OLIVIA WHITCROFT

"There is debate over the extent to which the law addresses the risks arising from the use of AI"

Where do *Quantum Leap* and the law meet? Right here in this column, as Olivia explains the legal ramifications of using AI

I wanted to test the accuracy of ChatGPT, so I asked it a question about something I know a lot about. No, not the law, but the best TV programme ever: *Quantum Leap*. The show follows formidable physicist Dr Sam Beckett, leaping through time and putting right what once went wrong. I posed a question to ChatGPT about Ziggy, *Quantum Leap*'s AI with an ego, and ChatGPT informed me: "Ziggy is the AI system created by Dr Beckett's friend and colleague, Admiral Al Calavicci." No, she isn't! I corrected ChatGPT: in fact, Sam created Ziggy.

I worry greatly about *Quantum Leap* misinformation, but I realise the consequences for society of this error are minimal. How about if I were relying on AI to conduct legal research, review contracts or to decide how to advise my clients? Earlier this year, it was reported that New York lawyers had used ChatGPT for case research and submitted fictitious cases to the court in relation to a personal injury claim. The lawyers involved were subsequently fined.

And aside from use by lawyers, how about AI used for autonomous vehicles, healthcare services or to identify criminals? The consequences of an error may be even more severe.

Olivia is principal of the law firm OBEP, which specialises in technology contracts, IP and data protection
𝕏 @ObepOlivia

Quality of output

While the use of AI for complex tasks can improve the accuracy and quality of outputs, there are also risks that outputs will be inaccurate, harmful, biased or unethical.

There are many recent examples. In October 2023, the UK Information Commissioner's Office issued Snap with a preliminary enforcement notice over the potential failure to properly assess the privacy risks posed by its generative AI chatbot "My AI", in particular for children.

> **"There are risks that outputs will be inaccurate, harmful, biased or unethical"**

In September 2023, it emerged that the Polish data protection authority was investigating a complaint made by a security and privacy researcher. He claimed that ChatGPT had generated inaccurate information about him, and OpenAI had failed to correct the errors upon request.

A month earlier, researchers from the UK and China published a paper in relation to pedestrian detection by driverless cars. The research found that systems were less accurate in detecting children compared to adults, and dark-skinned pedestrians compared to those with lighter skins.

Quality of input

Of course, the quality of the outputs of an AI system depends on the quality of the inputs. The *Quantum Leap* episode "Return of the Evil Leaper" features Alia, who seeks to counteract Sam's actions by putting wrong what once went right. Whenever Alia appears, Ziggy's calculations are put off track, as presumably Alia wasn't part of Ziggy's original training data.

If the training data has issues with quality, then the output is likely to carry the same issues. Training data may, for example, be inaccurate, biased or out-of-date, or may not take into account all factors relevant to the decisions the AI is making.

In 2018, reports revealed that Amazon had scrapped an AI recruiting tool that favoured men over women, the issue being that its training data was based largely on male applicants. There are many reports of facial recognition systems being less accurate at recognising black women than white men, having been trained on data sets primarily involving white men. The quality of database images can also be a factor, as some camera settings may be less effective at capturing darker skin tones.

In the recent complaint made to the Polish data protection regulator referred to above, the individual also claimed that OpenAI had failed to give him access to training data used to generate the allegedly inaccurate information about him.

Regulation – the EU approach

There is debate over the extent to which the law addresses, or should be created to address, the risks arising from the use of AI. The European Union has gone down the route of new regulation, with the EU AI Act currently going through the legislative procedure.

The EU AI Act classifies different AI models according to risk, and creates different obligations for each level of risk. AI systems creating an "unacceptable risk" are prohibited. This includes those that deploy subliminal techniques to distort people's behaviour and cause harm. AI systems creating a "high risk" include remote biometric identification, and those intended to be used as safety components of products or critical infrastructure. These are subject to a set of rules, including on risk management, data governance, transparency, human oversight, accuracy and security.

There are also transparency obligations for lower-risk AI systems that are intended to interact with people, such as chatbots, biometric categorisation and content-generating systems. Some AI systems that do not fall within any of these categories are deemed lower risk and will not be subject to the obligations.

BELOW **Snapchat could be required to stop processing data from My AI**

Principles – the UK approach

In March 2023, the UK government published a white paper for consultation: "A proinnovation approach to AI regulation" (the UK AI White Paper). The consultation was open until 21 June 2023. At the time of writing, the results of the feedback have yet to be published.

To support AI innovation, the UK government intends to issue a set of five principles to be implemented by existing regulators. The paper refers to laws that can already address some of the risks posed by AI, including equality, data protection, human rights, product safety, competition, consumer rights and tort laws.

The proposed principles are: safety, security and robustness; appropriate transparency and explainability (including enabling understanding of how decisions are reached); fairness (including not undermining legal rights or discriminating unfairly); accountability and governance (including effective oversight); and contestability and redress.

In November 2023, the Artificial Intelligence (Regulation) Bill (a Private Members Bill) was introduced into the House of Lords. It seeks to establish an AI Authority to oversee other regulators' approach to AI (taking into account the principles), construct regulatory sandboxes, and require businesses to have an AI responsible officer. It also envisages transparency from those training AI, including giving users health warnings, and clarifying use of others' data and intellectual property.

Human oversight and contestability

AI systems use probabilities to decide what action to take. In *Quantum Leap*, Ziggy regularly calculated the odds that Sam was there to fix a particular problem, or that an action would result in particular consequences. In "Return of the Evil Leaper", Sam leapt into the Midnight Marauder, and Ziggy calculated overwhelming odds that Sam was going to be killed if he tried to stop some chicken races. Sam decided to scoff at the odds and try to stop the races.

Humans have the ability to exercise discretion and make their decisions beyond the AI system's probabilities. They may spot obvious errors or bias, or take into account wider ethical or moral concerns. Human oversight, and the ability to contest outcomes, may therefore be important safeguards for some of the AI risks discussed above.

Human oversight is one of the EU AI Act's rules for high-risk AI, which includes that a human shall "be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system". Contestability and, linked to that, explainability, form part of the UK government's proposed principles in the UK AI White Paper. These aim to enable impacted parties to understand and contest harmful decisions or outcomes made by AI.

The UK GDPR already contains a prohibition on solely automated decision-making, meaning a decision made by a computer without meaningful human involvement. The restriction applies where the decision has legal effects for an individual or otherwise significantly affects them. This may include, for example, decisions about credit applications, recruitment, access to medical treatment or insurance premiums. There are exceptions to the rule, but the logic and consequences of the decision must be explained to the individual, and other safeguards must be in place, including the right to obtain human intervention and to contest the decision.

Under the UK Data Protection and Digital Information Bill (DPDI), which is still going through parliament as we go to press, the scope of the overall prohibition will be reduced, meaningthat many solely automated decisions will no longer be barred. During the consultation in 2022, concerns were expressed that the changes could have a disproportionately negative impact on people with protected characteristics, such as sex or race. An example was the claim that the 2020 A-level results algorithm produced different outcomes based on these characteristics. DPDI retains the same safeguards of human intervention and contestability, which seek to address some of these concerns.

Responsibilities

Organisations also need to consider who is legally responsible for AI-related harms, and how to manage these along an AI supply chain. In the UK, we need to look at existing laws to identify responsibility, such as controllers or processors under data protection laws, and manufacturers under product safety laws. The EU AI Act allocates responsibilities between "providers" and "deployers"; a provider being a party that develops an AI system, and a deployer being a party using an AI system.

ABOVE An AI error in a driverless car could have serious consequences

"Human oversight may be an important safeguard for some of the AI risks"

The UK AI White Paper recognises that AI supply chains can be complex and opaque, making risk management along the chain difficult. The paper suggests that assurance techniques and technical standards may assist with risk management, as part of its "Accountability and Governance" principle.

Contracts may assist to provide remedies along the supply chain. For example, if a business is deploying a new AI system developed by a technology company, the contract between the parties can allocate responsibilities between them. It may be that the business is providing some of the training data, in which case the technology company will not want to be liable for errors caused by low-quality data from the business.

Future risks?

I was delighted to watch the *Quantum Leap* reboot earlier this year. Though ChatGPT knew nothing about the reboot, as its last knowledge update was in 2021 (this was using GPT-3.5 rather than GPT-4). The reboot features a new leaper and a new Project *Quantum Leap* team, but still the same faithful AI, Ziggy. Or was she? Because – spoiler alert – as it turned out, Ziggy may well be the future mole allowing information to be leaked to Evil Leapers. Is this another potential AI risk to be addressed?

olivia.whitcroft@obep.uk

BELOW US TV show *Quantum Leap* has been rebooted for the 21st century