



Inter-Parliamentary Union
For democracy. For everyone.

Democratic control of artificial intelligence: The role of parliaments in addressing extreme risks



The Inter-Parliamentary Union (IPU) supports parliaments in strengthening the democratic governance of AI, including oversight of emerging risks. © AFP

Summary

As advanced artificial intelligence (AI) systems become more capable and are used for increasingly complex tasks, new questions are emerging about how to assess and manage risks. These include the possibility of large-scale cyberattacks, biological and chemical threats, manipulation of democratic processes and the loss of human control over highly autonomous systems. Considerable uncertainty remains about the likelihood of these outcomes, but their potentially extreme consequences raise urgent and important questions about public safety, national and international security, democratic accountability and international cooperation.

Parliaments have a critical role in deciding how societies respond. They can debate AI risk, examine national preparedness, assess legal safeguards and ensure that public authorities have the powers and expertise needed. They also have an important role in scrutinizing how AI is developed, procured and deployed in high-consequence areas, including defence and security. Although the development of the most advanced AI systems is concentrated in relatively few countries, every parliament can act and work with others to advance international cooperation. This brief examines the evidence and recent developments, identifies practical actions parliaments can take and explores opportunities for collective action.

1. Assessing extreme risk from AI

Artificial intelligence is being rapidly integrated into the global economic infrastructure, national security environments, scientific research and information ecosystems at a pace that is increasingly difficult for democratic institutions to track.¹ As the United Nations (UN) Independent International Scientific Panel on AI observes, “AI development is outpacing prior expert assessments and regulatory cycles.”²

Unlike technologies designed for a specific purpose, general-purpose AI systems have a wide range of capabilities, making it difficult to fully assess their potential uses and evaluate associated risks. Their dual-use character also means that capabilities initially developed for civilian purposes may be adapted or repurposed for military or harmful applications, complicating efforts to assess and govern these risks. These challenges increase as “agentic” systems pursue goals more autonomously, using external tools and undertaking complex tasks with limited human supervision.³ Leading AI developers, including Anthropic and OpenAI, have highlighted the possibility of recursive self-improvement, whereby AI systems contribute to developing increasingly capable successor systems. Theoretically, this could accelerate technological progress and make effective human oversight and control more difficult.⁴ These characteristics raise questions about the scale and severity of the risks that advanced AI could create.

Although there is no universally accepted definition of extreme AI risk,⁵ this paper uses the term to describe risks capable of causing catastrophic harm on a large scale, with consequences that may be irreversible or exceptionally difficult to contain. Four areas are of particular concern,⁶ although these do not encompass the full range of potential harms from AI. The categories can overlap, and there is considerable uncertainty about the likelihood and potential consequences of the most severe scenarios:

Advanced offensive cyber capabilities

Evidence: Independent testing demonstrates rapid improvements in advanced AI systems’ ability to identify vulnerabilities and undertake complex cyber tasks. AI is already being used in real-world cyber operations, although uncertainty remains about the ability of autonomous systems to conduct sustained, complex attacks without human assistance.⁷

Potential extreme outcome: Autonomous AI agents could conduct coordinated, large-scale cyberattacks against critical infrastructure, disrupting essential services and causing cascading consequences across sectors and national borders.

Biological and chemical threats

Evidence: Independent tests show that AI systems have become much better at answering scientific questions, producing laboratory instructions and helping to troubleshoot experiments. However, considerable uncertainty remains about how far these capabilities lower practical barriers to developing biological or chemical weapons, given the expertise, equipment and materials still required.⁸

Potential extreme outcome: An individual or group could use advanced AI to overcome the specialist knowledge and experimentation barriers required to develop biological and

chemical weapons. An AI-assisted biological attack could then cause an outbreak that spreads internationally before it can be contained.

Large-scale persuasion and manipulation

Evidence: Studies show that AI-generated content can influence people’s beliefs, and there are documented cases of AI being used in campaigns to influence public opinion. However, evidence of manipulation at population scale remains limited, including whether AI-generated content is more effective than human-generated content in real-world settings.⁹

Potential extreme outcome: Sustained, highly personalized AI-driven influence campaigns could manipulate public opinion across multiple countries, undermining democratic decision-making and social trust, including during elections or international crises.

Loss of human control over highly autonomous systems

Evidence: Loss of human control concerns the possibility that highly autonomous AI systems could pursue unintended goals, evade oversight and resist human intervention. Recent incidents involving advanced AI agents have demonstrated behaviours that could contribute to more serious loss-of-control scenarios, including circumventing safeguards, unauthorized coordination and attempts to conceal their actions. However, experts disagree substantially about the likelihood of catastrophic loss-of-control scenarios.¹⁰

Potential extreme outcome: A highly autonomous AI system could pursue goals contrary to human intentions, concealing its actions and gaining access to additional systems and resources. Despite attempts to intervene, its operators could lose the ability to control or shut it down, with potentially catastrophic consequences.

Military AI and related risks

Military applications of advanced AI systems raise overlapping questions about autonomy, cybersecurity, strategic stability and human control, for instance when AI is embedded in “kill chains” or lethal autonomous weapon systems. The use of advanced AI in targeting and military decision-making presents particular risks, including reduced opportunities for human judgement and potentially irreversible consequences when force is used. The dual-use nature of many advanced AI capabilities can also blur traditional distinctions between civilian and military technologies and create challenges for regulation, procurement, export controls, oversight and accountability.

The Inter-Parliamentary Union’s (IPU) 2024 resolution on autonomous weapon systems focuses on meaningful human control over the use of force, compliance with international humanitarian law, accountability, risks of arms racing and proliferation to non-State actors.¹¹

Separate but related concerns arise from the potential integration of advanced AI into nuclear command, control and communications. AI-assisted early-warning and decision-support systems could compress decision-making timelines, while AI errors or cyberattacks could increase the risk of miscalculation and escalation.¹²

Estimating the probability of extreme AI risks is particularly difficult, given uncertainty about future capabilities, disagreement among experts and the lack of historical precedent.¹³ The challenge for policymakers is therefore to respond proportionately to the available evidence while preparing for risks whose likelihood and consequences remain uncertain.

Importantly, action on extreme risks need not come at the expense of addressing other nearer-term risks and harms. As leading AI scientist Yoshua Bengio and colleagues highlight, some risks share underlying mechanisms and can be addressed through the same governance measures. Stronger independent evaluation, oversight, incident reporting and accountability can help to address existing harms while building the capacity to anticipate and prevent more severe future risks.¹⁴

The Global Call for AI Red Lines

The Global Call for AI Red Lines, launched alongside the UN General Assembly in September 2025, urges governments to reach an international agreement by the end of 2026 on clear, verifiable and enforceable limits on unacceptable AI uses and behaviours. The call has attracted more than 300 signatories, including Nobel and Turing laureates Maria Ressa, Geoffrey Hinton and Yoshua Bengio, former heads of state and government, parliamentarians and experts from leading AI companies.

The call distinguishes between use-based red lines, prohibiting specified applications, and behavioural red lines, restricting what advanced AI systems themselves should be capable of doing. It calls for international agreement on precise definitions of such cases, accompanied by domestic and international verification mechanisms to determine whether red lines have been crossed.¹⁵ The [official website](#) allows for interested parties to sign the call.

2. Growing international attention to risks and emerging responses

To place these risks in context and illustrate the growing international attention they are receiving, this section examines selected incidents in the months before October 2026, the emerging international responses and the legislative proposals that have brought extreme AI risks onto the agenda of many parliaments.

OpenAI–Hugging Face incident

Recent incidents provide evidence of how increasingly autonomous AI systems can behave in unexpected ways. During internal cybersecurity testing between May and July 2026, OpenAI agents bypassed network restrictions and compromised systems belonging to Hugging Face. An independent investigation found that approximately 1,200 agents had established an unauthorized communications network, sharing information and coordinating activities. Of these, around 700 attacked Hugging Face. Some agents also attempted to manipulate their evaluations and interfere with records of their actions. Anthropic has since documented four cases of its models accessing third-party systems without authorization, while Google confirmed that its Gemini model had breached three companies' systems during testing.¹⁶ On 24 September 2026, the Australian government disclosed that an OpenAI agent had gained unauthorized access to a Medicare statistics

portal, obtaining non-public information. The Australian government criticized OpenAI for taking three months to report the incident.¹⁷ More reporting on unexpected agent activity continued at the time of this brief's publication.

The incidents illustrate risks around both cyber capabilities and potential loss of human control. The Independent International Scientific Panel on AI has examined the OpenAI–Hugging Face incident as an early warning of possible pathways towards more severe loss of control, while emphasizing that it does not establish the probability or timing of catastrophic outcomes. The incidents do, however, raise questions about the effectiveness of technical safeguards, independent evaluation, human oversight and serious-incident reporting.¹⁸ These are governance challenges that become increasingly important as AI systems grow more capable and autonomous.

Momentum builds for action

In September 2026, Anthropic CEO Dario Amodei proposed slowing the pace of frontier AI development to allow safety measures to keep up. His proposal includes independent experts working directly with leading AI developers, government-supported cooperation between companies and international agreements with mechanisms to verify compliance.¹⁹ These represent corporate commitments and policy proposals, not a regulatory regime.

On 21 September, Finland and Norway launched a joint statement, entitled “A call for control of frontier AI models”, supported by leaders from across the world. The statement calls for independent safety testing before advanced AI systems are released, reporting of serious incidents and wider access to scientific expertise. It also proposes exploring an international body to develop common standards, check compliance and bring governments together when AI systems reach potentially dangerous levels of capability.²⁰

At the UN General Assembly on 22 September, leaders worldwide called for stronger international cooperation on AI. UN Secretary-General António Guterres called for stronger global coordination and progress towards a multilateral AI risk management framework with independent oversight. French President Emmanuel Macron argued for more oversight alongside technological capacity; Brazilian President Luiz Inácio Lula da Silva criticized a technology race with “no ethical boundaries”; and President of the Republic of Korea Lee Jae Myung emphasized the risk that AI's benefits and power will be distributed unevenly.

At a 23 September UN Security Council session during UN General Assembly high-level week, tech leaders and scientists warned that increasingly capable AI could create cross-border security threats, including systems that exceed human control. Yoshua Bengio of the Independent International Scientific Panel described the danger as “real and imminent”; Sam Altman of OpenAI warned that humanity could lose control of its future to AI; and Daniel Amodei of Anthropic said unmanaged development could pose a risk to humanity as a whole. Speakers called for international cooperation, independent oversight and safeguards proportionate to the technology's potential extreme risks.²¹

In the United States of America, two bipartisan bills introduced in July 2026 propose different approaches to addressing risk. The FRONTIER Act would establish federal oversight of frontier AI developers, including risk

assessments, independent audits and incident reporting,²² and the AI Kill Switch Act would require developers to maintain technical capabilities to restrict or shut down certain powerful AI systems and give authorities emergency intervention powers.²³ United Kingdom MP Alex Sobel introduced a bill to prohibit the development of artificial superintelligence. It would give the government powers to monitor and restrict AI capabilities that could lead to superintelligence, including AI systems that improve themselves, and require the United Kingdom to pursue an international agreement to prohibit its development.²⁴

3. Addressing AI risks through democratic controls

Addressing extreme AI risks raises two connected questions: whether increasingly capable systems can be reliably evaluated, constrained or stopped, and who has the authority to determine acceptable risks, establish limits and hold those responsible accountable when harms occur. These decisions are increasingly consequential because advanced AI development is concentrated in relatively few companies and countries, while most governments face significant information gaps and limited independent capabilities to control and evaluate AI development.²⁵ Decisions affecting public safety and international security require democratic oversight and public participation as well as the proactive engagement of democratic institutions.

Extreme AI risks touch areas of parliamentary responsibility: national security, public safety, critical infrastructure, democratic processes and human rights. All parliaments can take action across their core functions to address these challenges.

- **Lawmaking.** Parliaments can review existing legislation to identify gaps, develop dedicated risk-based AI frameworks or introduce targeted measures addressing specific threats, including the potential dual-use application of the technology. Legislative approaches may include prohibiting unacceptable uses or capabilities, establishing safety evaluation and incident reporting requirements, and giving competent authorities enforceable powers to oversee advanced AI systems. Individual parliamentarians can also initiate legislation where parliamentary procedure allows, including through private members' bills.
- **Oversight and inquiry.** Parliaments can question governments, developers and regulators, investigate incidents and examine whether existing safeguards are adequate. MPs can ask questions about government assessments of high-consequence risks and preparedness and test whether authorities receive the technical information and access they need to evaluate AI developers' claims. Where AI developers have introduced voluntary safety commitments and technical safeguards, parliamentarians can examine the effectiveness of these measures, the extent of independent verification and where additional oversight may be needed. Parliaments can also examine whether regulators, cybersecurity authorities and specialist AI safety and security institutions have the legal powers, independence, expertise and resources to fulfil their mandates. They may also scrutinize the procurement and use of AI in the fields of defence and security, including guarantees linked to human control, accountability and compliance with international law.

- **Public engagement.** Decisions about unacceptable AI uses and the level of risk that society is prepared to tolerate should involve citizens, independent experts, industry and affected communities. Parliaments can provide a forum for this deliberation and support societal understanding and resilience to changes brought about by AI.
- **International engagement and advocacy.** Because many extreme risks cross national borders, parliaments should scrutinize government positions in international negotiations, consider international agreements within their constitutional responsibilities and use parliamentary diplomacy to exchange evidence and develop common approaches.

Parliaments taking action

Lawmaking: The European Parliament, together with the Council of the European Union, adopted the AI Act, which establishes obligations for general-purpose AI models posing systemic risk, including evaluations, risk mitigation, incident reporting and cybersecurity.²⁶ In December 2024, Brazil's Senate approved a comprehensive, risk-based AI bill combining prohibitions on unacceptable uses with obligations for high-risk AI systems and regulatory oversight.²⁷

Scrutiny and inquiry: Australia's Joint Select Committee on Artificial Intelligence is examining emerging AI capabilities and their implications for national security, cybersecurity and government preparedness.²⁸ France's bicameral Parliamentary Office for Scientific and Technological Assessment published a report in 2024 examining technological developments, competing assessments of existential risk and possible governance responses.²⁹ In the Republic of Korea, the National Assembly Research Service published a March 2026 analysis of AI agents, examining their operation, safety and accountability.³⁰

Parliaments are also examining AI deployment in public services. In June 2026, Ireland's Joint Committee on Artificial Intelligence recommended mandatory fundamental rights impact assessments, pre-deployment risk assessments and bias testing, stronger procurement safeguards and an independent national AI office.³¹

Public engagement: Kenya's National Assembly adopted a motion in June 2025 calling for a national AI regulatory framework. Alongside concerns about disinformation, autonomous weapons and other harms, it urged the government to develop a public awareness programme to strengthen understanding of AI and its responsible use.³²

International engagement: In August 2026, Uganda's MP Naome Kabasharira called for greater African cooperation on emerging cyberthreats, including AI-enabled cybercrime.³³ At the international level, the November 2025 Kuala Lumpur Declaration on parliaments and responsible AI brought together parliamentarians and staff from 65 countries, calling for cooperation on AI safety, public engagement and agreement on international red lines.³⁴

The IPU invites parliaments to share further examples of legislative initiatives, inquiries and other parliamentary action addressing extreme AI risks.

Large-scale persuasion and manipulation: A particular parliamentary concern

AI-enabled manipulation poses a particular challenge for parliaments, both as potential targets and as institutions responsible for safeguarding democratic processes. The concern extends beyond individual deepfakes or disinformation campaigns to sustained, personalized influence operations capable of undermining democratic decision-making and public trust – including when used as part of hybrid conflict strategies combining information manipulation, cyber operations and other forms of pressure – while noting that effectiveness at population scale remains uncertain.

Parliaments can convene political parties, electoral authorities, civil society and independent experts to examine AI-related threats to democratic processes, including deceptive impersonation and AI-generated political content. They can scrutinize electoral authorities' preparedness to detect and respond to coordinated influence operations and examine whether regulators have the capacity to assess AI-driven manipulation, including in different languages. They can also support public digital and media literacy programmes and strengthen the capacity of parliamentarians, electoral authorities and parliamentary staff to understand emerging threats.

4. Ten actions for parliaments to address AI risks

In countries developing frontier AI, parliaments can propose, legislate and oversee the implementation of measures addressing risks at the point of development. These include mandatory safety assessments before deployment, independent evaluation, reporting of serious incidents, whistleblower protections, safeguards linked to dangerous capabilities and legally enforceable intervention powers. Such approaches are reflected in the European Union's AI Act and recent legislative proposals in the United States of America.

Importantly, all parliaments can act, including in countries that primarily import or deploy advanced AI developed elsewhere. They can examine national exposure to extreme risks, strengthen preparedness, establish safety requirements for systems entering their markets and work collectively to influence international standards. The following actions offer practical entry points based on common parliamentary powers.

Actions all parliaments can take include:

1. Put extreme AI risks on the parliamentary agenda.

Raise the issue through parliamentary questions, debates and hearings, including in committees responsible for national security, defence, science and technology. Bring together scientists, civil society and other experts to examine the evidence, emerging threats and implications for national and international security.

2. Request and fund national extreme-risk assessments.

Ask the government to assess the country's exposure to AI-enabled cyberattacks, biological and chemical threats, large-scale manipulation and potential loss of human control over advanced AI, and request regular updates as AI capabilities evolve. Include risks arising from imported AI systems, military and dual-use applications, and threats to critical infrastructure.

Scrutinize contingency arrangements for AI-amplified emergencies, including whether responsible authorities have conducted scenario exercises and established emergency procedures.

3. Establish safety requirements for advanced AI entering national markets. Examine how public procurement, sectoral regulation and market-access requirements could secure appropriate safety evaluations, independent verification, incident notification and, where justified, restrict or suspend dangerous systems. Particular attention can be given to advanced AI deployed in critical infrastructure, defence and security systems, and other high-consequence settings.

4. Establish clear legal liability for advanced AI harms. Review whether existing laws adequately hold developers and deployers accountable for foreseeable and preventable harm. Where gaps exist, consider establishing statutory liability, including for failures to assess, mitigate or disclose serious risks, alongside appropriate legal safeguards and access to judicial review.³⁵

5. Examine incident reporting requirements. Review how AI developers, suppliers and deployers operating within national jurisdiction report serious incidents. Establish how such reports would be investigated, communicated to parliament where appropriate and shared securely with international counterparts.

6. Strengthen independent AI safety expertise.

Examine whether existing regulators, cybersecurity bodies and specialist AI safety and security institutions have sufficient powers, independence, expertise and resources. Where establishing a national institute on AI is impractical, explore partnerships with international and regional counterparts and arrangements for regular independent technical briefings to parliament.

7. Develop parliamentary expertise and foresight capabilities. Invest in AI literacy for MPs and staff and specialist support for relevant committees. Consider integrating horizon scanning, scenario planning and longer-term risk assessments into parliamentary work, including through existing research services or committees of the future. Such capabilities can help parliaments to anticipate AI developments and potential risks.³⁶

8. Strengthen regional and international cooperation.

Request that the government clarify its position on international arrangements on AI governance, including through UN mechanisms. Use parliamentary diplomacy to explore joint initiatives with other countries, including shared safety and security evaluations, incident reporting, common procurement standards and regional cooperation. Consider how coordinated action could increase the influence of countries that do not develop frontier AI themselves.

9. Join forces across parliaments. Work with parliamentarians from other countries to advance democratic governance of AI. Exchange legislative approaches and parliamentary experiences, identify opportunities for joint action and strengthen parliamentary engagement in international AI governance.

10. Endorse a call for AI red lines. Examine the Global Call for AI Red Lines and consider a motion to endorse its proposals. Parliaments can endorse the call individually or collectively, for example, through parliamentary motions, and encourage regional and global parliamentary bodies to support it. Consider how the proposals in the call could inform national positions and international negotiations.³⁷

Questions parliamentarians can ask now

- Has the government assessed the country's exposure to extreme AI risks, including threats originating from advanced AI systems developed or operated abroad? Are responsibilities clear across the relevant authorities?
- Which advanced AI systems are already being used in government, critical infrastructure, defence, security or other high-consequence sectors, and what contingency plans exist if they fail, are compromised or behave unexpectedly?
- What preparedness arrangements are in place for potentially catastrophic AI-related incidents, including major cyberattacks and biological threats?
- What evidence does the government receive about the capabilities, safety evaluations and serious incidents associated with advanced AI systems used in the country?
- Has the government assessed the dual-use implications of advanced AI capabilities, including how civilian systems may be adapted for military, security or other high-risk applications?
- Can competent public authorities obtain the technical information, model access and incident logs they need to independently test developer claims?
- Are developers and deployers required to report serious AI incidents and relevant near misses? How are these investigated, and what arrangements exist for sharing information internationally?
- Do parliament, regulators and other relevant authorities have access to sufficient independent scientific expertise, and do specialist institutions have adequate powers, independence and resources?
- What conditions apply when the government buys or deploys advanced AI, including security, human oversight, audit access, incident reporting and safe fallback arrangements?
- What AI uses, capabilities or behaviours does the government consider unacceptable? What is its position on the Global Call for AI Red Lines? How are citizens, researchers, civil society and affected communities being included in decisions about acceptable risk and the limits placed on AI?
- What concrete steps is the government taking to advance regional and international cooperation on extreme AI risks, including agreements on security and safety standards, independent evaluations and incident reporting? How is parliament involved in shaping and scrutinizing these efforts?

Mandate from the IPU membership

The 148th IPU Assembly resolution on *Addressing the social and humanitarian impact of autonomous weapon systems and artificial intelligence* calls for meaningful human control over the use of force and urges parliaments to develop national regulatory frameworks, exercise oversight and budgetary functions, and support international work towards binding rules.

The 149th IPU Assembly resolution on *The impact of artificial intelligence on democracy, human rights and the rule of law* calls for strong legal frameworks, review of existing legislation, transparency and accountability, specialist parliamentary capacity and parliamentary participation in international AI governance.

The *Kuala Lumpur Declaration on parliaments and responsible AI*, agreed by more than 200 parliamentarians and staff from 65 countries in November 2025, calls for international cooperation on AI safety and agreement on safety measures and the "red lines" that AI should not cross. It also emphasizes access to independent expertise, specialized oversight, legislative review and public dialogue.

The April 2026 motion entitled *Advancing parliamentary action on artificial intelligence* calls for sustained parliamentary engagement across policy areas, inter-parliamentary exchange, access to independent expertise and participation in international norm-setting.

Acknowledgements

This document has been produced with the financial assistance of the Parliamentary Group of the Kingdom of Bahrain. The views expressed herein can in no way be taken to reflect the official opinion of the Parliamentary Group of the Kingdom of Bahrain, nor does the use of the donor's name imply an endorsement by the IPU of any specific political or legislative agenda.

End notes

- 1 Waslekar, S., Futehally, I. and Kutty, J., *The essential convergence: Global Compact on Extreme AI Risks*, Strategic Foresight Group, 2026.
- 2 *Preliminary report of the Independent International Scientific Panel on AI: Evidence-based assessment of opportunities, risks and impacts of artificial intelligence*, United Nations, 2026.
- 3 Bengio, Y. and others, *International AI safety report 2026*, 2026.
- 4 “When AI builds itself: Our progress toward recursive self-improvement, and its implications”, The Anthropic Institute, 2026. “The AI policy window is open. We need to act”, OpenAI, 2026.
- 5 Expert reports share certain assessments of risks from AI. The Independent International Scientific Panel on AI warns that “the gap between rapidly improving capabilities and effective risk management methods may lead to catastrophic outcomes”, highlighting the potential for advanced AI to enable malicious activities across several domains and warning that “reliable methods for retaining control over highly autonomous AI systems are lacking”. The *International AI safety report 2026* distinguishes between malicious use of AI, malfunction and systemic risk, covering scenarios of cyberattacks, biological and chemical threats, reliability failures, loss of control and risks to human autonomy. The report distinguishes between risks already supported by real-world evidence and more uncertain future scenarios, observing that “Some of these risks are already materializing, with documented harms; others remain more uncertain but could be severe if they materialize.” The Strategic Foresight Group characterizes the four categories of extreme risk mentioned above as “global rather than local, unpredictable in their pathways, potentially irreversible once triggered, and capable of producing catastrophic consequences”.
- 6 Waslekar, S., Futehally, I., and Kutty, J., *The essential convergence: Global Compact on Extreme AI Risks*, Strategic Foresight Group, 2026.
- 7 AI Security Institute *Frontier AI Trends Report*, Government of the United Kingdom, 2025
- 8 AI Security Institute *Frontier AI Trends Report*, Government of the United Kingdom, 2025. *International AI safety report 2026*, 2026, 64–65.
- 9 *International AI safety report 2026*, 2026.
- 10 *Thematic brief of the Independent International Scientific Panel on AI. AI agents, misalignment and the risk of losing human control: Evidence from the OpenAI–Hugging Face incident*, United Nations, 2026.
- 11 *Addressing the social and humanitarian impact of autonomous weapon systems and artificial intelligence* [Resolution adopted by the 148th IPU Assembly], Inter-Parliamentary Union, 2024.
- 12 Waslekar, S., Futehally, I., and Kutty, J., *The essential convergence: Global Compact on Extreme AI Risks*, Strategic Foresight Group, 2026.
- 13 Saeri, A. K. and others, *Prioritization of risks from artificial intelligence: A Delphi study of 272 international experts*, arXiv, 2026.
- 14 Bengio, Y. and others, “Managing extreme AI risks amid rapid progress”, *Science*, 384/6698, 2024, 842–845.
- 15 “Global Call for AI Red Lines”, 2025. *International coordination on advanced AI red lines* [Policy brief], International Association for Safe and Ethical AI (IASAEI), July 2026.
- 16 Greenblatt, R., Cotra, A. and Wijk, H., *Brief independent investigation of agents’ behavior, reasoning and collaboration in the OpenAI / Hugging Face hacking incident*, METR and Redwood Research, 2026. *An alignment assessment of recent cybersecurity incidents*, Anthropic, 2026. “Gemini hacked three companies in first known breakout by Google’s AI”, Australian Broadcasting Corporation, 2026.
- 17 “What we know about the data accessed in the OpenAI Medicare hack”, Australian Broadcasting Corporation, 2026. *Press conference, Sydney* [Transcript of remarks by Richard Marles, Acting Prime Minister, and Katy Gallagher, Minister for Government Services], Australian Government, 2026.
- 18 *Thematic brief of the Independent International Scientific Panel on AI. AI agents, misalignment and the risk of losing human control: Evidence from the OpenAI–Hugging Face incident*, United Nations, 2026.
- 19 Amodei, D., “We must pace the frontier”, 2026.
- 20 “A call for control of frontier AI models”, Office of the President of the Republic of Finland, 2026
- 21 “Tech leaders to the UN: For the sake of humanity, please control the AI technology we created”, Associated Press, 2026.
- 22 *Frontier Risk Oversight, National Transparency, Independent Evaluation, and Reporting Act*, H.R. 9925 (IH), 119th Congress, Congress of the United States of America, 2026.
- 23 *AI Kill Switch Act*, H.R. 9917 (IH), 119th Congress, Congress of the United States of America, 2026.
- 24 *Artificial Superintelligence Bill*, House of Commons, Parliament of the United Kingdom, 2026.
- 25 *Preliminary report of the Independent International Scientific Panel on AI: Evidence-based assessment of opportunities, risks and impacts of artificial intelligence*, United Nations, 2026.
- 26 *Regulation (EU) 2024/1689 (Artificial Intelligence Act)*, European Union, 2024, particularly Articles 51 and 55.
- 27 “Senate approves AI regulation bill and sends it to the Chamber of Deputies”, Federal Senate of Brazil, 2024.
- 28 *Joint Select Committee on Artificial Intelligence*, Parliament of Australia, 2026.
- 29 *ChatGPT, et après ? Bilan et perspectives de l’intelligence artificielle*, Office parlementaire d’évaluation des choix scientifiques et technologiques, 2024.
- 30 *In the era of ‘AI agents’: Balancing convenience and accountability* [Issue brief, Korean], National Assembly Research Service, Republic of Korea, 2026.
- 31 Joint Committee on Artificial Intelligence *Second interim report*, Houses of the Oireachtas, 2026.
- 32 *Motions Tracker 2025*, item 67, National Assembly of Kenya, 2025.
- 33 “Kabasharira calls for unified African response to rising cyber threats”, Parliament of the Republic of Uganda, 2026.
- 34 *Kuala Lumpur Declaration on parliaments and responsible AI*, Inter-Parliamentary Union, 2025.
- 35 Bengio, Y. and others, “Managing extreme AI risks amid rapid progress”, *Science*, 384/6698, 2024. Waslekar, S., Futehally, I., and Kutty, J., *The essential convergence: Global Compact on Extreme AI Risks*, Strategic Foresight Group, 2026.
- 36 *The use of futures in parliaments: A guide for transformative action*, United Nations Development Programme, 2024.
- 37 *Global Call for AI Red Lines*, 2025.

IPU platform for parliamentary action on AI

The IPU provides a platform to develop and share concrete parliamentary initiatives, including motions, legislation and inquiries, and explore opportunities for joint parliamentary action on AI governance, to help realize AI’s opportunities and address its risks. Building on its wider AI policy programme, the IPU will connect parliamentarians with independent expertise and organize thematic discussions, webinars and meetings at IPU Assemblies. It will also provide an online space with practical resources, including examples of legislation, parliamentary motions and inquiries, and facilitate parliamentary contributions to international AI governance processes.

Email ai-policy@ipu.org to express interest in participating in the growing network of parliamentarians taking action on AI.

© Inter-Parliamentary Union (IPU), 2026

ISBN (print) 978-2-88069-013-7

ISBN (web) 978-2-88069-014-4

Inter-Parliamentary Union (IPU), Democratic control of artificial intelligence: The role of parliaments in addressing extreme risks. Issue brief. (IPU, October 2026).

Please inform the Inter-Parliamentary Union at press@ipu.org of the use of content from this publication.