

C L I F F O R D
C H A N C E

 SoftBank
Group

**AI SAFETY AND ACTION:
HOW GOVERNANCE BUILDS TRUST**

AUTHORS



Brendan Kelleher
Chief Compliance Officer
SoftBank Group International

Brendan Kelleher

Brendan Kelleher is a Partner and Chief Compliance Officer at SoftBank Group International in Menlo Park, California. He leads SoftBank Group's governance and ethics agenda outside of Japan, including on matters of responsible AI. He is currently leading SoftBank's development of responsible AI governance, focusing on AI literacy, standards for AI use and deployment, and AI portfolio risk. He works closely with SoftBank companies in the fields of AI, robotics and semiconductors to establish best-in-class internal controls and AI governance standards. He is also the chief compliance officer of Stack AV, SoftBank's autonomous trucking company. Mr. Kelleher graduated from Brown University and Northwestern University School of Law.



Nicole Kidney
Senior Associate
Clifford Chance

Nicole Kidney

Nicole Kidney is a senior technology lawyer at Clifford Chance in London. Nicole specialises in AI, data law and international privacy matters. Her expertise spans a wide range of advisory and transactional matters, working with clients internationally to address strategic issues related to AI policy and investments, emerging technologies, privacy and online safety. Nicole studied at Trinity College Dublin and is a qualified solicitor in England and Wales, as well as being admitted to the Roll of Solicitors in the Republic of Ireland. She has gained in-house experience through a client secondment, and has worked in Clifford Chance's Singapore office.

AI SAFETY AND ACTION: HOW GOVERNANCE BUILDS TRUST

The current AI wave holds remarkable opportunities for innovation, societal advancement, and economic growth. To realise this promise, companies are securing massive amounts of compute, energy, data and other resources necessary for AI advancement. But the commercial and societal success of AI hinges largely on another resource not often discussed – trust.

Trust encompasses multiple dimensions, including public confidence, system reliability and transparency. But how can we cultivate this critical trust? The answer may lie in effective and targeted governance. If AI is not secure, reliable, and transparent, the risk of harm increases, users lose trust in AI tools, and adoption of the technology may stall.

By looking at key themes such as transparency and bias, and examining their evolution in other sectors, we can better understand the risks that need to be mitigated and, importantly, the benefits – to both safety and innovation – of implementing an effective governance program.

Trust in AI: Insights from the Paris AI Action Summit

The Paris AI Action Summit was the third meeting of heads of state and tech leaders on the topic of AI. While earlier events – Bletchley Park, UK in 2023 and Seoul in 2024 – focused on AI safety, the Paris Summit broadened the conversation by emphasising AI's role in driving innovation. Despite this shift, a central theme persisted in Paris – “trust in AI” is the foundation to both safety and progress.

Throughout the AI Action Summit and fringe events, participants stressed the importance of building societal confidence in AI as a means of accelerating innovation. Many proposed achieving this through a common goal – shared among developers, regulators and the public – of establishing AI for public good. Attendees identified the need for public transparency and accountability to improve understanding of, and trust in, AI. Some underscored the benefits of a common language on AI governance and safety to better align the expectations of all stakeholders.

Lessons from Other Sectors

Which brings us back to the crucial question – how can this essential trust be earned? History offers examples where governance has been used to foster trust in sectors initially faced with scepticism and fear. These governance frameworks not only protected the public but also unleashed a wave of innovation, transforming the once-hesitant products into integral parts of daily life.

The pharmaceutical industry has long been held to strict safety protocols. Ethics committees ensure that pharmaceutical products are not released without being

assessed by an impartial, diverse group charged with protecting the rights and safety of individuals. Similarly, rigorous clinical testing helps to ensure products come to market safely and post-market surveillance validates product safety once in widespread use. These governance mechanisms have helped to promote industry transparency, accountability and build trust.

The automotive industry's type approval regulations are another successful model for applying technical and safety standards to ensure reliability. Customers are reassured knowing that vehicles have undergone extensive testing to prove their safety, even when new advanced features are introduced.

The aviation industry has also earned trust by embedding transparency and accountability into its governance framework. Incident reporting and black boxes play a big role in this. When an incident occurs, airlines are required to document and investigate it. These reports are shared across the industry and reported publicly, creating a collective learning system and common language. Every airplane is equipped with a black box that records flight data to further help explain to regulators, airlines, manufacturers and the public what went wrong and how to prevent recurrence.

Each of these industries demonstrates that effective governance does not stifle innovation – it enables it. A similar approach in AI, combining proactive safety measures, transparent reporting and common industry standards could help bridge the trust gap.

Challenges to Building Trust in AI

While the benefits of building trust are well-recognised, the AI sector faces unique challenges. These need to be navigated carefully to ensure that governance is straightforward, effective and continues to foster innovation. Companies and regulators are grappling with these issues today, another common theme from the AI Action Summit.

Transparency and Confidentiality

Transparency is key to earning trust; but it comes with its own obstacles. To improve public confidence in AI, users and regulators increasingly expect developers to provide them with the necessary information to understand AI systems and their outputs. However, the inherent complexities of frontier AI models make them difficult to explain in a user-friendly way.

AI companies also have disincentives to revealing the inner workings of their products. For example, sharing the technical workings behind AI systems could expose vulnerabilities, allowing for malicious attacks, or facilitating reverse engineering.

Bias and Accuracy

The public has heard much about how AI systems can perpetuate human bias by reflecting prejudices inherent in their training data. Instances of AI hallucinations, and the consequences of relying on such inaccuracies, have also been widely reported in the press. Risks of bias and inaccuracies in AI outputs undermine public confidence in AI.

AI companies have made strides to address hallucinations, using strategies such as chain of thought reasoning and consistency checks to, in essence, double-check the work of models. But more public education is needed on the utility and accuracy of AI to counter its reputation for mistakes. This is an area where a common language may emerge.

While hallucinations are a technical issue addressed with technical solutions, bias is an issue systemic to society. Addressing it has proved challenging because bias is inherent in the data on which models are trained. Our concept of bias is also a moving target. Nevertheless, technical and procedural methods to mitigate bias can be deployed at all stages of AI development. Success on this front is essential to advance trust in AI.

User Empowerment

The enigmatic nature of AI can cause scepticism and misconceptions. This scepticism can often hinder adoption of AI products in everyday life. In addition, AI-related incidents can quickly exacerbate these fears and erode trust.

The more users understand why AI acts in a certain way and can rely on AI developers to investigate and remediate incidents, the more they will trust it. Companies should think carefully about how they can educate users and build AI literacy into their communication strategies. Similarly, enterprises can train their employees on AI's risks and opportunities, helping to develop a shared understanding and common AI language. Publicly prioritising safety through rigorous testing, incident reporting and clear accountability frameworks can further empower individuals to make informed decisions and use AI responsibly.

What's Next?

Building trust is not just a regulatory checkbox exercise but a strategic goal that benefits all stakeholders.

A “trust infrastructure”, established through effective and targeted governance, is critical and helps to address three challenges: first, managing the risk of harm as companies build secure products; second, increasing user adoption through promoting trust, safety and well-being; and third, clarifying developer and user expectations to create space for innovation. Governance then becomes an enabler of sustainable growth, aligning the objectives and strategic goals of all stakeholders in the AI ecosystem.

The result? A future where AI is not only powerful and transformative but also trusted and embraced by society.

CONTACTS



Nicole Kidney
Senior Associate
London
T: +44 207006 1302
E: nicole.kidney@cliffordchance.com



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



Phillip Souta
Global Director of
Tech Policy
London
T: +44 207006 1097
E: phillip.souta@cliffordchance.com

C L I F F O R D

C H A N C E

This publication does not necessarily deal with every important topic or cover every aspect of the topics with which it deals. It is not designed to provide legal or other advice.

www.cliffordchance.com

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

© Clifford Chance 2025

Clifford Chance LLP is a limited liability partnership registered in England and Wales under number OC323571

Registered office: 10 Upper Bank Street, London, E14 5JJ

We use the word 'partner' to refer to a member of Clifford Chance LLP, or an employee or consultant with equivalent standing and qualifications

If you do not wish to receive further information from Clifford Chance about events or legal developments which we believe may be of interest to you, please either send an email to nomorecontact@cliffordchance.com or by post at Clifford Chance LLP, 10 Upper Bank Street, Canary Wharf, London E14 5JJ

Abu Dhabi • Amsterdam • Barcelona • Beijing • Brussels • Bucharest • Casablanca • Delhi • Dubai • Düsseldorf • Frankfurt • Hong Kong • Houston • Istanbul • London • Luxembourg • Madrid • Milan • Munich • Newcastle • New York • Paris • Perth • Prague • Riyadh* • Rome • São Paulo • Shanghai • Singapore • Sydney • Tokyo • Warsaw • Washington, D.C.

*AS&H Clifford Chance, a joint venture entered into by Clifford Chance LLP.

Clifford Chance has a best friends relationship with Redcliffe Partners in Ukraine.