



The Development of the UN Scientific Panel on AI

Policy Paper



Mila

Quebec Artificial
Intelligence Institute



March 2025

Authors (in alphabetical order)

Jonathan Barry

Yoshua Bengio

Anna Jahn (corresponding author)

Noah Oder

Benjamin Prud'homme

Catherine Régis

Henri Vilandré

To cite this paper:

Bengio, Y., Régis, C., Jahn, A., Prud'homme, B., Barry, J., Oder, N., & Vilandre, H. (2025). The Development of the UN Scientific Panel on AI (Policy Paper). Mila - Quebec AI Institute.

Acknowledgements

We extend our sincere gratitude to the following participants of a Mila workshop in Montréal on January 27th, 2025, whose valuable insights and perspectives have greatly enriched this position paper: Felix Creutzig, Mariano-Florentino (Tino) Cuéllar, Anna Karin Eneström, Rebecca Finlay, Kevin Kohler, AJung Moon, Daniel Privitera, Craig Ramlal, Stephen Toope, Lucia Velasco. We are very grateful for conversations and interviews with a wide range of stakeholders without whom this publication would not have been possible, including: Kari De Pryck, Jaakko Kuosmanen, Taylor Owen, Toby Wardman. All recommendations expressed in this publication reflect the opinions of the authors, and the authors only.

Cover photo by Mathias Reding

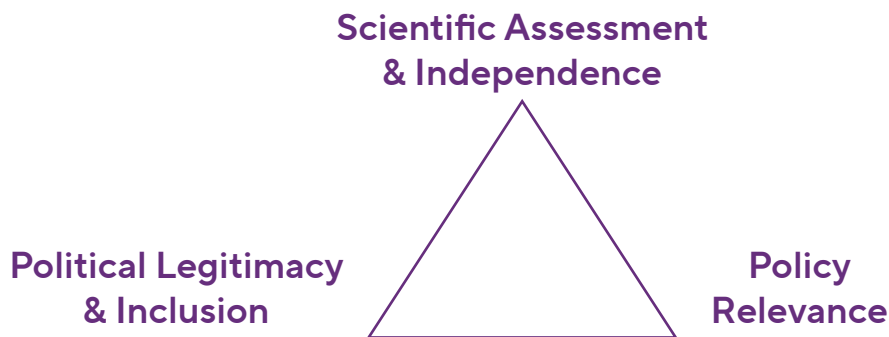
I. Context

In September 2024, the United Nations (UN) General Assembly adopted the Global Digital Compact (GDC), which includes a commitment to establish an Independent International Scientific Panel on Artificial Intelligence (AI) within the UN.¹ Accordingly, the President of the General Assembly appointed the Permanent Representatives of Costa Rica and Spain to the UN as co-facilitators for the process and consultations to identify the terms of reference and modalities for the establishment and functioning of the Panel and the Global Dialogue on AI Governance for adoption by the General Assembly.

On January 27, 2025, Mila - Québec AI Institute hosted a workshop to discuss design options for the proposed Panel (not the Global Dialogue). This paper builds on the results of this workshop, further integrating lessons learned from existing scientific panels and other literature on this subject.² It is our conviction that the analysis presented in this paper aligns the distinct challenges of ensuring effective, scientifically sound, and inclusive governance for this rapidly developing technology.

While the workshop was wide ranging, our recommendations build upon consensus around three guiding principles:

- **Scientific independence:** The Panel must build trust through the output of a scientific synthesis process that limits political interference or industry interests.
- **Political legitimacy and inclusivity:** The political legitimacy of the Panel’s reports is enhanced by the meaningful participation of Member States and other stakeholders. However, broadening inclusivity diminishes the Panel’s ability to deliver timely, tangible outputs. The Panel must strike a balance to maintain legitimacy and relevance.
- **Policy relevance:** The Panel’s policy relevance hinges on its ability to articulate the scientific evidence on AI targeted at policy makers. Scientists often prioritize rigor, whereas policymakers require timely, contextualized recommendations for action. Achieving a level of actionable, policy relevance will require robust and standardized ways to convey scientific consensus and uncertainty, as well as comprehensive, multidisciplinary analyses that speak to the socio-technical implications of AI.



¹Through the GDC, UN Member States committed to “establish, within the United Nations, a multidisciplinary Independent International Scientific Panel on AI with balanced geographic representation to promote scientific understanding through evidence-based impact, risk and opportunity assessments, drawing on existing national, regional and international initiatives and research networks.”

²Kohler, Kevin, et al. (2024). Blueprints: Design Options for the Independent International Scientific Panel on AI and the Global Dialogue on AI Governance. Simon Institute for Longterm Governance; Bengio, Yoshua,, et al. “International AI Safety Report.” arXiv, January 29, 2025. <https://doi.org/10.48550/arXiv.2501.17805>, UN High-level Advisory Body on AI’s: «Governing AI for Humanity» (September 2024), Carnegie Endowment for International Peace Organization’s «Proposal for an International Panel on Artificial Intelligence (AI) Safety (IP AIS)» (October 2023).

II. Lessons from International Scientific Panels Across Domains

The experiences of existing scientific panels provide valuable guidance for designing a UN Independent Scientific Panel on AI. This section focuses on three key cases: the [Intergovernmental Panel on Climate Change \(IPCC\)](#), the [EU's Scientific Advice Mechanism \(SAM\)](#), and the [International AI Safety Report](#). Each case study highlights critical trade-offs between scientific rigor, political legitimacy, and policy relevance, informing the design of an effective and independent AI panel.

1. The Intergovernmental Panel on Climate Change (IPCC)

The IPCC is one of the most widely recognised scientific panels, synthesizing climate science to inform international negotiations. It follows a structured, multi-tiered governance model, consisting of a bureau, multiple working groups, and a secretariat. Reports undergo extensive expert review and government consultations, ensuring scientific rigor and broad expert consensus. However, due to its intergovernmental nature, its approval processes are slow, complex, and consensus-driven, taking 5-7 years between assessments. This means that while the IPCC's work is highly credible and authoritative, it lags behind rapidly evolving scientific and policy debates.

Policy relevance is a further challenge. The IPCC strictly maintains policy neutrality, providing scientific findings but avoiding policy considerations. While this enhances its independence, it also means that governments must interpret its findings without structured guidance on actionable responses.

We regard the trade-off between scientific rigor and responsiveness as the key takeaway from the IPCC. While strong review mechanisms and broad consensus-building ensure credibility, slow processes and the absence of clear policy guidance limit the IPCC's ability to influence urgent decision-making. Finding ways to balance rigorous scientific assessments with the need for timely, actionable outputs is a central challenge for a scientific panel addressing fast-moving fields like AI.

2. The EU's Scientific Advice Mechanism (SAM)

The EU's SAM provides independent scientific input to the European Commission across a broad range of scientific topics. Unlike more narrowly focused panels, SAM can address any scientific question deemed relevant by policymakers, ensuring flexibility in responding to emerging challenges. It operates through a group of senior independent advisors, a large network of scientific academies and learned societies, and a small secretariat embedded within the Commission, which facilitates communication between scientists and policymakers.

A key feature of SAM is its demand-driven approach. It only works on topics requested by the Commission, ensuring strong policy relevance but limiting its ability to set its own research agenda. To enhance its impact, SAM employs a dual-reporting system, producing both a scientific evidence report and a policy report that translates findings into actionable insights. This layered approach strengthens the connection between scientific assessments and policy decisions.

We regard the effective use of a distributed scientific network and structured policymaker engagement as the key takeaway from SAM. By leveraging its network of European scientific institutions, SAM ensures multidisciplinary expertise and rigorous evidence synthesis, while its structured policy translation mechanisms make scientific findings more accessible and actionable. The trade-off here is between scientific independence and policy responsiveness: while SAM's model ensures direct political relevance, its reliance on policymaker requests may result in neglect of politically sensitive topics.

3. International AI Safety Report

The International AI Safety Report provides an independent, science-based assessment of AI risks and capabilities. It follows a layered model with multiple steps of drafting and review:

- An multidisciplinary writing group of technical experts drafts sections based on the guidance of the Chair;
- An multidisciplinary group of Senior Advisors selected by the Chair provides feedback focused on their areas of expertise at two intervals as the Report is drafted;
- An expert advisory panel composed of representatives from 30 nations, the EU, the Organisation for Economic Co-operation and Development (OECD) and the UN provides strategic guidance and feedback. The panel focuses exclusively on scientific synthesis, assessing AI risks and capabilities without engaging in policy recommendations;
- Outside experts, including from civil society and industry, are asked to provide feedback on an ad hoc basis throughout the drafting process.

This Report is scientifically independent, as the Chair has editorial control over the Report's content. However, this independence and the narrow subset of countries represented on the advisory panel may reduce its political legitimacy.

We regard the agile, timely and scientifically independent output of the International AI Safety Report as the key takeaway. This scientific independence enhances the Report's credibility, but limited political engagement can reduce its policy impact. Structured engagement with policymakers and greater inclusivity may enhance both the legitimacy and the influence of the Report. The challenge lies in ensuring broad representation and structured government engagement without compromising the independence and agility of scientific assessments.

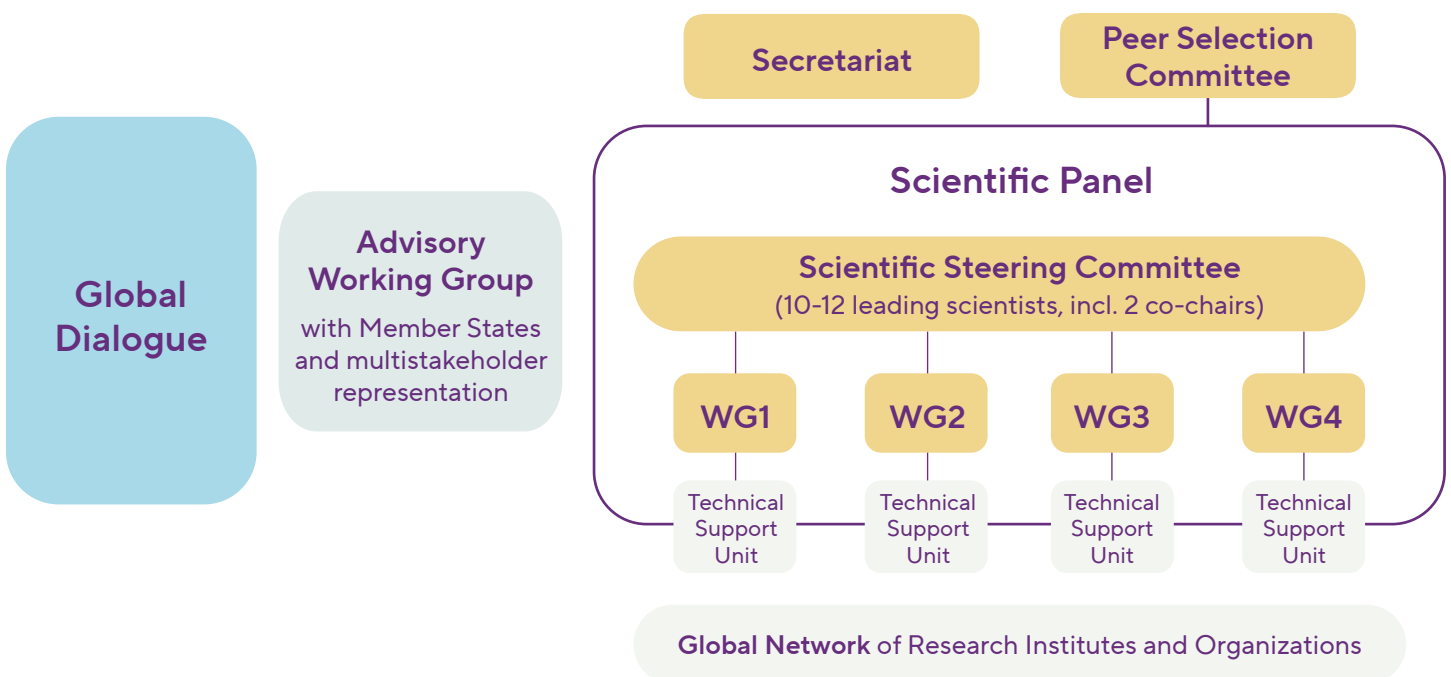
III. Recommendations for the Scientific Panel on AI

Recommendation 1: A Governance Structure that Ensures Scientific Rigor

AI governance requires a multidisciplinary approach that reflects the needs of Member States while balancing scientific independence and legitimacy. To strike that balance, **we recommend the following governance elements:**

- An independent and multidisciplinary Scientific Steering Committee of 10-12 members with 2 co-chairs;
- An independent Peer Selection Committee to select the scientific steering committee;
- A limited number of Working Groups focused on distinct AI-relevant subjects led by Vice-Chairs;
- A permanent Secretariat for operational, communication, and continuity purposes;
- An advisory Working Group composed of Member States to provide some oversight on policy relevance;
- Inclusive nominations through Member States and self-nominations;
- A focus on agile governance to allow the panel to follow and adapt to the fast pace of AI.

For a visualization of this structure, please see below.



Scientific Panel

We propose a Scientific Panel comprising two elements: a Scientific Steering Committee and a larger group composed of Working Group members, supported by Technical Support Units. The Scientific Steering Committee would form the core of the leadership and take much of the decision-making on the report, particularly those that affect the overall scientific direction, including decisions on Working Group report subjects. It would be composed of 10-12 leading scientists with multidisciplinary expertise. A dual chair structure would allow for representation from both the Global North and South, as well as different disciplines.

The larger Scientific Panel—which includes the members of the Working Groups—can convene to make decisions on the report’s multidisciplinary, division of labour, and other governance issues. Given their subject matter expertise, the Working Group Vice-Chairs (2 per Working Group) should have autonomy in hiring writers who meet a minimum qualification threshold.

We are also leaving out specific recommendations on Working Groups in the understanding that these are subject to change based on the pace and trajectory of AI. However, we do strongly recommend a Working Group focused specifically on **AI safety**. This Working Group should address issues related to frontier and advanced AI models and capabilities, particularly in light of the insufficient industry transparency regarding the advancement of AI capabilities and the risks they pose. This Working Group could draw heavily on existing scientific reporting in this area, notably the International AI Safety Report. Other Working Groups should consider covering subjects related to risks and opportunities related to SDGs (Sustainable Development Goals).

To preserve the report’s independence, we propose a separate committee to choose the members of the Scientific Steering Committee based on clearly defined criteria (academic impact, peer recognition, conflict-of-interest checks). This Peer Selection Committee would be made up of high-level multidisciplinary scientists who would not join the report in any other capacity. This would, as much as possible, provide credibility and independence to the selection process.

Nominations

We recommend an inclusive nomination process for the Scientific Panel. This would include opportunities for Member States to nominate potential participants, while also allowing self- or third party-nominations. Such a process would ensure access to the widest possible set of talent.

Secretariat

Both the IPCC and International AI Safety Report function with secretariats that provide continuity and connections to other relevant governing structures. The secretariats also fulfill other functions, including communications, legal tasks, budgeting, and coordination within the governance structure through meetings and formal consultations.

Advisory Working Group of Member States

To ensure Member States have robust input mechanisms to the Panel, we recommend establishing an Advisory Board comprising Member States (30-35 Member States with a focus on geographical representation through the 5 UN regions) with consideration for further membership for stakeholders through an accreditation process. As noted above, this Advisory Board would facilitate Scientific Panel nominations and requests for topics of study. Furthermore, this Board would be the connection between the Panel and the Global Dialogue on AI Governance, ensuring that the issues identified by the Dialogue can be independently studied in depth. Given that Member States shape and respond to the subjects studied by the Panel, this Board would also include a non-binding feedback mechanism on the reports.

Recommendation 2: Making Science More Relevant to Policymakers

The Panel can maximize its impact through direct relevance to policymakers. Therefore, **we recommend building policy relevance through three strategies:**

1. Provide a mechanism for Member States to communicate proposed areas of study;
2. Ensure reports are frequent and relevant, with a particular focus on quickly adapting to a changing AI landscape;
3. Provide clear policy options in a separate companion document to the annual report with applicability across UN membership.

AI will affect societies at an unprecedented speed and with an uncertain trajectory, leaving many governments unprepared for the potential impacts of AI.³ The UN is uniquely positioned to understand the needs of Member States, providing a lens on AI that differs from other organizations in the space. By providing relevant and timely information to Member States, the UN can bridge the preparedness gap on AI.

This **preparedness gap** is particularly acute for states that do not have high AI capacity or may have other policy priorities. In such cases, states may be unable to make policies from a policy-agnostic scientific report effectively. The International AI Safety Report is policy agnostic. Likewise, the IPCC strives to be ‘policy relevant but not policy prescriptive’ and has a relatively slow reporting cycle. These factors limit the IPCC’s policy efficacy.

Feedback Mechanism from Member States

While the outputs of the Panel must be insulated from undue political pressure, Member States should have a mechanism to communicate their needs. As such, **we recommend providing a formal mechanism for Member States to request areas of study.** Given the anticipated high volume of requests and resource constraints, these requests should not be binding. Working Groups should retain independent discretion over their research priorities.

³ We conceive of this as a “preparedness gap”, which represents the difference between the fast-paced development of AI and the current policy frameworks and strategies that governments have in place to address its challenges.

UN Reporting to Match the Speed of AI Progress

The speed of AI progress can hinder effective policymaking. Therefore, we recommend:

- **Releasing an annual report** with up-to-date information across disciplines;
- **Making the report a living document** with factual updates of some types of data between reporting periods;
- **Releasing focused reports** by Working Groups outside of the reporting cycle to respond to member state requests and address key subjects at a higher level of granularity than the annual report.

In the case of a well-funded, high-capacity secretariat, **we recommend:**

- Maintaining a live dashboard of key AI indicators with data ranging from investment by country, reported AI safety incidents by severity, data centre energy use, and other relevant data; leveraging existing trackers and key indicators by the global research network (e.g. OECD's incident reporting, African Responsible AI Observatory, Stanford HAI index, etc.)
- Proactively communicating across media to help all levels of governments understand and engage with findings.

These measures will allow for agility by the Panel to respond to changing conditions in the AI landscape. Feedback from the International AI Safety Report emphasized the difficulty in incorporating important advances in AI near and following the publication deadline. While these issues will persist under a resource-constrained organization, measures to provide timely updates to policymakers can support its policy relevance.

Policy Options, Not Policy Prescriptions

For states to act rapidly with informed policy, **we recommend including policy options** to complement the annual scientific report.

To maintain scientific independence, the policy options should not be drafted by the writers of the report or scientific Working Groups. Likewise, the policy options should be presented in a separate document or set of documents. We propose to convene a Working Group of AI policy experts to draft policy options from scientific outputs. This would leverage specific policy expertise while maintaining the group's independence.

Policy experts must not be prescriptive but instead offer a range of options that reflect the diverse needs of Member States, facilitating the adaptation of policies for local contexts.

While the Global Dialogue on AI Governance will likely deliberate on international governance policies like standards, funding to reduce the global AI divide, and human rights, the Panel's policy options are meant to be independent resources that States can take into consideration when establishing international or domestic policies.

Recommendation 3: Moderating Industry Participation

Scientific independence must be the foundation of the Panel’s credibility. Therefore, **we recommend moderating the participation of industry** through three main avenues:

1. Prohibit industry scientists from participation on the Panel or in the writing process of the reports;
2. Develop a conflict of interest policy with clear guidelines on disclosure for industry-affiliated scientists;
3. Build transparent avenues for industry feedback that balances independence with the necessity for information from companies.

Below, we will build on each of the sub-recommendations and outline the fundamental tension between leveraging industry participation and maintaining scientific independence.

Independence and Industry: A delicate balance

As outlined above, AI is unique in its speed and industry dominance. Following the advent of ChatGPT, industry has attracted a disproportionate amount of AI talent and vastly more capital and compute resources than academia—and this is growing.⁴ While published papers are estimated to hover around less than 50% with industry-affiliated authors,⁵ this undercounts the unpublished research within companies. Companies’ extensive funding opportunities also mean that even those who do not work directly for an AI company are likely to receive some type of industry funding for research.

AI companies also hold much of the data necessary to conduct timely AI safety research. These data include performance benchmarks for unreleased models, usage statistics, and deployment of new methods like chain-of-thought reasoning. This means that the public information available to academics lags behind the true capabilities of cutting-edge models, reducing the time policymakers have to react to new developments and associated risks.

Like any industry, AI companies have incentives to shape outcomes favourable to their interests. To preserve scientific independence, the Panel must strike a balance between leveraging the talents of leading scientists and collecting timely data while insulating itself from misaligned industry incentives.

Who is Industry?

As noted, many scientists have varying levels of industry affiliation. In order to attract the best talent, while maintaining scientific independence, **we recommend prohibiting those with a direct financial relationship to an AI company from substantive participation**. We define a direct financial relationship to include drawing a regular salary and ownership stakes in a company. This definition creates a clear barrier from both the perception and reality of industry influence. Those who have other types of industry participation—like fellowships or one-time consulting contracts—would not be prohibited. This is a fairly standard policy in independent science advisory bodies.

⁴ Ahmed, Nur, Muntasir Wahed, and Neil C. Thompson. “The Growing Influence of Industry in AI Research.” *Science* 379, no. 6635 (March 3, 2023): 884–86. <https://doi.org/10.1126/science.ade2420>.

⁵ Ibid.

Conflict of Interest Policy

All types of industry affiliation may erode the Panel’s objectivity and appearance of independence under certain circumstances. This necessitates a strong conflict of interest policy with transparent reporting and recusal mechanisms that may include:

- Disclosure of fellowships and consulting fees with AI companies;
- Disclosure of equity stakes and other investments in AI companies exceeding a significant threshold;
- Disclosure of gifts, offers of employment, and participation in industry events;

Engaging with Companies

As noted above, some level of industry participation supports the Panel’s scientific output, necessitating lines of communication with AI companies. This includes data collection, technical expertise, and participation as stakeholders who influence and are influenced by subjects under the Panel’s purview. Industry participation should be disclosed, structured, and inclusive to all relevant companies to reduce the perception of adverse influence. Forms of industry participation could include:

- Collecting aggregated data about advanced AI models through a trusted entity⁶ to allow scientists timely access to data without exposing companies’ intellectual property;
- Structured Expert Dialogues modeled after those by the IPCC to allow for transparent exchanges between scientists and industry with a clear public record;
- Sharing draft subsections of reports with industry experts at limited and consistent intervals for non-binding feedback as was done by the International AI Safety Report.

IV. Final Thoughts

In conclusion, the establishment of an Independent International Scientific Panel on Artificial Intelligence (AI) by the UN is a crucial step toward fostering an evidence-informed global and inclusive governance of AI. These recommendations aim to embed within the Scientific panel’s framework the necessary elements to uphold and balance scientific independence, legitimacy and policy relevance. Mila remains available to collaborate on further elaborating on the recommendations made, and support their implementation

⁶ For example, the Frontier Model Forum or the International Network of AI Safety Institutes.