

Who's Who

in

AI Safety & Governance

August 2026

SOURCE: Microsoft Co-pilot

Frontier-Model Technical Safety Leaders

These individuals design or oversee *model-level safety systems* and alignment research.

Leader	Organization	Focus
Dario Amodei	Anthropic – Co-founder & CEO	• Frontier-model safety frameworks (RSP, ASL-3/4)
Demis Hassabis	Google DeepMind - Chairman	• Frontier Safety Framework (FSF), global governance
Ilya Sutskever	Safe SuperIntelligence – Co-founder & CEO	• Alignment research, Superalignment team
Jan Leike	Anthropic – Head of Alignment Science	• Scalable oversight, alignment methods
Stuart Russell	UC Berkeley – Professor	• Provably beneficial AI, alignment theory

SOURCE: Microsoft Co-pilot

AI Governance Technical Leaders

These leaders build *formal AI safety frameworks, capability indices, and operating points* bridging technical and policy domains.

Leader	Organization	Key Contributions
Chai Toh	SAFE AI Foundation – CEO & Chairman	• AGI/ASI Taxonomy, AGI Capability Index, SAIOP
Brad Smith	Microsoft – Vice Chair & President	• Global AI policy, safety standards
Bruce Reed	White House – Deputy Chief of Staff	• U.S. AI Executive Order, national safety strategy
NIST AI RMF Team	U.S. NIST	• AI Risk Management Framework
UK Frontier Safety Framework Authors	UK DSIT	• Frontier Safety Framework (FSF)

SOURCE: Microsoft Co-pilot

Industry Governance Leaders

These executives integrate safety governance into corporate operations and standards.

Leader	Organization	Focus
Kent Walker	Google – Chief Legal Officer & President of Global Affairs	• Responsible AI governance • Safety assessment tools • Global AI policy • AI law and governance frameworks
Alexandr Wang	Scale AI – Founder & CEO	
Nick Clegg	Meta – President of Global Affairs	
Deirdre Mulligan	UC Berkeley – Professor	

SOURCE: Microsoft Co-pilot

Policy & Advocacy Leaders

These figures shape international AI-safety norms and public governance.

Leader	Organization	Focus
Dan Hendrycks	Center for AI Safety	• Risk taxonomy, safety categories
Allan Dafoe	DeepMind / Oxford	• Institutional governance research
OECD / UNESCO Teams	International Bodies	• Global AI standards and principles

SOURCE: Microsoft Co-pilot