

[21] Quoted in M Y Vardi, "Artificial Intelligence: Past and Future" (2012) 55 *Communications of the Association for Computing Machinery* 5.

[22] Organisation for Economic Cooperation and Development, *What is AI? Can you make a clear distinction between AI and non-AI systems?* (2024), https://oecd.ai/en/wonk/definition. The definition for "AI system" in the Data (Use and Access) Act 2025 is very similar to this definition. This definition is also similar to that adopted in the EU's AI Act.

[23] J Zerilli and A Weller, "The Technology" in M Hervey and M Lavy KC (eds) *The Law of Artificial Intelligence* (2nd ed 2024) paras 2-003 to 2-007.

[24] P Villalobos, A Ho, J Sevilla, T Besiroglu, L Heim, and M Hobbhahn, "Will we run out of data? Limits of LLM scaling based on human-generated data" (4 June 2024) *arXiv preprint*, https://arxiv.org/pdf/2211.04325.

[25] M Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (2019) pp 89 to 108.

[26] OpenAI, *Hello GPT-4o* (13 May 2024), https://openai.com/index/hello-gpt-4o/.

[27] Stanford Institute for Human-Centered AI, *The AI Index 2025 Annual Report* (2025) pp 83 to 84, https://hai.stanford.edu/ai-index/2025-ai-index-report.

[28] R Bommasani, D A Hudson, E Adeli, R Altman, S Arora, S v Arx, M S Bernstein, J Bohg, A, Bosselut, E Brunskill, and E Brynjolfsson, "On the Opportunities and Risks of Foundation Models" (16 August 2021) *arXiv preprint*, https://arxiv.org/pdf/2108.07258.

[29] For example, see the summary of AI technical performance in the Stanford Institute for Human-Centered AI's AI Index Annual Report: Stanford Institute for Human-Centered AI, *The AI Index 2025 Annual Report* (2025) pp 81 to 159, https://hai.stanford.edu/ai-index/2025-ai-index-report.

[30] M Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (2019) pp 27 to 36 and G Sanderson, *But what is a neural network?* (5 October 2017, updated 13 March 2025), https://www.3blue1brown.com/lessons/neural-networks.

[31] AI is sometimes broken down into "narrow" and "general" kinds. The former refers to AI that has a specific, and potentially singular, task, such as filtering out spam e-mails. The latter refers to AI that can complete an array of tasks.

[32] Florence G'Sell, *Regulating Under Uncertainty: Governance Options for Generative AI* (September 2024) pp 37 to 39, https://cyber.fsi.stanford.edu/content/regulating-under-uncertainty-governance-options-generative-ai.

[33] M Mitchell, "Why AI is Harder Than We Think" (28 April 2021) *arXiv preprint*, https://arxiv.org/pdf/2104.12871.

[34] Although "chain-of-thought" technology is being developed to seek to improve AI performance on complex problems. "Chain-of-thought" requires AI models to break down their outputs into multiple steps, which has the appearance of a "chain-of-thought". Stanford Institute for Human-Centered AI, *The AI Index 2025 Annual Report* (2025) pp 15, 87, and 111 to 112, https://hai.stanford.edu/ai-index/2025-ai-index-report. See also: J Zerilli, J Danaher, J Maclaurin, C Gavaghan, A Knott, J Liddicoat, and M Noorman, *A Citizen's Guide to Artificial Intelligence* (2021) Chapter 2.

[35] For a recent example, in the courts of England & Wales, see: *Ayinde, R (on the application of) v The London Borough of Haringey and Hamad Al-Haroun v Qatar National Bank QPSC & Anor* [2025] EWHC 1383 (Admin). We discuss AI hallucinations and non-existent legal authorities below.

[36] Anthropic, A Lynch, C Larson, and S Mindermann, "Agentic Misalignment: How LLMs could be insider threats" (20 June 2025), https://www.anthropic.com/research/agentic-misalignment.

[37] J A McDermid, Y Jia, and I Habli, "AI for Lawyers: A Gentle Introduction" in E Lim and P Morgan (eds) *The Cambridge Handbook of Private Law and Artificial Intelligence* (1st ed 2024) p 30.

[38] M Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (2019); and J Zerilli, J Danaher, J Maclaurin, C Gavaghan, A Knott, J Liddicoat, and M Noorman, *A Citizen's Guide to Artificial Intelligence* (2021).

[39] Although this is arguably also true of natural persons. It is not possible to discern why a

person acted as they did from looking at synapses firing in their brain (or some other biological or physical description of their brain). Accordingly, some commentators argue that there is a double standard regarding our demands for transparency of AI, compared with our expectations of transparency regarding the behaviour of natural persons. See: J Zerilli, A Knott, J Maclaurin and C Gavaghan, "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" (2019) 32 *Philosophy & Technology* 661.

[40] There are many, many books and articles written on AI. For useful further general reading, see: M Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (2019); J Zerilli, J Danaher, J Maclaurin, C Gavaghan, A Knott, J Liddicoat, and M Noorman, *A Citizen's Guide to Artificial Intelligence* (2021); R Susskind; *How to Think About AI: A Guide for the Perplexed* (2025); S Russell, *Human Compatible: AI and the Problem of Control* (2020). For technical reading, see: P Norvig and S Russell, *Artificial Intelligence – A Modern Approach* (4th ed 2020); F Chollet, *Deep Learning with Python* (2nd ed 2021); C M Bishop, *Pattern Recognition and Machine Learning* (2006); I Goodfellow, Y Bengio, and A Courville, *Deep Learning* (2016); and D Barber, *Bayesian Reasoning and Machine Learning* (2012).

[41] For a useful sample, see: M Hervey and M Lavy KC (eds) *The Law of Artificial Intelligence* (2nd ed 2024); B McGurk KC and J Tomlinson, *Artificial Intelligence and Public Law* (2025); D J Baker and P H Robinson (eds), *Artificial Intelligence and the Law: Cybercrime and Criminal Liability* (2021); and E Lim and P Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (2024). There is also a vast set of peer reviewed articles on AI and the law, in all areas. Various public authorities have also released AI related guidance, for example: Competition & Markets Authority, *AI Foundation Models: Initial Report* (18 September 2023), https://assets.publishing.service.gov.uk/media/65081d3aa41cc300145612c0/Full_report_.pdf; Bank of England and Financial Conduct Authority, *Artificial intelligence in UK financial services* (November 2024), https://www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financial-services-2024; Information Commissioner's Office, *How do we ensure fairness in AI?* (March 2023), https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-fairness-in-ai/; and Government Digital Service, *AI Playbook for the Government* (10 February 2025), https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government.

[42] R Abbott, *The Reasonable Robot* (2020) p 34.

[43] J Zerilli, J Danaher, J Maclaurin, C Gavaghan, A Knott, J Liddicoat, and M Noorman, *A Citizen's Guide to Artificial Intelligence* (2021) p 72.

[44] Google DeepMind, *AlphaGo*, https://deepmind.google/research/breakthroughs/alphago/. See also: M Mitchell, *Artificial Intelligence: A Guide for Thinking Humans* (2019) pp 199 to 208.

[45] Stanford Institute for Human-Centered AI, *The AI Index 2025 Annual Report* (2025) pp 145 to 148 and 208 to 209, https://hai.stanford.edu/ai-index/2025-ai-index-report.

[46] S Pichai, D Hassabi, and K Kavukcuoglu, "Introducing Gemini 2.0: our new AI model for the agentic era" (11 December 2024) *Google*, https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024/#ceo-message.

[47] B Adkins and L Page, "Competition Law" in M Hervey and M Lavy KC (eds) *The Law of Artificial Intelligence* (2nd ed 2024) paras 12-036 to 12-044.

[48] V Krakovna, J Uesato, V Mikulik, M Rahtz, T Everitt, R Kumar, Z Kenton, J Leike, and S Legg, *Specification gaming: the flip side of AI ingenuity* (21 April 2020), https://deepmind.google/discover/blog/specification-gaming-the-flip-side-of-ai-ingenuity/.

[49] A Meinke, S Bronson, J Scheurer, M Balesni, R Shah, and M Hobbhahn, "Frontier Models are Capable of In-context Scheming" (2024) *arXiv preprint*, https://arxiv.org/abs/2412.04984; and see: P S Park, S Goldstein, A O'Gara, M Chen, and D Hendrycks, "AI deception: A survey of examples, risks, and potential solutions" (10 May 2024) 5 *Patterns*.

[50] Florence G'Sell, *Regulating Under Uncertainty: Governance Options for Generative AI* (September 2024) pp 63 to 64, https://cyber.fsi.stanford.edu/content/regulating-under-uncertainty-governance-options-generative-ai.

[51] P Morgan, "Tort Law and AI: Vicarious Liability" in E Lim and P Morgan (eds) *The Cambridge Handbook of Private Law and Artificial*

Intelligence (1st ed 2024) pp 135 to 171; Z Porter, P Ryan, P Morgan, J Al-Qaddoumi, B Twomey, J McDermid, and I Habli, "Unravelling Responsibility for AI" *arXiv preprint* (2024), https://arxiv.org/abs/2308.02608.

[52] Road Traffic Act 1988, s 2B.

[53] S Steele, "Legal Causation and AI" in E Lim and P Morgan (eds) *The Cambridge Handbook of Private Law and Artificial Intelligence* (1st ed 2024) pp 202 to 205.

[54] S Steele, "Legal Causation and AI" in E Lim and P Morgan (eds) *The Cambridge Handbook of Private Law and Artificial Intelligence* (1st ed 2024) pp 204 to 205.

[55] Some offences may be committed on the basis of recklessness or dishonesty. Recklessness usually requires awareness of a risk coupled with an unreasonable decision to take that risk (*R v G* [2003] UKHL 50; [2004] 1 AC 1034). Whether conduct was dishonest is a matter for the court to assess in the light of the knowledge or beliefs of the defendant (*Ivey v Genting* [2017] UKSC 67; [2018] AC 391; and *R v Barton & Booth* [2020] EWCA Crim 575; [2021] QB 685).

[56] Perjury Act 1911, s 1(1).

[57] At common law, the test is whether the natural person represents "a directing mind and will of the corporation" (*Tesco Supermarkets Ltd v Nattrass* [1971] UKHL 1; [1972] AC 153). Under the Economic Crime and Corporate Transparency Act 2023, ss 196 to 198, liability for certain economic offences can be attributed to a corporate body or partnership where a "senior manager" commits a relevant offence acting within the actual or apparent scope of their authority. Provisions in the Crime and Policing Bill, currently before Parliament, would extend this to all offences. These new provisions draw on the Law Commission's work on corporate criminal liability: Corporate Criminal Liability: an options paper (2022).

[58] For instance, a company can be criminally liable for failure to prevent bribery (Bribery Act 2010, s 7) or fraud (Economic Crime and Corporate Transparency Act 2023, s 199) where that offence is committed by an employee or other relevant person. The former implements recommendations of the Law Commission in Reforming Bribery (2008) Law Com No 313; the latter draws on one of the options we set out in Corporate Criminal Liability: an options paper (2022).

[59] *R v G* [2003] UKHL 50; [2004] 1 AC 1034.

[60] M Dsouza, "Don't Panic: Artificial Intelligence and Criminal Law 101" in D J Baker and P H Robinson (eds) *Artificial Intelligence and the Law: Cybercrime and Criminal Liability* (2021), pp 257 to 259.

[61] For example, one academic considers that criminal law already has the resources to cope with AI technology, at least in the medium term: M Dsouza, "Don't Panic: Artificial Intelligence and Criminal Law 101" in D J Baker and P H Robinson (eds) *Artificial Intelligence and the Law: Cybercrime and Criminal Liability* (2021).

[62] For very useful discussions of AI supply chains, see: I Brown, "Allocating Accountability in AI Supply Chains" *Ada Lovelace Institute* (29 June 2023), https://www.adalovelaceinstitute.org/resource/ai-supply-chains/; S Küspert, N Moës, and C Dunlop, "The Value Chain of General-Purpose AI" *Ada Lovelace Institute* (10 February 2023), https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/; and Florence G'Sell, *Regulating Under Uncertainty: Governance Options for Generative AI* (September 2024) pp 46 and following, https://cyber.fsi.stanford.edu/content/regulating-under-uncertainty-governance-options-generative-ai.

[63] AI Action Summit, *International AI Safety Report: The International Scientific Report on the Safety of Advanced AI* (January 2025) pp 30 to 36, https://www.gov.uk/government/publications/international-ai-safety-report-2025.

[64] J Cobbe, M Veale, and J Singh, "Understanding Accountability in Algorithmic Supply Chains" (2023) *ACM Conference on Fairness, Accountability, and Transparency*. See also: T Lawton, P Morgan, Z Porter, S Hickey, A Cunningham, N Hughes, I Iacovides, Y Jia, V Sharma, and I Habli, "Clinicians Risk Becoming 'Liability Sinks' for Artificial Intelligence" (2024) 11 *Future Healthcare Journal* 100007.

[65] A Hopkins, S H Cen, I Struckman, A Ilyas, L Videgaray, and A Mądry, "AI Supply Chains: An Emerging Ecosystem of AI Actors, Products, and Services" (28 April 2025) *arXiv preprint*, https://arxiv.org/pdf/2504.20185.

[66] [1932] UKHL 100; [1932] AC 562.

[67] Though given it is doubtful the Consumer Protection Act 1987 applies to software, it is unlikely that a producer of an AI system that was solely in the form of software could be liable in product liability currently.

[68] J Zerilli, A Knott, J Maclaurin, and C Gavaghan, “Algorithmic Decision-Making and the Control Problem” (2019) 29 *Minds and Machines* 555, pp 562 to 563.

[69] Some models are more transparent than others. For example, Meta describes some of its Foundation Models as “open-source”: “Introducing Meta Llama 3: The most capable openly available LLM to date” (18 April 2024) *Meta*, https://ai.meta.com/blog/meta-llama-3/. However, what “open-source” means in the context of AI models is a matter of debate: M Webb, “What do we mean by open-source AI?” (2 August 2024) *JISC National Centre for AI*, https://nationalcentreforai.jiscinvolve.org/wp/2024/08/02/what-do-we-mean-by-open-source-ai/.

[70] 881 N.W.2d 749 (Wis. 2016). See: “State v. Loomis” (2017) 130 *Harvard Law Review* 1530.

[71] 881 N.W.2d 749 (Wis. 2016) at para 133 (Abrahamson J).

[72] Though, as already noted, when asking for an explanation of a person’s behaviour, we do not expect an explanation in terms of their neurons and synapses. Accordingly, some commentators consider there is a double standard regarding our demands for transparency of AI, compared with transparency of natural persons. See: J Zerilli, A Knott, J Maclaurin, and C Gavaghan, “Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?” (2019) 32 *Philosophy & Technology* 661.

[73] Anthropic, “Interpretability”, https://www.anthropic.com/research#interpretability

[74] R Williams, “Rethinking Administrative Law for Algorithmic Decision Making” (2022) 42 *Oxford Journal of Legal Studies* 468 pp 479 to 480.

[75] M Oswald, “Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power” (2018) 376 *Philosophical Transactions of the Royal Society*.

[76] J Zerilli, J Danaher, J Maclaurin, C Gavaghan, A Knott, J Liddicoat, and M Noorman, *A Citizen’s Guide to Artificial Intelligence* (2021) pp 85 to 86.

[77] J Zerilli, J Danaher, J Maclaurin, C Gavaghan, A Knott, J Liddicoat, and M Noorman, *A Citizen’s Guide to Artificial Intelligence* (2021) pp 85 to 86. See also: M Oswald, “Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power” (2018) 376 *Philosophical Transactions of the Royal Society*.

[78] For example, see: *R (on the application of Ames) v Lord Chancellor* [2018] EWHC 2250 (Admin); [2018] Lloyd’s Rep. F.C. 545, where the Legal Aid Agency was found to have unlawfully made an offer in respect of counsel fees where it refused to disclose a calculator it used to determine the fees offer. The failure to disclose the calculator was found to be a breach of the agency’s duty of transparency and clarity.

[79] The Law Society of England and Wales, *Algorithms in the Criminal Justice System* (June 2019).

[80] B Custers, “AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law” in B Custers and E F Villaronga (eds) *Law and Artificial Intelligence* (2022).

[81] K Quezada-Tavarez, P Vogiatzoglou, and S Royer, “Legal challenges in bringing AI evidence to the criminal courtroom” (2021) 12 *New Journal of European Criminal Law* 531.

[82] Note that the Information Commissioner’s Office provides detailed guidance on explaining decisions made by AI. The difficulty raised here is that in (at least) some cases people simply will not be able to fully understand automated decisions made in relation to them. See the discussion in: R Williams, “Rethinking Administrative Law for Algorithmic Decision Making” (2022) 42 *Oxford Journal of Legal Studies* 468, 474 to 476.

[83] The EU’s AI Act distinguishes between AI systems based on the level of risk, with a separate risk category for general-purpose AI specifically.

[84] For an example in England & Wales, see: *Ayinde, R (on the application of) v The London Borough of Haringey and Hamad Al-Haroun v Qatar National Bank QPSC & Anor* [2025] EWHC 1383 (Admin).

[85] *Ayinde, R (on the application of) v The London Borough of Haringey and Hamad Al-Haroun v Qatar National Bank QPSC & Anor* [2025] EWHC 1383 (Admin).

[86] One recent study found that in relation to diagnosing potential diseases from chest x-rays, an AI system, working independently, accurately diagnosed 92% of cases, compared with 74% for radiologists working independently, and 72% for radiologists working with AI assistance. See: A Agarwal, A Moehring, P Rajpurkar, and T Salz,

"Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology" (2023) *National Bureau of Economic Research Working Paper Series* 31422, http://www.nber.org/papers/w31422; and P Rajpurkar, and E J Topol, "The Robot Doctor Will See You Now" (2 February 2025) *The New York Times*, https://www.nytimes.com/2025/02/02/opinion/ai-doctors-medicine.html.

[87] This was unfortunately the result in Australia as a result of the implementation of an automated debt assessment and recovery tool implemented in relation to a welfare payment compliance program. The scheme was unlawful and caused serious harm, resulting in a high-profile Royal Commission, with 57 recommendations. See: *Royal Commission into the Robodebt Scheme* (7 July 2023), https://robodebt.royalcommission.gov.au/.

[88] As noted above, many public authorities have already published such guidance. Government has also published "dynamic" guidelines for the use of AI by the public sector: Government Digital Service, *AI Playbook for the Government* (10 February 2025), https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government.

[89] Copyright and AI: Consultation (2024) Cm 1205.

[90] As noted above, under the Data (Use and Access) Act 2025, the Secretary of State is required to publish an impact assessment and policy report on its options for managing AI and copyright, within nine months of the act coming into force on 19 June 2025. See sections 135 to 137.

[91] *Bartz v. Anthropic PBC*, No 3:24-cv-05417, (N.D. Cal. 2025); and *Kadrey v. Meta Platforms, Inc.*, No. 3:2023-cv-03417, (N.D. Cal. 2025).

[92] Art 6(1)(f) UK GDPR states that the ground of legitimate interests can only be relied upon where the legitimate interests are not overridden by the "interests or fundamental rights and freedoms of the data subject".

[93] A Obermeyer, B Powers, C Vogell, and S Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations" (25 October 2019) *Science*, https://www.science.org/doi/10.1126/science.aax2342.

[94] T Whittaker, R Williams, A Suterwalla, and W Perry, "Public Law and Procurement Law" in M Hervey and M Lavy KC (eds) *The Law of Artificial Intelligence* (2nd ed 2024) para 14-008.

[95] For an example of the challenges in meeting that duty, see: *R (on the application of Bridges) v Chief Constable of South Wales Police* [2020] EWCA Civ 1058; [2020] 1 WLR 5037 (the use of automated facial recognition technology was unlawful in circumstances where, among other things, reasonable steps had not been taken to determine whether the system had a racial or gender bias). Government has provided guidance to public authorities on the use of AI systems: Government Digital Service, *AI Playbook for the Government* (10 February 2025), https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government.

[96] Given the volume of research and analysis on AI and the law, there are plenty of such proposals. We refer to the authorities set out at endnote 41 for discussion of many such proposals.

[97] For a very early example, see: L B Solum, "Legal Personhood for Artificial Intelligences" (1992) 70 *North Carolina Law Review* 1231. See also: S Chesterman, "Artificial Intelligence and the Limits of Legal Personality" (2020) 69 *International and Comparative Law Quarterly* 819; G Teubner, "Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law", Lecture delivered 17 January 2007, *Max Weber Lecture Series*; P Čerka, J Grigiene, and G Sirbikytė, "Is it possible to grant legal personality to artificial intelligence software systems?" (2018) 33 *Computer Law & Security Review* 685; and M Fenwick and S Wrbka, "AI and Legal Personhood" in L A DiMatteo, C Poncibò (eds) *The Cambridge Handbook of Artificial Intelligence* (2022).

[98] J J Bryson, M E Diamantis, and T D Grant, "Of, for and by the People: The Legal Lacuna of Synthetic Persons" (2017) 25 *Artificial Intelligence and Law* 273; and N Banteka, "Legal Personhood and AI: AI Personhood on a Sliding Scale" in E Lim and P Morgan (eds) *The Cambridge Handbook of Private Law and Artificial Intelligence* (2024).

[99] *Bumper Development Corporation v Commissioner of Police of the Metropolis* [1991] 1 WLR 1362.

[100] Te Awa Tupua (Whanganui River Claims Settlement) Act 2017 (New Zealand) s 14(1).

[101] *Hanson Robotics*, https://www.hansonrobotics.com/sophia/.

[102] "An Artificial Intelligence Has Officially Been Granted Residency" *Futurism* (11 June

2017), https://futurism.com/artificial-intelligence-officially-granted-residency.

[103] European Parliament Resolution with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), Recital AC.

[104] "Open Letter to the European Commission: Artificial Intelligence and Robotics" *Robotics Open Letter*, https://robotics-openletter.eu/. As of the date of this Discussion Paper, there were 285 signatories.

[105] For example, see: J Turner, *Robot Rules: Regulating Artificial Intelligence* (2019) Chapter 5; S Chesterman, "Artificial Intelligence and the Limits of Legal Personality" (2020) 69 *International and Comparative Law Quarterly* 819; and G Teubner, "Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law", Lecture delivered 17 January 2007, *Max Weber Lecture Series*.

[106] N Banteka, "Legal Personhood and AI: AI Personhood on a Sliding Scale" in E Lim and P Morgan (eds) *The Cambridge Handbook of Private Law and Artificial Intelligence* (2024) p 628. See also: J Turner, *Robot Rules: Regulating Artificial Intelligence* (2019) p 197; S Chesterman, "Artificial Intelligence and the Limits of Legal Personality" (2020) 69 *International and Comparative Law Quarterly* 819; and L B Solum, "Legal Personhood for Artificial Intelligences" (1992) 70 *North Carolina Law Review* 1231.

[107] See, for example, Harman J in *Re Crestjoy Products Ltd* [1990] BCC 23 at 26.

[108] For example, people with significant control include those who control more than 25% of the shares: Companies Act 2006, Schedule 1A.

[109] R Abbott, *The Reasonable Robot* (2020) pp 50 to 66; and M Lavy KC and I Munro, "Liability for Economic Harm" in M Hervey and M Lavy KC (eds), *The Law of Artificial Intelligence* (2nd ed 2024) paras 7-039 to 7-041.