# UK launches global coalition to tackle AI alignment risks



The UK has launched a major international coalition aimed at governing artificial intelligence more responsibly, marking a significant step in global AI safety. Spearheaded by the AI Security Institute, the £15 million initiative—known as the Alignment Project—brings together governments, academic institutions, civil society organisations and tech giants including Amazon Web Services, Anthropic and Cohere.

The coalition’s central mission is to tackle the challenge of AI alignment: ensuring advanced systems behave predictably, ethically and in accordance with human values. Its formation follows growing concern over the “Wild West” nature of AI development, as increasingly powerful models influence decisions in critical areas such as healthcare, hiring and finance.

The UK’s leadership has been underlined by the signing of its first legally binding international treaty on AI risk, which calls for cooperation to prevent misuse while safeguarding democracy, human rights and the rule of law. The coalition includes the Canadian AI Safety Institute and UK Research and Innovation, reinforcing its global scope. It aims to develop shared safety standards and foster research into mitigating AI unpredictability—such as the generation of false data and biased decision-making.

A 2023 class-action lawsuit against US-based HR software company Workday spotlighted the legal risks of unchecked AI, with claims that automated hiring tools had discriminated against underrepresented groups. Cases like this underscore the need for transparent and accountable systems.

The AI Security Institute is also working with the Home Office and industry partners to explore how AI can support economic growth while protecting national security. The institute’s research collaborations position the UK as a global hub for safe and trustworthy AI development.

For businesses, the message is urgent. Risk and compliance teams are being advised to act now. Best practices include AI impact assessments, bias and explainability audits, ethical training programmes, and strong governance frameworks for procurement and monitoring. These are increasingly expected by regulators, particularly in high-risk sectors such as law, finance and healthcare.

The coalition signals growing international momentum towards safe and ethical AI innovation. By leading this charge, the UK is shaping a future where AI is developed and deployed with the safeguards necessary to protect society while unlocking economic value.

In a rapidly advancing field, alignment remains one of AI’s most complex scientific and policy challenges. But through collaboration and strong governance, the UK and its partners are helping to ensure the next wave of AI progress is safe, accountable and aligned with public interest.

Created by [Amplify](https://www.hbmadvisory.com/amplify): AI-augmented, human-curated content.

## Bibliography

1. <https://vinciworks.com/blog/global-coalition-formed-to-tame-wild-west-ai/> - Please view link - unable to able to access data
2. <https://www.gov.uk/government/news/ai-security-institute-launches-international-coalition-to-safeguard-ai-development> - The UK’s AI Security Institute has launched an international coalition to safeguard AI development. This initiative brings together the Canadian AI Safety Institute, Amazon Web Services, Anthropic, and civil society organisations. The project focuses on AI behaviour and control, aiming to ensure AI systems act predictably and align with human values. Backed by over £15 million, the fund supports the UK’s position as a leader in AI safety and expands global efforts to tackle AI alignment challenges.
3. <https://www.gov.uk/government/news/uk-signs-first-international-treaty-addressing-risks-of-artificial-intelligence> - The UK has signed its first legally binding international treaty addressing the risks of artificial intelligence. The treaty commits parties to collective action to manage AI products and protect the public from potential misuse. It includes safeguards against risks to human rights, democracy, and the rule of law, ensuring that AI development is monitored and managed within strict parameters. The UK continues to play a key role as an international leader in safe, secure, and trustworthy AI.
4. <https://www.gov.uk/government/news/prime-minister-launches-new-ai-safety-institute> - The UK Prime Minister has launched a new AI Safety Institute, a state-backed organisation dedicated to advancing AI safety. The institute aims to evaluate and ensure the safety of advanced AI models, bringing together international partners, tech companies, and civil society. It focuses on AI behaviour and control, ensuring AI systems act predictably and align with human values. The initiative is backed by over £15 million and supports the UK’s position as a leader in AI safety.
5. <https://www.gov.uk/government/news/tackling-ai-security-risks-to-unleash-growth-and-deliver-plan-for-change> - The UK’s AI Safety Institute has been rebranded as the AI Security Institute, strengthening protections against AI risks to national security and crime. The institute has partnered with the Home Office to research a range of crime and security issues that could harm UK citizens. A new agreement with AI giant Anthropic focuses on AI opportunities to help grow the economy as part of the government’s Plan for Change. The initiative aims to balance AI development with safety and security considerations.
6. <https://www.gov.uk/government/news/uk-signs-first-international-treaty-addressing-risks-of-artificial-intelligence> - The UK has signed its first legally binding international treaty addressing the risks of artificial intelligence. The treaty commits parties to collective action to manage AI products and protect the public from potential misuse. It includes safeguards against risks to human rights, democracy, and the rule of law, ensuring that AI development is monitored and managed within strict parameters. The UK continues to play a key role as an international leader in safe, secure, and trustworthy AI.
7. <https://www.gov.uk/government/news/ai-security-institute-launches-international-coalition-to-safeguard-ai-development> - The UK’s AI Security Institute has launched an international coalition to safeguard AI development. This initiative brings together the Canadian AI Safety Institute, Amazon Web Services, Anthropic, and civil society organisations. The project focuses on AI behaviour and control, aiming to ensure AI systems act predictably and align with human values. Backed by over £15 million, the fund supports the UK’s position as a leader in AI safety and expands global efforts to tackle AI alignment challenges.