Gli GLOBAL LEGAL INSIGHTS

AI, Machine Learning & Big Data 2025

Seventh Edition

Contributing Editor:

Charles Kerrigan

CMS LLP

glg Global Legal Group

TABLE OF CONTENTS

Preface

Expert Analysis Chapters

Jurisdiction Chapters

Autonomous AI: Who is responsible when AI acts autonomously and things go wrong?

Erica Stanford

CMS LLP

This chapter explores what "autonomous AI" is, what can happen and what can go wrong, and examines the assignment of responsibility or liability when an AI system causes unpredictable harm, how legal systems in key jurisdictions are beginning to regulate it, and some best practices to avoid the worst-case scenarios.

What is autonomous AI?

Autonomous AI generally refers to an AI system that can act or make decisions without ongoing human intervention or approval. Here, "autonomous" refers to AI's ability to operate with minimal or no real-time human oversight, adapt behaviour as it learns or encounters new data, and make its own decisions. Once purely theoretical, this concept has evolved over recent years to a practical reality, and its decisions and outcomes can carry real-world consequences.

Traditional software systems typically follow deterministic fixed rules encoded by human developers. Human programmers define the parameters and instructions line by line, and the software behaves as expected unless it malfunctions due to a traceable defect, which can generally be found and attributed to a responsible human.

By contrast, AI systems use machine learning models – often neural networks – to learn from data and adapt their behaviour over time. This learning capacity can bring benefits beyond human potential, such as real-time navigation by self-driving vehicles or personalised medical diagnostics. It also means the system may evolve in unexpected and unforeseen ways, potentially operating as a "black box" whose decisions and outputs give no explanation as to their reasoning and where even its creators cannot always fully anticipate, understand, or explain its outputs, especially if the environment shifts or data inputs deviate from those in initial training.

Simplistically, the more data AI is trained on and the more computer power available, the more intelligent AI gets. Whilst intelligence with regard to AI is both debated and subjective, the more AI can parse through enormous amounts of data, understand that data, and make decisions at speeds and volumes beyond human ability, the less humans can follow or understand its reasoning, especially where the AI systems do not give transparency as to their decision-making or "thought" process. This combination of complexity and opacity makes it harder to calculate foreseeability, liability, and how to safeguard against harm when no human operator is continuously "in the loop".

Autonomous AI systems – with their capacity for real-time data processing, advanced pattern recognition, and the ability to operate with minimal human intervention – are transformative. They also create new legal and ethical challenges. Central among these is determining who is responsible, who to hold to account, and how to hold humans or corporate entities to account when AI systems operate seemingly of their own accord and in ways even their developers did not fully predict, with potentially harmful or disruptive outcomes.

In many ways, even "autonomous AI" is still bounded by human involvement: the ways AI models are trained and deployed and the priorities set for their training and deployment are shaped by corporate policies, data curation, regulatory constraints, time and budget constraints, and potential biases and preferences. Increasingly, however, these AI systems, while shaped by humans, evolve beyond their initial training and can appear to have a mind of their own. Unexpected outcomes might be attributed to emergent properties that arise from the complexity of their models, the data they are trained on or their operating environment, or because adversarial actors intentionally manipulate their inputs in a process known as data poisoning. Unexpected outputs, such as the hallucinations often seen through Generative AI Large Language Model (LLM) chatbots, are common, but a system that drives a vehicle, moderates social media content, or diagnoses medical conditions might make subtle changes to its decision-making process after deployment, learn from its accumulated experience or environment in ways that its creators or users find difficult to interpret, and cause real-world consequences.

Some examples of consequences

In 2018, an autonomous test vehicle operated by Uber in Arizona fatally struck a pedestrian pushing a bicycle at night. Investigations revealed that the vehicle's AI detected the pedestrian but did not identify her as a hazard demanding emergency braking. A human safety driver was present but was streaming a video instead of concentrating on the road.[1] In this instance, prosecutors charged the negligent driver rather than Uber as a corporate entity. Tesla's "Autopilot" feature has been linked with hundreds of crashes – some fatal – prompting debate over whether the drivers were negligent for failing to supervise the system properly or whether Tesla's marketing and technical design created undue reliance on what is perhaps a semi-autonomous instead of a fully autonomous driving feature.[2] In both situations, whilst the AI's interpretation of data and its decisions were central to causing harm, the broader context of human oversight, corporate design decisions, regulatory frameworks, and checks for worst-case scenarios – or the potential lack thereof – were a large part of the cause of the outcome.

Even in non-physical instances, AI acting autonomously and unexpectedly can still have significant real-world consequences. Several episodes illustrate how the technology can behave in discriminatory or harmful ways without direct malicious intent by its human creators. When Microsoft launched its "Tay" chatbot in 2016, adversarial users used inflammatory content on Twitter to manipulate it into making racist and offensive comments. There was no formal lawsuit – although Taylor Swift's lawyers reportedly tried to sue, it seems they were more concerned over the likeness to the singer's name[3] – however, the incident still caused some reputational damage. A recruitment tool trialled by Amazon downgraded CVs referencing female applicants, having "learned" that historical hiring data associated successful tech candidates with predominantly male-dominated backgrounds.[4] Amazon stopped the project before any legal claims emerged. In finance, in the 2010 "Flash Crash", high-frequency trading algorithms – each functioning according to its own complex logic – interacted in a feedback loop that sent the market plummeting nearly 1,000 points in minutes. In this instance, market regulators were unable to fully untangle responsibility in a context where no single entity appeared directly at fault.

Autonomous AI systems could also bring about real-world consequences – perhaps sooner than many might expect – in governmental, policing, judicial, security, or surveillance contexts. Law enforcement agencies in various jurisdictions are already experimenting with predictive policing tools that claim to anticipate criminal activity based on historical data. The trouble is that historical data may be racist or biased, as can be training methods, which can result in algorithms embedding and potentially exacerbating systemic biases that disproportionately target certain groups or communities. Some jurisdictions have begun to trial risk-assessment software to aid bail or sentencing decisions in courts, which could increase algorithmic discrimination and due process violations, especially if judges or other individuals or systems over-rely on the opaque outputs of machine learning models. Government agencies introducing AI for immigration and welfare determinations risk increasing the risk of wrongful denials or approvals when human oversight diminishes. In the military and in defence, the development of autonomous weapons and surveillance drones means that lethal decisions could, in effect, be delegated to algorithms. This delegation could be especially dangerous if AI starts making other decisions or the algorithms or data they rely on are hacked or otherwise manipulated or changed without the knowledge or testing of their handlers. AI's capacity for large-scale data analysis and pattern recognition can help with intelligence gathering. This capability can also easily slide into mass surveillance if unchecked, undermining civil liberties and human rights. In law and commerce, AI-driven software may draft legislation, manage entire supply chains, or recommend policy reforms, potentially centralising power in systems whose logic and decision-making remain partially untraceable and inscrutable.

What can cause AI to act unpredictably and cause things to go wrong?

One problem is that the more data and parameters involved, the more unpredictable AI can become. Large neural networks, the foundation of many AI models, can contain millions or billions of parameters, making them intrinsically difficult to interpret. Even if the general functioning of the model is transparent and understood, the precise pathways through which it processes a given set of inputs may remain opaque. Humans also cannot keep up with the speed and volume at which AI is able to process information. Another challenge that adds to its unpredictability is that AI does not always return the same output for the same prompt or ask. These factors do not necessarily mean that AI behaviour is entirely ungovernable; rather, it may be foreseeable that a self-driving vehicle could fail to recognise an obstacle or that a chatbot could be manipulated into generating offensive content and, arguably, attributing responsibility should involve anticipating worst-case scenario outcomes.

Data poisoning – where data is slowly and imperceptibly altered or manipulated until the AI system using it misconstrues or misclassifies critical information and gives different results – is one of the most potentially dangerous manifestations of AI's vulnerability and how AI can be manipulated to give unpredictable or malicious outcomes. To give some examples of how data poisoning could cause life-threatening incidents, a self-driving car might be trained on image datasets that include doctored photographs of road signs. If these doctored images "teach" the AI to register a stop sign as a speed limit sign under specific conditions, the resulting behaviour on public roads could cause a crash. Data poisoning in the datasets used to train autonomous weapons, medical devices or diagnosis tools or in surveillance, policing, or justice could have equally harmful consequences. A hacking group or rogue nation-state could cause chaos by manipulating the datasets driving social media or government outcomes, which might be their goal. Microsoft's chatbot fiasco illustrates another form of poisoning, carried out simply by malicious prompts rather than hacking. Sophisticated attackers may, at any stage, discover and exploit vulnerabilities, and this is to be expected and should be planned for.

Data poisoning raises questions about where blame and liability lie. It is easy to blame the malicious actor alone. In principle, malicious actors remain criminally and civilly liable for tampering with AI. However, there is little benefit to attributing blame to an organised crime gang, rogue nation-state, or state-sponsored hacking gang that may lie in a jurisdiction outside of international reach, such as North Korea. This leaves the harmed party to seek compensation from the AI provider or operator, who may contend that an unforeseeable, sophisticated cyber-attack amounts to a superseding cause. However, courts and regulators are likely to examine whether adversarial attacks were widely known to be feasible in that sector. If so, providers or operators who took inadequate precautions or who neglected proper precautions may be held liable for failing to maintain sufficient cybersecurity. This is like established norms in data breach law, where organisations can be held responsible if they did not implement protective measures consistent with current best practices.

Regulatory trends suggest that companies will be expected to demonstrate that they have taken "reasonable" steps to address foreseeable hazards, even if the specific manifestation of a machine-learning error was not anticipated.

While direct blame might be attributed to a malicious actor, there should also be some responsibility at the organisational level. Should the developers be held accountable for failing to embed sufficient safeguards, the compliance or cyber security team for insufficient testing, the leadership team for turning a blind eye to a lack of security checks, or the organisation for implementing or allowing a culture that may prioritise speed or profit over safety checks? Or should the regulator be held accountable for creating an environment where these risks were able to take place? Given that data poisoning or adversarial manipulation can be foreseen as a common hazard in any AI system, a reasonable duty of care includes extensive pre-emptive adversarial testing and rapid patching of discovered flaws.

Challenges of assigning responsibility to seemingly autonomous AI

The legal question that follows from the possibility of AI acting unpredictably is: who should be held responsible when things go wrong? Even if the autonomous system initiates the harmful act, a human or organisation must still be at fault. Machines cannot stand trial or pay damages. They have no moral agency, and it would be counterproductive for victims to have to pursue compensation from a non-human entity in situations such as wrongful arrests or serious accidents. At some stage, humans are responsible for how and why the AI system did what it did. Assigning personhood to AI would risk letting companies off the hook, in effect encouraging them to disclaim control over the "independent" machine.

The relevant question becomes who among the network of human and organisational actors – programmers, managers, compliance, security or legal teams, corporate boards, end-users, or regulators – ought to bear ultimate responsibility. Determining responsibility can be complex even for simple digital tools, but the complexity is intensified when AI evolves, adapts, or just acts unexpectedly post-deployment. A self-learning algorithm might pick up harmful patterns or biases as it processes new data, or it might fall prey to adversarial attacks that its creators did not anticipate. In these scenarios, the harm a machine causes may no longer be traceable to a single line of code, a single design decision, or a single oversight by a compliance team. This can produce a temptation to blame the AI itself, but the legal consensus in most jurisdictions is that responsibility should remain anchored to the human and corporate entities who developed, deployed, or supervised it.

Where large, well-resourced companies are involved, victims might seek to recover damages by alleging flaws in the AI system under traditional product liability or negligence principles. If a self-driving car's sensor suite malfunctions or a chatbot defames someone, an injured party could initiate a claim by arguing that the manufacturer or operator owed a duty of care and failed to meet it. This might mean the developer did not conduct adequate tests before release or did not find and patch security vulnerabilities that made data poisoning possible. The person who suffers harm – whether physical, reputational, or financial – may not need to prove the precise technical cause, only that the defendant fell below the standard of reasonable care for that industry or that the product was "defective" in the sense used by product liability statutes.

It becomes harder to attribute liability if the developer can plausibly claim that they had no way to predict a certain outcome or that malicious tampering was to blame. Claimants then need to show that the particular risk was still reasonably foreseeable, that the developer did not implement best practices or recommended safeguards, or that the user was misled about the technology's capabilities. If the AI was highly experimental or unregulated, and the user or regulator was aware of the risks, the matter becomes more complicated. Existing legal principles are now being tested by AI's capacity to learn and adapt. Historically, the notion of a "defect" has been easier to identify in a static product: a broken design or a manufacturing fault. Responsibility becomes even more unclear when the "defect" emerges only after an AI system shifts its parameters based on new data, and an AI continues to learn after deployment.

There is also the matter of development risk defences, which allow companies to say that the defect was unknown and unknowable at the time of deployment. This could be invoked in situations where a machine-learning system behaves in ways no one had anticipated. The very premise of machine learning is, however, that unexpected behaviours are a predictable category of risk, even if one cannot predict the specific manifestation. Organisations might, therefore, be expected to implement robust fail-safes, real-time monitoring, or ways to revert to a safe fallback mode when anomalies arise. Even so, the prevailing stance in law is that developers and operators cannot simply disclaim responsibility by pointing to the AI's autonomy. Courts increasingly treat AI's capacity for post-deployment adaptation as a normal and expected factor of machine learning. As such, responsible parties must guard against known risks, track the system's performance, and address vulnerabilities that come to light.

The complexity of data supply chains also makes it harder to clearly attribute responsibility. AI systems may rely on layered algorithms, open-source libraries maintained by a global community of developers, or data from third-party sources. If a harmful flaw arises in a library or dataset that the primary developer never examined closely, is it fair to assign liability to the developer, the open-source contributors, or the end-user who integrated the component? Traditional doctrines such as indemnification or contributory liability can help to allocate responsibility, but they can leave victims uncertain about where to direct a legal claim.

Responsibility gaps

One problem in assigning responsibility is the potential for "responsibility gaps", where the complexity and semi-autonomous nature of AI leads every stakeholder to disclaim liability and try and pass the blame onto someone else. A developer might say they merely coded the underlying algorithm, a data curator might argue they had no knowledge of how the data would be used, and a corporate executive might insist that direct oversight lay elsewhere. This diffusion of accountability is exacerbated by the fact that responsibility in AI is already hard to attribute. Additionally, many AI-driven systems operate across jurisdictions and industries with variable regulatory controls.

A variety of proposals have emerged to pre-empt or minimise the problem of "responsibility gaps", one is to channel liability towards a single entity – often the AI's developer or the operator deploying it – analogous to how nuclear law channels responsibility to the nuclear facility operator, with mandatory insurance to ensure victims are compensated. Another is to mandate pre-market certification of high-risk AI systems, resembling the rigorous testing protocols for new aircraft. This might include ongoing audits, forced disclosure of training data and model performance metrics, and a capacity to mandate product recalls if evidence of dangerous AI behaviour emerges. Another approach is to embed "human in the loop" designs in critical systems. Alternatively, if the AI is genuinely fully autonomous, legal frameworks would treat it as a product or service for which the deploying organisation bears entire responsibility.

One possibility is a regime that sets out layered responsibilities for each actor in the AI development chain, from those providing raw datasets to those making final deployment decisions. Such a regime could circumvent the risk of a "responsibility gap" by ensuring that liability attaches proportionally to each contributor's degree of control or benefit from the AI system. In some proposals, developers would be responsible for ensuring the algorithmic model meets certain safety and transparency thresholds, data suppliers would need to demonstrate that their datasets are free from known biases or tampering, and final deployers would bear responsibility for ensuring the model is used only in contexts for which it is suitable. Mandatory auditing and licensing requirements could help ensure diligence.

What the law says: legal frameworks around the world

Jurisdictions around the world have taken varied approaches in attempting to regulate or oversee the development of autonomous AI. In the United Kingdom (UK), the legislative focus has largely been on incorporating AI into existing frameworks while offering sector-specific updates. The UK Government's 2023 White Paper on AI sets out five principles – safety, transparency, fairness, accountability, and contestability – and expects regulators to apply these within their existing remits. Though the UK does not possess a dedicated AI liability statute, it does have pertinent legislation for specific applications. For example, the Automated and Electric Vehicles Act 2018 requires insurers to cover accidents caused by automated vehicles and then send claims back to the manufacturer if a defect in the autonomous system caused the harm. There are discussions around further legislation identified in some sources as the Automated Vehicles Act 2024, which would clarify that once a vehicle is in full self-driving mode, the liability would attach to an "authorised self-driving entity" rather than the human occupant. Outside the automotive sector, the UK continues to rely on product liability under the Consumer Protection Act 1987 (implementing EU Directive 85/374/EEC) and negligence law, with courts likely to test whether an AI developer or deployer acted with reasonable care.

The EU's regulatory trajectory is shaped by comprehensive rules and the EU AI Act's risk-based classification system for AI systems. Applications deemed "high-risk", such as certain healthcare applications and self-driving vehicles, must meet stringent requirements for accuracy, transparency, and human oversight. Failure to comply can lead to substantial fines. Companion legislation sometimes referred to as the New Product Liability Directive (2024/2853) expands the definition of "product" to include intangible AI software and introduces provisions to alleviate claimants' burden of proof when challenging complex AI. Courts can, for example, order disclosure of technical documentation to address the black-box challenge and may presume defectiveness if the defendant cannot satisfactorily demonstrate otherwise. An AI Liability Directive was proposed to harmonise fault-based civil liability but did not proceed due to political hurdles. Nevertheless, the current mix of updated product liability rules, the AI Act, and national tort laws across EU Member States collectively bring a comparatively robust environment in which those harmed by AI systems can seek redress.

In the United States, legal principles can vary from state to state, and there is currently no unified AI liability regime at the federal level. Liability claims involving AI are predominantly channelled through traditional doctrines of negligence, product liability, and consumer protection, distributed across state jurisdictions. When self-driving cars crash, US courts increasingly look at whether the manufacturer misled consumers about the capabilities of the technology or failed to include safety features that a reasonable industry player would have implemented. Tesla has faced numerous lawsuits alleging that Autopilot's marketing, combined with software design, contributed to crashes in ways that might be construed as product defects or as a "failure to warn".[5] States often require companies testing driverless cars to shoulder liability, while federal agencies like the National Highway Traffic Safety Administration set vehicle safety standards. Agencies such as the Food and Drug Administration regulate medical AI devices, and the Securities and Exchange Commission addresses algorithmic trading. Recent non-binding efforts, including the White House's "Blueprint for an AI Bill of Rights" and the National Institute of Standards and Technology's AI Risk Management Framework, suggest the US Government's aspiration to clarify best practices, but these do not themselves create enforceable legal obligations. Even so, US case law is slowly accumulating. Litigation outcomes remain varied, reflecting the complexities of fault in partially automated driving environments where drivers retain some measure of oversight. Some states are also experimenting with laws requiring transparency in AI-driven hiring processes or anti-discrimination obligations in automated decision-making.

Some core legal doctrines repeatedly emerge. Negligence, in a common law sense, requires a duty of care, a breach of that duty, and causation leading to harm that is reasonably foreseeable. The nature of black-box

AI, however, is that a specific error or emergent behaviour might not have been anticipated, even if the general risk of malfunction was known, and a malfunction, even if not a specific one, could be predicted. Courts may resolve this by taking a higher-level view, reasoning that using a self-learning algorithm in a sensitive context is a choice with foreseeable hazards and that the onus is on the developer or deployer to mitigate these risks appropriately. Product liability doctrine, which includes strict liability for defective products, simplifies the injured party's burden by requiring proof of defectiveness rather than proof of negligence. It still prompts questions about whether stand-alone software constitutes a product and, if so, how a "defect" is to be defined in the context of constantly evolving AI.

Despite these legal frameworks, ascertaining responsibility still presents challenges. One challenge is that courts are used to seeing causation and fault as discrete events or decisions. When an AI system continues to learn after deployment, the cause of a harmful result can be diffuse and, at best, unclear. This is made harder in jurisdictions that do not have a well-defined stance on whether intangible software is a "product" covered by liability regimes and that may require statutory updates or creative judicial interpretation to capture AI-based tools.

An ongoing question is who is best placed to foresee harm and enforce best practices, which may be any of the leadership, developers, compliance, security or cybersecurity teams, product managers, auditors or even users.

Possible solutions and practical guidance

One pragmatic step is to implement solid governance and oversight frameworks within organisations. This might include teams or committees incorporating legal, technical, compliance, ethics and risk experts to review proposed AI deployments for safety and ethical considerations. These reviews can identify potential biases in training data, highlight vulnerabilities to adversarial attacks, and ensure there is a protocol for intervention if the AI starts malfunctioning.

Contractual arrangements can also help. A company that purchases or licenses AI technology from an external provider might demand warranties on reliability, indemnities if the software fails in predictable ways, and obligations to patch or update vulnerabilities quickly. Disclaimers of liability may still be tested in court if serious harm occurs, but agreements that spell out each party's responsibilities can at least align expectations and create an audit trail of which organisation controlled each aspect of the AI's lifecycle.

Insurance may be a solution. For instance, product liability insurance can be expanded to encompass software-based autonomous systems, while cyber insurance can cover malicious attacks or data poisoning. Insurance for high-risk AI scenarios – such as surgical robots, automated trading platforms, or large-scale recommendation engines – could offer a measure of certainty for victims seeking compensation. Insurance providers are also likely to demand robust risk assessments, which puts commercial pressure on organisations to adopt safer AI practices. If a particular use of AI is deemed too high-risk, the premiums might become prohibitive unless the organisation can show strong safety measures. Over time, this could push developers to invest more heavily in explainability, testing, and adversarial defence.

Good organisational culture plays a role. Employees might sidestep thorough testing or skip routine audits if a corporation incentivises rapid releases and minimal oversight, whereas companies that embed a genuine "safety first" or "ethics first" ethos are more likely to surface and address problems early. Organisations that reward or incentivise staff who discover security holes or raise ethical concerns before real-world harms occur create an environment that fosters collaborative accountability rather than creating scapegoats after the fact. Such a culture aligns with the idea of "proactive compliance", where a business tries to anticipate legal and ethical duties rather than waiting for a court to impose liability.

Many machine-learning systems degrade over time or shift unexpectedly when user behaviour, market conditions, or relevant data distributions change. Regular reviews not only reduce risk but also generate an audit trail demonstrating the organisation's commitment to best practices, which can be crucial if it ever faces legal scrutiny. Advanced logging and explainability tools aid in post-incident investigations and can help show what the AI "saw" and how it reached a conclusion. A carefully maintained "black box" of AI decision-making can be as critical to liability defence as it is for diagnosing root causes of failures. Education and interdisciplinary collaboration also play a growing role. Lawyers may need to develop an understanding of AI's technical aspects, while data scientists are being encouraged to acquaint themselves with legal and ethical guidelines.

"Explainable AI" (XAI) is often offered as a partial technical fix for ascribing liability. The assumption is that by providing explanations of how an AI arrived at a decision after it happened, one can identify faults or at least see whether a developer or end-user missed critical warning signs. But explainability by itself is not good enough. A model might highlight which input features or "weights" influenced a decision, but that does not always reveal the full reasoning process or broader design flaws. Focusing on an individual explanation can also distract from systemic issues and may deflect attention from broader organisational responsibilities, such as managers pressing for rushed rollouts without adequate safety testing or product teams overlooking robust adversarial testing protocols. Realistically, the black-box nature of modern machine learning will persist, even with sophisticated explainability tools, so that blame can still be elusive.

Conclusion

For now, the clearest guiding principle is that any apparent "autonomy" of AI does not cancel out the fundamental requirement that human or corporate actors remain accountable. The consistent message from regulators and courts is that, even for autonomous AI, ultimate responsibility must remain anchored to human decision-makers.

●●●

References and further reading

Berber, A. and Srećković, S. When something goes wrong: Who is responsible for errors in ML decision-making? *AI & Soc* 39, 1891–1903 (2024). Available at: https://doi.org/10.1007/s00146-023-01640-1

Downer, J. (3AD). The Limits of Knowledge and the Sociology of Inevitable Failure. *American Journal of Sociology*, [online] 117(3), pp 725–762. Available at: https://www.jstor.org/stable/10.1086/662383

Frazer, H. and Suzor, N. (2024). Locating fault and responsibility for AI harms: A systems theory of foreseeability, reasonable care and causal responsibility in the AI value chain. *Law, Innovation and Technology, 17(2). (In Press).* Available at: https://doi.org/10.1080/17579961.2025.2469345

Porter, Zoe; Ryan, Philippa; Morgan, Phillip; Al-Qaddoumi, Joanna; Twomey, Bernard; McDermid, John; and Habli, Ibrahim. (2023). *Unravelling Responsibility for Ai*. Available at SSRN: https://ssrn.com/abstract=4871675 or https://dx.doi.org/10.2139/ssrn.4871675

Porter, Zoe; Al-Qaddoumi, Joanna; Ryan, Philippa; Morgan, Phillip; McDermid, John; and Habli, Ibrahim. (2023). *Unravelling Responsibility for Ai*. Available at: https://dx.doi.org/10.2139/ssrn.4871675

Santoni de Sio, F. and Mecacci, G. Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them. *Philos. Technol.* 34, 1057–1084 (2021). Available at: https://doi.org/10.1007/s13347-021-00450-x

Yazdanpanah, V.; Gerding, E.H.; and Stein, S. *et al.* Reasoning about responsibility in autonomous systems: challenges and opportunities. *AI & Soc* 38, 1453–1464 (2023). Available at: https://doi.org/10.1007/s00146-022-01607-8

Endnotes

[1] https://www.bbc.co.uk/news/technology-54175359

[2] https://www.washingtonpost.com/technology/2023/06/10/tesla-autopilot-crashes-elon-musk

[3] https://www.theverge.com/2019/9/11/20860659/taylor-swift-microsoft-tay-chatbot-lawsuit-sue-lawyers-legal-action

[4] https://www.technologyreview.com/2018/10/10/139858/amazon-ditched-ai-recruitment-software-because-it-was-biased-against-women

[5] https://www.theverge.com/2024/4/26/24141361/tesla-autopilot-fsd-nhtsa-investigation-report-crash-death

Erica Stanford

Tel: +44 207 067 3437 / Email: erica.stanford@cms-cmno.com

Erica Stanford is the bestselling, award-winning author of *Crypto Wars: Faked Deaths, Missing Billions and Industry Disruption*, which received a "Highly Commended" accolade at the Business Book Awards. She holds a non-legal position at international law firm CMS, where she advises on digital assets and AI in a non-legal capacity. Erica has authored chapters on misinformation and disinformation, AI in public policy and governance, and AI in law firms and legal technology in the forthcoming 2025 textbook on AI law and regulation. She has also co-authored important works on ethical AI and AI regulation, notably featured in *GLI – AI, Machine Learning & Big Data 2024*. Erica is currently pursuing a Master's degree in Data and AI Ethics at the University of Edinburgh.

CMS LLP

Cannon Place, 78 Cannon Street, London EC4N 6AF, United Kingdom
Tel: +44 207 367 3000 / URL: www.cms.law

Global Legal Insights – AI, Machine Learning & Big Data provides analysis, insight and intelligence across 22 jurisdictions, covering:

- Trends
- Ownership/protection
- Antitrust/competition laws
- Board of directors/governance
- Regulations/government intervention
- Generative AI/foundation models
- AI in the workplace
- Implementation of AI/big data/machine learning into businesses
- Civil liability
- Criminal issues
- Discrimination and bias
- National security and military

globallegalinsights.com