



POLICY BRIEF

The Future of the AI Summit Series: One Year Later

OXFORD AI DIPLOMACY LAB

A 2026 UPDATE

Lucia Velasco, Reed Schuler, Robert Trager,
Yoshua Bengio, Julia Morse, Benjamin Prud'homme,
Joslyn Barnhart, Jonathan Barry, Roxana Radu,
Stephen Casper, Ranjit Lall, Duncan Cass-Beggs,
Charles Martinet, Daniel Zhang

The Future of the AI Summit Series: One Year Later

A 2026 Update

**Lucia Velasco¹, Reed Schuler¹,
Robert Trager¹, Yoshua Bengio², Julia Morse^{3,1}, Benjamin Prud'homme²,
Joslyn Barnhart^{4,3}, Jonathan Barry², Roxana Radu⁵, Stephen Casper⁶, Ranjit Lall¹,
Duncan Cass-Beggs⁷, Charles Martinet¹, and Daniel Zhang⁸**

¹Oxford Martin AI Governance Initiative, University of Oxford

²Mila – Quebec AI Institute

³University of California, Santa Barbara

⁴Google DeepMind

⁵Blavatnik School of Government, University of Oxford

⁶Harvard Kennedy School

⁷Global AI Risks Initiative, Centre for International Governance Innovation

⁸Independent

Corresponding author: lucia.velasco.j@gmail.com

Full citation:

Velasco, Lucia, Reed Schuler, Robert Trager, Yoshua Bengio, Julia Morse, Benjamin Prud'homme, Joslyn Barnhart, Jonathan Barry, Roxana Radu, Stephen Casper, Ranjit Lall, Duncan Cass-Beggs, Charles Martinet, and Daniel Zhang. 2026. "The Future of the AI Summit Series: One Year Later." Policy Brief. Oxford Martin School AI Governance Initiative.

Juan Manuel Santos

Former President of Colombia,
Nobel Peace Laureate and
Chair of The Elders



Artificial intelligence is developing at an extraordinary pace. This technological transformation holds great promise across domains in which the world urgently needs progress. At the same time, without guardrails, AI threatens to destroy livelihoods, deepen inequalities, entrench authoritarianism, and exacerbate conflicts.

These risks are not fundamentally technical challenges. They are about governance: who decides how our societies develop. Their solutions require political leadership, both nationally and internationally. That leadership must keep pace with technological change. This is not impossible, as we have been led to believe. Governments can and must ensure that as the power of artificial intelligence grows, we use it for shared global purposes, and do not allow its worst risks to materialise.

One important avenue for cooperation on addressing the governance gap has been the AI Summit Series, launched at Bletchley Park in 2023. These summits bring together world leaders alongside the companies developing AI technologies, with a focus at their founding on the governance and safety of advanced AI. Providing a nimble forum for cooperation, they can serve as an important complement to the United Nations and other multilateral efforts. Yet over time, the summits have shifted away from urgent governance challenges.

I welcome this report's valuable contribution to the debate on how to restore the summits' focus on governance and strengthen their outcomes. We need continuity from one summit to the next, a dedicated track for governance, a standing capability to support and connect successive hosts, and mechanisms for accountability and implementation. These ingredients are crucial to the success of the summits, and to developing effective global AI governance.

Multilateralism is currently going through its worst crisis since the end of World War II. That is precisely why this call to strengthen existing dialogues is timely. Periods of geopolitical uncertainty demand investment in new ideas and preparation for the future, so that we are ready when windows of political possibility open. The timing of those windows cannot be predicted, and they may not remain open for long.

The Elders encourage far-sighted political leadership on the importance of effective guardrails for AI (despite the strong voices arguing against this). We want to strengthen the AI Summit Series, as a central component of more inclusive multilateral management of the risks and benefits of AI.

We urge political leaders everywhere to find the courage to ensure that AI serves humanity, and contributes to a safer, more peaceful, and prosperous future for all.

TABLE OF CONTENTS

The Future of the AI Summit Series: One Year Later

Executive Summary	05
An Evolving Landscape	06
The Stakes: Why Coordination Still Matters	14
The Opportunity	17
Structural Options Worth Considering	20

Executive Summary

One year and two more AI summits have passed since we published our January 2025 memo, *The Future of the AI Summit Series*.¹ In the interim, the 193 Member States of the United Nations² have approved the creation of a Global Dialogue on AI Governance.

Over the last year, AI capabilities have continued their rapid development. Frontier models have grown more capable and deployment is accelerating across applications, ranging from basic scientific research to military systems. There is active debate about the future trajectory of these models, and the public and policymakers are urgently grappling with the implications and the responses they demand across society.

The AI Summit Series has continued to grow in participation and visibility. The Paris AI Action Summit (February 2025) drew participation from over 100 countries, a significant expansion from Bletchley and Seoul. The Delhi Summit held in February 2026 expanded this further, drawing approximately 600,000 in-person attendees from all over the world, and producing the New Delhi Declaration endorsed by 91 nations. The six-day format incorporated a public-facing AI Impact Expo, a CEO Roundtable, a Leaders' Plenary, and seven thematic working groups. Switzerland has been confirmed as the 2027 host in Geneva, with the UAE taking the lead in 2028, giving the series its first visible succession chain. By measures of attention and attendance, the series has succeeded. Yet the series has not developed structural maturity to match these stakes. Relatively ad hoc hosting arrangements remain in place. The secretariat function remains informal and host-dependent. Agenda-setting continues to start anew with each host's priorities rather than building systematically on previous work. The series has not translated its growing visibility into proportionate policy relevance.

Despite headwinds for global cooperation, important work remains on the governance of advanced AI, and summit hosts have a clear opportunity to advance reforms that sharpen this work. The original memo offered recommendations across five core design elements: hosting arrangements, secretariat format, participant selection, agenda-setting, and summit frequency.

We proposed a two-track participation framework (a Governance Track and an Action Track on public-interest uses of AI), a hybrid secretariat model, a coordinated succession approach to agenda-setting, and interim technical meetings to complement annual summits. Of these, the hosting succession has seen organic but informal progress, with Switzerland (2027) and the UAE (2028) announced as future hosts during the Delhi Summit itself, improving the ability of hosts to plan effectively. The others remain unimplemented in any structural sense. Delhi's seven-track, impact-focused design illustrated both the appeal of broadening the series and the cost: focused policy work on advanced AI governance was dispersed across a wider agenda. The opportunity now is clear: by adopting the two-track participation framework in full, the series can build on what it has already achieved while preserving a dedicated space for focused work on governing the frontier.

¹ Velasco, Trager, et al. (2025). "The Future of the AI Summit Series," Oxford Martin School AI Governance Initiative Policy Memo.

² A/RES/79/325, adopted by consensus on 27 August 2025, establishing the Global Dialogue on AI Governance as an inclusive platform for all countries to collectively address AI governance.

An Evolving Landscape

01

I. An Evolving Landscape

From Governance to Diffusion: Shifting Focus

The Paris AI Action Summit ushered in a significant shift in orientation. The Bletchley and Seoul summits had a tight focus on the governance implications and opportunities of advanced AI, producing concrete outputs like the Bletchley Declaration, the Frontier AI Safety Commitments, and the International Scientific Report on the Safety of Advanced AI. The Bletchley Declaration was seen as laying the groundwork for future cooperation to increase transparency, monitoring, and mitigation efforts across a variety of risks. Paris adopted a much broader frame, introducing multiple thematic tracks spanning innovation, culture, the future of work, and public interest AI. The French government positioned AI as a domain of opportunity as much as or more than risk, and significantly expanded summit participation.

The Paris Summit also marked an inflection point in global politics. The United States and United Kingdom declined to sign the AI Action Summit Declaration despite its broadly framed content, citing concerns about global governance and the risks of excessive regulation. This departure from the consensus of Bletchley and Seoul signalled that the original animating spirit, a shared commitment among leading AI powers to collective frameworks to manage challenges together, may no longer hold. Since Paris, many have seen a wider fraying of the international order, including recent sharp divergences at the Annual Meeting of the World Economic Forum in Davos.

Observers have characterized the Paris Summit shift in varying ways. Some viewed these changes as a necessary evolution, a widening of the aperture to ensure the gathering does not remain narrowly focused on risks. Others viewed it as a dilution, arguing that the expansion of focus came at the cost of focused policy engagement. What seems clear is that focused policy-oriented work on the governance of advanced AI has receded from the summit series' agenda. The Delhi Summit confirmed this trajectory. India organized the summit around working groups under three broad pillars (People, Planet, Progress), with 'Safe and Trusted AI' as one strand among seven. The Frontier AI Impact Commitments, voluntary pledges by 13 companies on usage transparency and multilingual evaluation, were the closest output to the governance-track work of Bletchley and Seoul, but addressed deployment and inclusion rather than governance and regulatory questions, such as the safety frameworks, incident reporting, and risk thresholds that characterized earlier summits. The shift is now a pattern across consecutive summits.

Meanwhile, there have been other developments in international engagement. The United Nations launched its Global Dialogue on AI Governance, providing an inclusive forum for wide-ranging discussions between national governments. Other forms of coordination on technical AI risks have grown. The second International AI Safety Report, again published to coincide with the summit, offered a comprehensive review of the latest scientific research on the capabilities and risks of general-purpose AI systems, and further consolidated the summit as its venue. Led by Turing Award

winner Yoshua Bengio and authored by over 100 AI experts, the report is backed by over 30 countries and international organizations, representing the largest global collaboration on AI safety to date.³ These are valuable contributions. Yet the summit series retains a distinctive potential: the ability to incubate specific avenues of policy collaboration among willing and capable parties, and to move more quickly and deeply on them than broader fora allow. That potential remains underdeveloped.

The summit series' distinctive potential lies not in competing with the UN's inclusive global legitimacy, but in serving as the venue where technically capable parties can begin to work towards specific, implementable agreements.

The summit drew 600,000 in-person attendees, secured over \$200 billion⁴ in AI investment commitments, and produced the New Delhi Declaration endorsed by 91 nations, including the United States and China.⁵ Yet the Declaration's breadth and non-binding character were preconditions for this inclusivity. The Frontier AI Impact Commitments, signed by 13 model developers, addressed usage transparency and multilingual evaluation rather than the safety frameworks or risk thresholds of earlier summits. The governance-focused working group ('Safe and Trusted AI') operated as one strand among seven, with no distinct deliverables or privileged access to heads of state. Observers noted that policy challenges and risks were discussed with greater urgency than ever in corridor conversations but did not make it to official outcomes. India's summit was a political success for the series' visibility and an important expansion of participation, but it confirmed that without structural focus or political prioritization, the development agenda tends to subsume the governance agenda.

Competing Visions of AI Governance

2025 saw the consolidation of three distinct models of AI governance. The current United States approach emphasizes market-led innovation and hostility to regulation, with the administration threatening subnational efforts to regulate and condemning multilateral governance and other efforts to regulate AI use abroad.

³ Yoshua Bengio et al., [International AI Safety Report 2026](#), DSIT 2026/001 (London: Department for Science, Innovation and Technology, 2026), accessed April 29, 2026.

⁴ Ron Schmelzer, "[India To Invest \\$200 Billion in AI Infrastructure: AI Impact Summit](#)," Forbes, February 20, 2026, accessed April 29, 2026.

⁵ Press Information Bureau, Government of India, "[India AI Impact Summit 2026 Witnessed Extensive Participation of Global Leaders and Industry Captains](#)," press release, March 10, 2026, accessed April 29, 2026.

Delhi confirmed the series' core structural risk: without separation of tracks, the development agenda crowds out the governance questions that most require coordinated international policy responses.

The European approach⁶, now embodied in the AI Act's implementation, represents the most advanced binding regulatory framework: certain prohibitions took effect in February 2025, obligations for general-purpose AI models in August 2025 (accompanied by a Code of Practice), with the full high-risk system requirements due in August 2026. Yet the EU is already recalibrating under competitiveness pressure. The Commission's November 2025 Digital Simplification Package⁷ proposed targeted amendments to the AI Act, and in March 2026 the Council agreed a position to delay high-risk AI system rules to December 2027. The InvestAI initiative⁸, launched at Paris with a target of mobilizing 200 billion euros, and the AI Continent Action Plan signal that Europe's strategic emphasis is shifting from regulation towards industrial policy. The tension between the EU's rights-based regulatory model and its competitiveness ambitions remains unresolved.

China has moved aggressively on domestic regulation, in keeping with its tight controls on the technology sector and its information ecosystem. The State Council's AI Plus Action Plan⁹ (August 2025) set targets for 70% AI penetration across key sectors by 2027 and 90% by 2030, making deployment the centrepiece of national strategy. In parallel, the AI Governance Framework¹⁰ (September 2025) set out principles and risk management guidelines, mandatory labelling measures for AI-generated content took effect in September 2025, and three national standards on generative AI security were issued in April 2025. China is thus pursuing a dual-track approach domestically: rapid deployment paired with sector-specific binding rules, particularly on content and information management. At the same time, China has made new multilateral institutional proposals that we discuss below.

These differing approaches reflect fundamentally different conceptions of the relationship between states, markets, and technology, and about optimal global arrangements. The summit series was founded in a moment when these differences seemed bridgeable. Today, there are differing views regarding whether the summit should simply avoid any areas that would bring these clashes into focus, or whether it should play a longer game, working to nurture collaboration among some that could meet a future political moment.

⁶ "Implementation Timeline," EU Artificial Intelligence Act, accessed April 29, 2026.

⁷ European Commission, "Digital Omnibus Regulation Proposal," Shaping Europe's Digital Future, accessed April 29, 2026.

⁸ European Commission, "EU Launches InvestAI Initiative to Mobilise €200 Billion of Investment in Artificial Intelligence," press release, February 10, 2025, accessed April 29, 2026.

⁹ State Council of the People's Republic of China, "Opinions on Deepening the Implementation of the 'AI Plus' Initiative," August 26, 2025, accessed April 29, 2026.

¹⁰ National Information Security Standardization Technical Committee (TC260), "AI Safety Governance Framework 2.0," September 15, 2025, accessed April 29, 2026.

The American Pivot to AI Nationalism

In July 2025, the United States released “[America’s AI Action Plan](#),” built on three pillars: accelerating innovation through deregulation, building domestic AI infrastructure (“Build, baby, build”), and “leading in international diplomacy and security” through countering China’s influence in international standard-setting bodies.¹¹ The Plan was accompanied by executive orders¹² accelerating federal data centre permitting and barring federal procurement of AI systems deemed to contain “ideological biases.”¹³ In parallel, the administration launched the Stargate project (a \$500 billion public-private partnership for AI data centres), and concluded bilateral Technology Prosperity Deals with the UK (September 2025), Japan and South Korea (October 2025) that align AI standards and export frameworks among allies. In March 2026, the White House released a legislative blueprint urging Congress to adopt a federal AI framework built on preemption of most categories of state laws, while NIST launched an AI Agent Standards Initiative (February 2026). The cumulative effect blends domestic deregulation and hostility to multilateral governance with bilateral deals aimed at exporting and leveraging US primacy in the AI technology stack. The Plan frames AI governance as a domain of nearly pure strategic competition rather than cooperation or a mix of the two, directing agencies to “vigorously” advocate American positions in international fora and work only with “like-minded” allies on “shared values.” Where the U.S. has occasionally explored cooperative work on governance topics, such as at the June 2026 G7 Leaders’ meeting, competition with China has been the primary frame.

This approach represents a fundamental reorientation from the Bletchley era. Where the 2023 summit sought to establish shared frameworks that might include Chinese participation, current US policy treats international AI governance primarily as a battlefield for national advantage, and one in which it should fight regulation around the world as barriers to US companies. The Plan’s skepticism towards multilateralism creates significant headwinds for the substantive international coordination the summit series was designed to facilitate.

China’s Alternative Architecture

At the World Artificial Intelligence Conference in Shanghai in July 2025, Chinese Premier Li Qiang announced China’s intention to establish a “[World AI Cooperation Organization](#)” (WAICO), with headquarters provisionally planned for Shanghai. President Xi Jinping further advocated the idea at the November 2025 APEC summit¹⁴, framing WAICO as an international public good that would coordinate global AI development, establish universally accepted governance rules, and promote open-source sharing.

¹¹ Alex Friedland, “[Trump’s Plan for AI: Recapping the White House’s AI Action Plan](#),” Center for Security and Emerging Technology, July 22, 2025, accessed April 29, 2026.

¹² Davis Wright Tremaine LLP, “[The AI Mandate: Trump Administration’s Executive Orders and AI Infrastructure](#),” Artificial Intelligence Law Advisor, July 27, 2025, accessed April 29, 2026.

¹³ Covington & Burling LLP, “[July 2025 AI Developments Under the Trump Administration](#),” Inside Government Contracts, August 18, 2025, accessed April 29, 2026.

¹⁴ Ministry of Foreign Affairs of the People’s Republic of China, “[Jointly Forging a Sustainable and Brighter Future](#),” speech, October 31, 2025, accessed April 29, 2026.

The WAICO proposal was accompanied by a detailed [13-point roadmap](#) for global AI cooperation that emphasized bridging the digital and AI divide for developing nations, a direct appeal to Global South countries that may feel marginalized by existing governance structures. WAICO's proposed structure includes a technology-sharing platform, an equity adjustment mechanism for developing countries, corporate voting rights within its governance model, and a tiered membership pathway. Whether or not it fully materializes, the proposal signals that China views the summit series as currently constituted as insufficiently representative of its interests and is prepared to construct alternative institutional architecture with concrete institutional design features that the summit series itself lacks.

Other Institutional Architectures

The UN Global Dialogue on AI Governance

The UN General Assembly formally established the Global Dialogue on AI Governance in August 2025, following the mandate in the Global Digital Compact adopted at the 2024¹⁵ Summit of the Future. The Dialogue held its launch meeting in September 2025¹⁶ and held formal sessions in July 2026 in Geneva. Its creation changes the institutional landscape: there is now an inclusive, legitimacy-conferring forum on governance questions with broad participation. Structurally, however, the format is primarily high-level national remarks, with no immediate plans for more detailed policy development or intent to launch negotiations. This reinforces our view that the summit series' comparative advantage lies not in replicating universal membership but in incubating specific cooperation among capable and willing parties, coordination that can later be socialized through UN processes or in other appropriate plurilateral or multilateral venues. The UN Global Dialogue's first formal session¹⁷ launched in July 2026 in Geneva¹⁸, immediately preceding the ITU's AI for Good Global Summit. Switzerland's role as host of the July events and as the confirmed host of the 2027 summit creates a structural link between the two processes that should be deliberately exploited and made more effectively complementary, as described in the 2025 memo. The two processes serve different functions, and should reinforce rather than crowd out one another. The Dialogue offers universal membership and legitimacy; the summit series offers depth among technically capable parties.

As noted above, WAICO positions itself as an alternative multilateral structure for AI, with a particular emphasis on South–South cooperation and technology transfer. Its appeal to countries seeking to diversify away from US- or EU-centric frameworks underlines the risk that the summit series could be bypassed if it fails to demonstrate value for a broad range of actors, including those outside traditional Western alliances.

¹⁵ United Nations, "[Global Dialogue on AI Governance](#)," accessed April 29, 2026.

¹⁶ United Nations, "[High-Level Multi-Stakeholder Informal Meeting to Launch the Global Dialogue on Artificial Intelligence Governance](#)," UN Web TV video, September 24, 2025, accessed April 29, 2026.

¹⁷ United Nations, "[Roadmap: Global Dialogue on AI Governance](#)," accessed April 29, 2026.

¹⁸ International Telecommunication Union, "[Summit 26 – Unlock AI's Potential to Serve Humanity \(AI for Good Global Summit 2026\)](#)," accessed April 29, 2026.

The summit series could be bypassed if it fails to demonstrate value for a broad range of actors, including those outside traditional Western alliances.

AISI Network

The network formerly known as the International Network of AI Safety Institutes (AISII Network) has been renamed the International Network for Advanced AI Measurement, Evaluation, and Science and expanded its technical work on evaluation, red-teaming, and model-testing methodologies. Its membership, currently including institutes or equivalents from Australia, Canada, France, Japan, Kenya, the Republic of Korea, Singapore, the United Kingdom, the United States, and the European AI Office, remains narrower than the summit series' political footprint. While the network previously had a more prominent role in shaping policy directions, it has retreated to a quieter, more technical posture, and despite its growth, has predominantly OECD-based membership, with few exceptions. India announced the establishment of the India AI Safety Institute (IndiaAISII) at the Delhi Summit, tasked with research, standard-setting, benchmarking, and technical guidance. This expands the global footprint of national safety institutes but does not resolve the structural disconnect: the AISII Network continues to operate on its own technical track, with no formal role in summit agenda-setting, commitment review, or intersessional work. The pipeline from the network's technical findings to policy discussion remains ad hoc.

A Fractured Geopolitical Moment

The summit series thus operates in a geopolitical context much less favorable to international cooperation than at the moment of its founding. The post-Bletchley optimism about building shared frameworks has given way to a more fractured landscape, marked not only by competition on AI, but by broader anxieties about fraying international order. Even the G7, traditionally a venue for coordinated action among like-minded democracies, has struggled to maintain unity on a range of topics.

Global anxiety has grown over frontier model development being dominated by the United States and, to a lesser extent, China. Other nations and regions, including Europe, the Gulf states, India, and various coalitions of middle powers, are actively discussing how to avoid being left behind. Proposals have circulated for compute-sharing consortiums, sovereign AI initiatives, energy growth initiatives, and regional partnerships aimed at ensuring broader access to the infrastructure required for meaningful participation in frontier AI development. Concerns persist about how differential rates of AI progress may reshape strategic stability, economic leverage, and the distribution of global power. At the same time, the assumption that frontier capabilities would remain concentrated in a handful of identifiable labs has been tested by the emergence of open-weight models with near-frontier performance, most

notably DeepSeek’s model. While frontier-class capabilities can be nearly reproduced at lower cost by a wider range of actors, company-level voluntary commitments, the summit series’ primary governance tool since Seoul, apply only to the original developers, not to downstream deployers or fine-tuners. This does not eliminate the case for company commitments, but it means the governance agenda must also engage with deployment-phase governance, compute access controls, and the regulatory treatment of open-weight models.

These competitive dynamics reflect fundamental interests, with nations reasonably seeking to avoid dependence on foreign providers for critical economic and security capabilities. But while broad dynamics do not favor collaboration, history suggests that even intense competition need not preclude cooperation on specific issues of mutual concern. Cold War arms control demonstrates that even adversaries can sometimes find narrow spaces for agreement when shared risks are sufficiently stark, even while competing fiercely elsewhere. The analogy is imperfect. Arms control involved bilateral negotiations between roughly symmetric powers, while AI governance involves a far more complex multi-actor landscape, but the underlying logic of narrow cooperation amid broad competition remains instructive.

The Stakes: Why Coordination Still Matters

02

II. The Stakes: Why Coordination Still Matters

Despite these headwinds, the case for international coordination on AI remains strong. Model capabilities continue to grow steadily. The recent advent of Anthropic's Mythos rang global alarm bells on cybersecurity vulnerabilities, and reminded us that as the frontier of the technology expands, so too do the risks.

There is broad value in international engagement on AI-related opportunities and challenges. Funders might collaborate on strategies to harness AI in catalyzing scientific research. Advocates might share lessons on preventing harms from non-consensual deepfake imagery. Governments might discuss opportunities to expand access to health benefits through AI tools. The summit series offers multi-stakeholder convening value, helping communities find one another and share approaches.

But there is also a distinctive case for governments coordinating on a subset of global risks that cannot be adequately managed at national scale. There is extensive writing elsewhere about such risks, but some that call for greater international coordination include:

● Systemic financial risks ⌘	AI systems are increasingly integrated into financial markets. The potential for AI-driven dynamics to generate or amplify systemic risks, through correlated trading strategies, novel manipulation, or unpredictable interactions, may create shared vulnerabilities no single jurisdiction can address alone.¹⁹ Financial institutions like the Financial Stability Board have begun to evaluate these risks. According to a recent publication from Oxford Martin AI&I on the role of investors on AI governance, “Overall, investors would like to place more focus on AI governance but face the following challenges – limited AI expertise, lack of research and resources on AI governance for investors, breadth of risks and ill-defined nature of the topic, lack of evidence for financial materiality, and caution about appearing as anti-AI.”
● Global workforce implications 👤	Serious labour displacement and extreme inequality between those who benefit from AI-driven productivity and those negatively affected by it may present challenges that transcend borders. Migration pressures, social instability, and divergent national responses could generate significant international tensions. The international community could benefit from additional collective work, building on existing data and monitoring of current trends, and considering proactive and adaptive policy responses.

¹⁹ Julian Jacobs and Andrew Sutton, “[Risks of an Intelligent Financial System](#),” OMFIF Digital Monetary Institute, January 29, 2025, accessed April 29, 2026.

●	Military and national security applications 
	The integration of AI into military systems, including nuclear command and control, autonomous weapons, and strategic decision-making, raises profound questions about stability and escalation. These technologies may disrupt existing strategic calculations and decades of deterrence frameworks, creating risks of catastrophic miscalculation. The absence of shared understandings or confidence-building measures is a gap existing arms control frameworks have not filled.
●	Loss of control and misalignment 
	As AI systems become more capable and are deployed in higher-stakes contexts, the risk that they may behave in ways not intended or anticipated becomes more consequential. The scientific community has continued to emphasize these challenges and urge policymaker action. The technical challenges of alignment are not unique to any single nation; nor are the potential consequences of failure.
●	Malign use and catastrophic terrorism 
	Advanced AI capabilities could lower barriers to the development of biological, chemical, or other weapons capable of causing mass casualties. The potential for non-state actors to exploit such capabilities creates shared security concerns requiring international attention. These challenges are already a major focus for AI developers and national institutes.

Some of these risks (and this list is by no means comprehensive) are priority technical research topics and intense focus areas for most major AI laboratories. Yet many of these risks likely require deeper policymaker focus, greater development of policy toolkits for mitigation, and more global coordination on a range of approaches, including verification. The international governance architecture for addressing these questions and potential responses remains underdeveloped. While the summit series was founded in part to fill this gap, the gap persists.

The Opportunity

03

III. The Opportunity

International coordination on complex topics requires sustained investment. Such coordination often suffers downturns during divisive times, even when underlying risks may be growing. Yet diplomatic engagement and informal coordination during such periods can build the foundation for rapid action when windows of opportunity appear. Relationships formed, concepts developed, and frameworks sketched during difficult periods can enable dramatically different levels of action when political conditions shift. Consensus cannot be built in a day, and we argue that the work of sketching out potential global responses to these challenges should continue with its gaze focused over the horizon, not stopping simply because consensus is not yet within our grasp.

The work of sketching out potential global responses should continue with its gaze focused over the horizon, not stopping because consensus is not yet within our grasp.

In this context, the summit series' value proposition requires refinement. Its distinctive potential lies not in competing with the UN's inclusive global legitimacy or WAICO's appeals to developing economies, but in serving as the venue where technically capable parties, including leading AI labs, can begin to work towards specific, implementable agreements, or refer high-priority areas of concern to other venues. This incubator function is unlikely to be replicated by universal-membership institutions. While the UN process brings legitimacy and global representation, and an important airing of high-level views, a plurilateral summit series with expert participation and intersessional work can bring focus, speed, and capacity to engage deeply with technically complex issues among parties with relevant capabilities. These approaches are complements, not substitutes.

There are a number of substantive opportunities for summit hosts and participants. The series should convene participants to focus on particular topics, exploring frameworks and probing avenues of cooperation, examining external proposals, launching workstreams for intersessional work on promising areas, and slowly identifying areas of potential convergence even when full agreement is not achievable. These efforts are important in order to deepen the continuity of policy work between summits, ensuring each gathering builds on the last.

Multi-stakeholder engagement, which to some extent is in the DNA of the summit series, also remains critical. AI developers hold crucial insights into potential risks, technical mitigation strategies, and implementation and policy opportunities and challenges. Their involvement is important for keeping governance discussions grounded in technical trends. Transparency and accountability regarding previous

commitments, including the Frontier AI Safety Commitments secured at Seoul, should be better reweaved into the dialogues. Currently there is no standing mechanism within the summit series to collectively review implementation of these commitments. Future hosts could make reporting on prior commitments²⁰ a standing agenda item, and explore options for independent assessment, peer review, or joint incident-reporting mechanisms.²¹ The Delhi Summit illustrates this gap. The Seoul Frontier AI Safety Commitments were not formally reviewed at any point. India's focus on deployment and impact meant that accountability for safety commitments fell off the agenda entirely. The 13 Frontier AI Impact Commitments announced at Delhi were new voluntary pledges on different topics (usage transparency and multilingual evaluation) rather than a reckoning with the status of prior commitments. This pattern, new commitments at each summit with no review of old ones, risks making the series a generator of aspirational language rather than a mechanism for cumulative governance.

New commitments at each summit with no review of old ones risk making the series a generator of aspirational language rather than a mechanism for cumulative governance.

The gap between policymaker action and public anxiety is growing. Surveys consistently show public concern about AI risks and a lack of confidence that governments are managing them effectively. The summit series can help address this government accountability deficit through substantive policy development that takes public priorities seriously, including through growing civil society engagement (which the Government of India has been doing through stakeholder meetings).

²⁰ Zach Stein-Perlman, "[AI Companies' Commitments](#)," AI Lab Watch, accessed April 29, 2026.

²¹ Organisation for Economic Co-operation and Development, [Towards a Common Reporting Framework for AI Incidents \(Paris: OECD Publishing, 2025\)](#), OECD Artificial Intelligence Papers, No. 34, accessed April 29, 2026.

Structural Options Worth Considering

04

IV. Structural Options Worth Considering

The structural options we outlined one year ago remain relevant, though they must now be considered in light of the challenging international environment. The combination of US disengagement from multilateral AI governance, Chinese promotion of alternative institutions, and European consolidation around its own regulatory approach is likely to create serious obstacles to rapid consensus development. Future hosts may need to pursue structural reforms through narrower coalitions of committed parties, potentially accepting that some major AI powers will not participate, at least for now. The Delhi Summit provides a partial proof of concept: despite US public rejection of global AI governance, the US endorsed the non-binding New Delhi Declaration and participated in the CEO Roundtable, and China also endorsed the Declaration while remaining minimally present. The 91-nation endorsement demonstrates that soft, non-binding outputs can attract broad sign-on even in a fractured environment. But it also demonstrates the tradeoff: the Declaration's breadth and non-binding character were preconditions for this inclusivity. For the series to produce more specific, implementable agreements, the drafting and negotiation work will need to happen in a smaller circle of technically capable and politically willing parties, with broader endorsement sought after the fact. The Switzerland-UAE hosting succession offers a natural opportunity to pilot this approach. The table below lists structural reform recommendations that we believe will prove valuable over time.

Measure	Core description	Purpose in the summit series 2025	2026 Update
Formalized hosting arrangements	Combine a bidding system with regional rotation, with hosts selected and announced well in advance (e.g. two years) to allow adequate preparation.	Reduce uncertainty, improve planning, and ensure a fair, transparent process for selecting hosts among leading AI countries.	Use rotation and advance notice to keep a functioning core even if some major powers disengage in particular years.
Hybrid secretariat model	Establish a small semi-permanent core team that provides institutional memory and operational continuity, complemented by a host-appointed team for each summit.	Preserve host flexibility while avoiding repeated “reinvention” of processes, documentation, and follow-up between summits.	Team tracks implementation of past commitments, interfaces systematically with other fora (UN, AISI Network, WAICO where appropriate).
Two-track participation framework	Maintain a focused Governance Track on advanced AI governance, and a broader Action Track on public-interest uses of AI.	Balance effectiveness (small group able to agree on technical and risk issues) with legitimacy and inclusiveness on broader socio-economic questions.	Allow Governance Track to proceed among willing parties even if countries opt out, while using Action Track to keep engagement on broader set of issues.
Coordinated succession and agenda-setting	Use a “troika” of previous, current, and next hosts and a Governance Track steering committee to develop and periodically update a multi-year roadmap for the series.	Ensure each summit builds on the last, avoid abrupt agenda swings, and sustain progress on long-term files (e.g. safety standards, incident reporting).	Use the roadmap to keep a standing slot for accountability on previous commitments (e.g. Seoul Frontier AI Safety Commitments) and for responding to emerging risks.
Annual summits plus interim technical-ministerial meetings	Maintain an annual leaders’ summit, complemented by smaller technical or ministerial meetings to track progress and respond to fast-moving developments.	Preserve high-level political attention while creating space for detailed technical work, peer learning, and monitoring between headline events.	Use interim meetings for joint evaluation work, red-teaming coordination, and exchanges between national AI safety institutes and regulators.
Regular engagement with the UN and AISI Network	Systematically align with the UN Global Dialogue on AI Governance and the AISI Network through information-sharing, timing coordination, and joint sessions where appropriate.	Avoid duplication, increase legitimacy, and channel summit outputs into broader, more inclusive processes; draw on specialized technical expertise.	Position the summit series as an incubator: test and refine cooperation among capable parties, then feed successful arrangements into UN and regional processes, while guarding against exclusion of key actors not in the AISI Network.

TABLE 1

Such structural evolution would not resolve underlying geopolitical tensions. But it would help ensure that when windows of opportunity appear, the summit series is better positioned to act, with greater institutional capacity, accumulated trust, and substantive preparation to translate political will into outcomes.

Future hosts could also advance incremental versions of these or other reforms, experimenting with modest steps towards greater institutionalization (for example, even designating an informal technical working group to support summit efforts, before a formally established secretariat). The Switzerland-UAE co-hosting arrangement announced at Delhi is the first concrete opportunity to test the coordinated succession model. Both countries have positioned themselves as bridging actors between major AI powers. A shared planning unit between the 2027 and 2028 hosts, a jointly agreed set of agenda items that carry across both summits, and a mechanism for reviewing implementation of Delhi commitments at Geneva would demonstrate that continuity is achievable and create institutional precedent for successors. Outside of host action, there may also be space for civil society and other stakeholders to usefully develop more robust engagement approaches, formulating proactive agenda recommendations and engaging with national governments well in advance of summits, as occurs in other international processes.

Looking Ahead: A Realistic Assessment

The summit series has reached a turning point. The optimism of Bletchley, when the world's leading AI powers signed a joint declaration on managing frontier risks, has given way to a more fragmented and uncertain landscape. The United States, represented in Delhi by Michael Kratsios, Director of the White House Office of Science and Technology Policy, who declared²² that the US 'totally rejects global governance of AI,' has pivoted towards deregulation and strategic competition. China endorsed the New Delhi Declaration but was largely absent from substantive sessions, with its delegation minimized during Chinese New Year. The European Union has consolidated its own distinct approach. The UN Global Dialogue now provides an inclusive but not yet action-oriented forum.

In this context, the summit series must be realistic about what it can achieve. It cannot substitute for bilateral channels between Washington and Beijing on strategic stability questions. It cannot provide the universal legitimacy the UN confers. It cannot compel cooperation among unwilling parties.

What it can do is maintain a venue where technically capable parties, including governments, labs, and researchers, continue substantive work on specific governance challenges: interoperability of governance frameworks, incident-reporting mechanisms, technical standards for evaluation, coordination on emerging risks, and exchanges of views regarding policy mitigations, both national and global. It can serve as an incubator for agreement frameworks that might later be negotiated by other fora. It can preserve and build relationships and institutional capacity for rapid action when political windows open.

²² Michael Kratsios, "[Remarks by Director Michael Kratsios at the India AI Impact Summit](#)," The White House, February 20, 2026, accessed April 29, 2026.

The work of consensus-building is slow, but can make important, behind-the-scenes progress even when it seems most unlikely. Future hosts, including Switzerland, and those who follow, can advance structural reforms incrementally to help position the series to matter when conditions shift.

The work of consensus-building is slow, but can make important, behind-the-scenes progress even when it seems most unlikely.

As AI reshapes society, the global public is increasingly demanding a thoughtful response, and we argue that the structure of governance conversations must continue to mature alongside the technology and its deployment.

This update supplements our January 2025 policy memo, “The Future of the AI Summit Series.” The original memo’s detailed analysis of design options and implementation considerations remains current and should be consulted alongside this update.

Given the number of contributors, authorship does not imply agreement with every recommendation.

OXFORD **AI DIPLOMACY** LAB

