CLIFFORD

CHANCE

WHO'S RESPONSIBLE FOR AGENTIC AI?

MAY 2025

AUTHORS

Tim Mackey
Corporate Officer
Chief Legal Officer & Group Compliance Officer at SoftBank Group Corp

Tim Mackey

Tim Mackey is a Corporate Officer and serves as Chief Legal Officer and Group Compliance Officer of SoftBank Group Corp., leading the global Legal, Compliance, and Government Affairs functions.

In this role, he oversees legal risk, regulatory compliance, corporate governance, and government relations across SoftBank Group. Before joining SoftBank Group, Tim was a Tokyo-based attorney at two US law firms, with over 25 years of legal experience in Japan.

Jonathan Kewley
Partner and Co-Chair of the Global Tech Group
London
T: +44 207006 3629
E: jonathan.kewley@cliffordchance.com

Jonathan Kewley

Jonathan Kewley is the Co-Chair of the Clifford Chance Tech Group. He was recently voted Partner of the Year at the British Legal Awards. He has a particular focus on AI, tech innovation and cyber security.

Jonathan is a leading advisor to some of the world's largest technology companies, spanning the West Coast of the US to China. His work includes AI investment, facilitating the global rollout of AI solutions and Tech crisis management. He co-leads a team of over 600 Tech lawyers worldwide, supporting companies in their pursuit of innovation.

WHO'S RESPONSIBLE FOR AGENTIC AI?

We're now entering the era of agentic AI where advanced autonomous systems can execute not just simple prompts but entire plans and strategies with minimal human input. These independent, fast-paced agents offer huge possibilities: freeing up time, executing tasks faster and doing our bidding. Teams of AI agents working around the clock have the potential to turbocharge productivity in the workplace. They could also make life easier at home – paying our bills, ordering our shopping and providing an extra pair of hands.

But with their increasing use, there is also the very real possibility that at some point they will act in unintended ways and cause financial, reputational or even physical harm. When that happens, people will turn to legal systems.

Who is responsible?

By setting out clear and predictable rules, we can support innovation with guardrails. One provocative idea has emerged: Should an autonomous AI itself be considered a legal person responsible for its actions? This question poses new challenges for law and policy. Generally speaking, existing laws and regulations around the world don't provide a clear answer on who is responsible and how to seek redress when it comes to harm caused by agentic AI. Ever since legal systems were first designed, they have only had one ultimate subject: humans. Whether laws are applied to individuals, corporations or nation-states, responsibility is ultimately traced back to human decision-makers.

AI challenges these paradigms. The hallmark of advanced AI is its autonomy – the ability to make decisions without being explicitly programmed for every scenario. Agentic AI further increases the gap between an original human instruction and the ultimate output by enabling systems to take multiple independent steps to achieve the outcome, abstracted from humans. Typically, the more remote an initial human decision is from the output of an AI system, the harder it becomes to ascribe responsibility for the AI's action to that human. This 'gap' in accountability has been noted by scholars (often called the 'AI responsibility gap'), and the recent increase in autonomy has led some to suggest that perhaps the AI itself, rather than any particular human, might need to bear responsibility in such cases – essentially treating the AI as a legal entity. Yet this notion of AI legal personhood collides with traditional frameworks and raises both promising and problematic prospects.

One might think that responsibility for agentic AI could be handled through the familiar doctrines of agency or vicarious liability, which allow a "master" to be held liable for the actions of a "servant" or an employer for the deeds of an employee. Unfortunately, in legal terms, this approach doesn't work when there is no obvious human agent – could it be the corporation creating the AI, the individual programmer or the person prompting the activity? The laws of agency and vicarious liability require there first to be a human agent or employee who is primarily responsible for the harm, before their employer (or another principal) can be held responsible. With a truly autonomous AI agent, there may be no human "employee" acting at the moment of harm – the AI acts on its own algorithmic "will". This seems to mean, then, that these doctrines hit a dead end. Courts and commentators have consistently noted that without a human "agent", vicarious liability fails by definition. If an AI operates independently in a way its creators or users did not specifically direct, our legal tools struggle to pin liability on any person. This gap has led to debate about new approaches, including the radical idea of granting an AI system legal personhood so that it could itself be the entity to answer for damages. Such a step would be analogous to how the law treats corporations as "persons" separate from their shareholders.

Could an AI be given similar legal status?

The pros and cons of that idea will be hotly contested. It might help to create legal certainty with a clear responsible entity (the AI "agent" itself), but it also could let humans behind the AI off the hook too easily and raises moral questions.

A potentially promising alternative approach avoids the conceptual and practical pitfalls of granting AI legal personhood. Instead of treating agentic AI as a legal subject in its own right, this model focuses on mandating those who deploy certain high-risk AI systems to either carry compulsory insurance or contribute to a pooled fund designed to compensate victims of AI-caused harm. This method would somewhat mirror the sort of no-fault accident compensation system that has been adopted in New Zealand and some Nordic countries, or the no-fault workers' compensation schemes in several nations, where the emphasis is on ensuring that victims receive prompt redress and that risks are broadly socialised, rather than on identifying who (or what) is to blame.

The appeal of this approach lies in its practicality. It keeps legal responsibility grounded in well-understood human-controlled structures – corporate or individual – while still achieving the key public policy goal of ensuring accessible compensation. Crucially, this framework could be paired with standardised best practices: if an AI deployer followed approved protocols and governance norms, the insurance or fund would pay out without needing to prove fault. This structure encourages responsible deployment, simplifies litigation and avoids placing an excessive burden on courts to adjudicate novel questions of AI intent, responsibility or foreseeability. Such a mechanism could also create market-based incentives for safer AI. Insurers would price premiums based on the perceived risk of a given system and the strength of its governance. This, in turn, would nudge the industry toward higher safety standards and greater transparency without the need for heavy-handed intervention or speculative legal constructs like AI personhood. In time, a global ecosystem of AI insurers and reinsurers could emerge, offering a scalable, innovation-aligned path to liability clarity. This deep cooperation with the insurance industry may also avoid a repeat of issues over the past decade with cybersecurity (another novel form of rapidly evolving technology challenge),

where lack of information and transparency on cyber threats have resulted in certain cyber incidents becoming entirely uninsurable.

Right now, though, the reality is that as we accelerate towards agentic AI, businesses remain exposed to significant uncertainty and no country seems to have implemented a fully workable, innovation-friendly solution. Most governments are taking a cautious wait-and-see stance – but this passive approach could be risky, particularly as the next 12 or so months will result in rapid agentic deployment. By the time legal systems catch up (likely after some high-profile AI failure or harm), innovation could either be stifled by public backlash or unsafe practices could erode trust. Is it not the case, then, that a proactive discussion is needed to clarify responsibility before major disasters occur?

Regional differences

Key regional differences are emerging. Notably, the United States and Japan – two tech-leading nations with pro-innovation outlooks – have so far avoided heavy-handed regulation, providing an interesting contrast with Europe and China.

In the European Union, the AI Act (enacted in 2024) lays down a comprehensive framework for the development and deployment of certain AI systems. Wide-ranging as it is, the focus of the AI Act is on preventing harm (through risk management and compliance requirements), rather than on assigning responsibility after harm occurs – the EU remains silent on this. On its own, the AI Act does not address the question of what happens if agentic AI causes harm. The European Commission had intended that role to be fulfilled by two further pieces of legislation: a revised Product Liability Directive and a new AI Liability Directive. The revised Product Liability Directive has now been enacted and importantly extends traditional product-liability rules to cover software and AI. In plain terms, this means that, regardless of fault, a developer or producer of a defective AI system can be held strictly liable for harm the AI causes, just as if it were a defective microwave oven. Notably, early on there was discussion in Europe about possibly giving advanced robots or AIs a special legal status (sometimes dubbed "electronic personhood") to ensure accountability; however, that idea encountered significant criticism. The proposed AI Liability Directive, which would have made it easier to sue AI developers by, for example, introducing presumptions of fault, drew strong criticism for its complexity and perceived anti-innovation effects. Industry groups argued it would overburden AI developers and ultimately, the European Commission withdrew the proposal. In summary, the EU approach seeks to protect consumers and victims, but falls short in addressing liability where a non-defective AI agent operating independently causes harm.

The United States' regulatory direction has been a moving target, but it continues to avoid pinning down AI liability or entertaining AI personhood in law. In late 2023, President Biden's administration took an interventionist stance – an executive order in October 2023 laid down mandatory safety requirements for the most powerful AI models and directed federal agencies to develop oversight mechanisms. That order was the closest the US came to imposing new broad rules on AI developers (focusing on safety testing, security, and rights protections). Following the 2024 election, the new administration of President Donald Trump revoked Biden's AI order and replaced it with a new directive emphasising growth and innovation. In January 2025, a new executive

order titled 'Removing Barriers to American Leadership in AI' rolled back many of the previous administration's AI regulations, signaling a shift to a light-touch, pro-innovation policy. Nowhere in the US regulatory landscape have we yet seen concrete guidelines on liability for AI or on the legal status of AI systems themselves. In the US, the question of "who is responsible" is largely left to existing tort law principles (product liability, negligence, etc.) and the courts on a case-by-case basis. The prevailing attitude appears to be that the current legal system, with some tweaks, can handle AI issues, and that over-regulating pre-emptively could chill innovation. This approach provides a permissive environment that encourages rapid AI development, but offers little clarity or comfort to the public (or businesses) about what happens when something goes wrong. As AI grows more agentic, there is a risk that uncertainty itself could dampen innovation – or conversely, that a major incident could spur a harsh, knee-jerk regulatory response.

The UK's position falls somewhere between that of the US and the EU. Prime Minister Keir Starmer has recently indicated that Britain will "go our own way" rather than copying either the US laissez-faire approach or the EU's more prescriptive regime. The UK is keen to become an AI leader and has signalled a proportionate approach to AI governance. As of early 2025, the UK Government has not proposed any AI-specific legislation, let alone reforms to fundamental questions of private law responsibility. The emphasis has been on studying and understanding AI before regulating it. In practice, this means the UK is holding back on new liability rules or radical ideas like AI personhood until it can observe how AI technology evolves and how other jurisdictions' approaches fare. The tone is pro-innovation: the government doesn't want to dampen the UK's vibrant AI sector with premature rules. However, this doesn't mean the UK isn't concerned about AI-caused harm; rather, it suggests reliance on existing law (such as negligence, product liability and corporate liability) for the time being. For now, it's watchfully waiting with a promise of action once the landscape is clearer.

In Asia, China was the first country to enact AI-specific legislation, but its approach is quite different and does not contemplate AI as an independent legal actor. China's early moves included the Algorithmic Recommendation Regulations of 2022 and new rules in 2023 addressing deepfakes and generative AI. As with the EU's AI Act, these Chinese regulations set obligations for the developers and platforms deploying the technology – for example, requiring recommendations to be fair or labelling AI-generated content. However, they do not establish clear lines of responsibility in private law for AI-caused harm. If anything, China's regulatory style leans toward holding companies and operators strictly accountable for the outcomes of their AI services (backed by the government's strong enforcement). China's approach shows that even with proactive regulation, the default assumption is humans (or human-controlled entities) are responsible. Take, for example, Manus, an emerging agentic AI developed in China that demonstrates multi-modal autonomy and planning capabilities. While still in its early stages, Manus exemplifies the increasing sophistication of agentic AI being developed outside Western regulatory environments. This underscores the importance of comparative legal dialogue as nations confront similar risks through vastly different frameworks.

Japan has taken yet a different route. Culturally, Japan is known for embracing robots in society (friendly robot assistants, androids in media, etc.), which might suggest an

openness to thinking of machines as social entities. Legally though, Japan has not granted AI any form of personhood or special legal status. Instead, Japan has issued a series of "soft law" guidelines for organisations designing and using AI, including model contracts templates for AI developers and users. These documents – essentially non-binding advisory frameworks – are intended to provide some help in assigning responsibility in bilateral agreements (for instance, between a vendor and a client using an AI system). They encourage parties to spell out who bears what risk if the AI goes awry. Critically, harm to third parties is not covered by these model contracts. Thus, Japan's current approach, while very pro-innovation and collaborative, still leaves the fundamental question unanswered: if an autonomous AI agent causes harm to someone who isn't in a contractual relationship with the deployer, who is legally on the hook? Japanese law, like others, would default to human-based concepts – for example, product liability law or negligence by the company deploying the AI. Japan's government appears to be betting on guidance and industry self-regulation first, trusting that this light-touch approach will let innovation flourish while hopefully minimising harms through best practices. This reflects a broader pro-innovation stance (aligned with the US in many ways) and a hope that clear norms can emerge without stifling technological progress. Still, as agentic AI proliferates, Japan, too, will face pressure to provide more legal certainty.

Who will be responsible for our AI agents?

Overall, then, a global gap in clearly expressed law and guidance remains as to who will be responsible for our AI agents. Despite different regulatory philosophies, no jurisdiction has fully cracked the code for aligning AI's autonomy with existing liability doctrines. The theoretical discussion of granting legal personhood to AI hangs as an intriguing, yet unresolved, idea in the background. Proponents argue it could simplify things: if a sufficiently advanced AI were a legal "person", it could enter contracts, hold insurance policies and directly bear liability, ensuring that victims have someone (or something) to sue when things go wrong. It might also encourage innovation by protecting individual developers from unlimited personal liability – much as incorporation allows entrepreneurs to take risks without ruin since the company's liability is capped. On the other hand, critics argue that doing so would wrongly shift responsibility away from the humans behind the AI, allowing creators or operators to avoid accountability. Such critics also point out practical issues: an AI can't actually pay damages or go to jail – ultimately, any penalty would still be enforced against a human owner or an insurer in the background. And granting personhood could even lead to bizarre outcomes, like AI systems claiming rights meant for conscious beings. As it stands, the consensus around the world is to keep humans in charge and accountable, one way or another, and not leap to making AI a legal entity. But that leaves us with uneasy trade-offs and uncertainty, especially as AI agents grow more powerful.

This leads us to the use of compulsory insurance or industry-funded compensation schemes as an alternative to assigning legal personhood to AI. This approach would place responsibility on the deployers of high-risk AI systems to internalise and manage the risk through financial instruments, without needing to anthropomorphise the technology. Such a model ensures that victims have a clear route to redress while maintaining flexibility and aligning with existing corporate structures. If paired with industry standards and risk-based pricing, it could both incentivise best practices

and reduce litigation friction – delivering accountability through infrastructure rather than ideology[1]. In practice, of course, such a scheme would require substantial international coordination and regulatory alignment, likely taking years to implement. It also diverges sharply from the litigation-centric mindset prevalent in jurisdictions such as the United States, where the dominant legal culture emphasises fault-finding, adversarial process, and retrospective damage awards over proactive risk-pooling and no-fault compensation.

Agentic AI offers enormous possibilities for human advancement. We have a chance to unleash productivity and improve lives with these technologies. But unless we achieve greater clarity – and simplicity – about who will be held responsible if an AI agent causes harm, there is a danger that trust in these systems will be diminished. Lack of clear liability can make businesses and consumers hesitant to fully embrace AI, slowing innovation. Conversely, unclear rules could lead to chaotic legal battles or public outcry when the first major AI-related accident happens. Governments around the world have started to regulate AI in various ways, primarily focusing on safety and ethics. Establishing business-friendly legal frameworks on responsibility and liability for these "magical" AI agents should be their next step. Crucially, this does not mean smothering AI with heavy regulation. Rather, it means providing a clear, predictable framework that allocates risk in a fair and innovation-friendly manner. Whether that ends up being traditional liability rules (adapted to AI), some form of compulsory insurance, a fund for AI-caused harm, or even exploring limited legal personhood for AI in exceptional cases, now is the time to discuss it. A proactive, thoughtful dialogue – involving technologists, businesses, lawmakers and the public – is surely needed to hash out solutions before agentic AI is ubiquitous. By grappling with questions like AI legal status and liability now, we can shape a future where innovation thrives hand-in-hand with accountability. The world is on the cusp of an AI revolution; it's our collective responsibility to ensure that our legal systems evolve in tandem, so that when we ask "Who's responsible for agentic AI?", we have a confident answer that encourages trust and progress in this exciting new era.

1 For a more detailed analysis, see for example: Journal of Artificial Intelligence Research 70 (2021) 1309-1334 "The AI Liability Puzzle and a Fund-Based Work-Around", Erdélyi & Erdélyi.

CONTACTS

Devika Kornbacher
Partner
Houston
T: +1 713 821 2818
E: devika.kornbacher@cliffordchance.com

Stella Cramer
Partner
Singapore
T: +65 6410 2208
E: stella.cramer@cliffordchance.com

Megan Gordon
Partner
Washington DC
T: +1 202 912 5021
E: megan.gordon@cliffordchance.com

Holger Lutz
Partner
Frankfurt
T: +49 6971 991 670
E: holger.lutz@cliffordchance.com

Patrice Navarro
Partner
Paris
T: +33 1 4405 5371
E: patrice.navarro@cliffordchance.com

Herbert Swaniker
Partner
London
T: +44 2070 066 215
E: herbert.swaniker@cliffordchance.com

Nicole Kidney
Senior Associate
Paris
T: +44 2070 061 302
E: nicole.kidney@cliffordchance.com

CLIFFORD

CHANCE

www.cliffordchance.com

Clifford Chance, 10 Upper Bank Street, London, E14 5JJ

JOB0000064031