AI Governance Day – From Principles to Implementation

2024 Report

Disclaimers

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of ITU concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The mention of specific companies or certain manufacturer products does not imply that they are endorsed or recommended by ITU in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by ITU to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader.

The opinions, findings and conclusions expressed in this publication do not necessarily reflect the views of ITU or its membership.

ISBN

Please consider the environment before printing this report.

AI Governance Day - From Principles to Implementation

2024 Report

Table of contents

List of figures and tables

Figures

Tables

AI Governance Day – From Principles to Implementation

0. Executive Summary

The first ever "**AI Governance Day**" subtitled "**From Principles to Implementation**" was held in Geneva, Switzerland, on 29 May 2024. Convened by the United Nations' International Telecommunication Union (ITU), the event brought together a kaleidoscope of participants from around the world. This included government leaders, policymakers, researchers, and technologists from both developed and developing countries. The event's multistakeholder composition aligned with ITU's mission to provide an inclusive, neutral, and globally representative platform for artificial intelligence (AI). The day was marked by vibrant discussions and collaborations aimed at transforming AI governance principles into actionable frameworks.

In the area of AI, the transition from principles to actionable governance is a challenge. These principles, while vital, have remained at a high level, often too abstract in their application for direct application in the ever-evolving AI landscape.

Recently, a significant shift has been observed, as regulatory bodies worldwide have begun to codify these principles into concrete regulations, creating foreseeable regulatory pressure on the development of AI. For example, China has instituted an Algorithm Registry in December 2022, and in October 2023, the President of the United States issued an "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence". The European Union AI Act has perhaps gone furthest: it is set to enter into force in the second half of 2024 and to become progressively enforceable over a 36-month period.

In the rapidly evolving landscape of AI, an "**AI governance paradox**" has emerged: while regulation is often seen as lagging behind technological advancements, there is an equally critical yet less recognized issue where the current state of technology and tools available does not allow for monitoring, checking, and controlling AI systems. This gap poses risks and underscores the need for advancing tools capable of ensuring effective governance.

"AI Governance Day" tackled the step of moving from regulatory frameworks to implementation. How are countries and regions navigating the dual objectives of maximizing AI's benefits while minimizing its risks? Participants shared experiences on what works, what does not work (yet), identified hurdles, and discussed what needs to happen next on the path towards effective regulatory implementation.

The first half of AI Governance Day – the morning sessions – were interactive and dynamic discussions among **200 senior leaders** from governments, the private sector, international organizations, academia, the technical community, and civil society.

Three themes were discussed:

- **Theme 1: What is the landscape of AI governance, and how will it evolve?**

  This report highlights the debate at the session about national vs. regional and international AI governance and includes a listing of the major multilateral as well as national initiatives as of the end of May 2024 – please see Chapter 4.3.1.

- **Theme 2: How do we implement AI governance frameworks?**

  The session discussed how technical standards wuold be an important piece in measuring whether AI governance frameworks are successfully implemented. As the report highlights, the current technical methods are often insufficient to measure the implementation of AI governance frameworks and to provide feedback. The report then goes deeper into "compute governance", arguably an area where measurements can be more straightforward and quantifiable as opposed to the governance of data or the governance of algorithms.

- **Theme 3: How do we ensure inclusion and trust?**

  Inclusion starts with connecting the unconnected. In 2023, according to ITU statistics, a third of the world's population is still not connected to the Internet. Furthermore, many people and many countries that are being impacted by AI are not integrated in ongoing discussions to set AI policies. The report also provides details on the discussion of open vs. closed sourcing of AI models.

The second half of AI Governance Day, the public afternoon sessions, built on the morning discussions and featured panel discussions.

- ITU's Secretary-General Doreen Bogdan-Martin recalled that the conundrum of how to govern a technology if we do not yet know its full potential is not new: twenty years ago, the Internet was met with a similar mix of shock, awe, and skepticism. But what we have learned from the WSIS (World Summit on the Information Society) process was that **we can take steps toward governance even if we are building the plane as we fly**. From the discussions in the morning, Ms Bogdan-Martin's learnings, based on the fundamental premise of including every nation and every stakeholder group in governance efforts, included three key pieces which must be part of any AI governance effort: (1) development of technical standards; (2) putting core UN values at the heart of AI governance; and (3) development through capacity building.

- In the session "**Leaders speak**: Insights and key findings on AI governance implementation", three Ministers, from Namibia, Colombia and Bangladesh, underscored the importance of a coordinated, inclusive, and human-centric approach to AI governance.

- "**State of play** of major global AI governance processes" featured the Council of Europe, the European Commission, USA, China, Japan (G7 Hiroshima process), and Republic of Korea (follow-up host of the UK AI Safety Summit in November 2023).

- "**Leveraging the UN system** to advance global AI governance efforts" spotlighted the Executive Heads of the International Labour Organization (ILO), the Food and Agriculture Organization (FAO), the World Intellectual Property Organization (WIPO), the United Nations University (UNU), and the ITU.

- In "The **government's AI dilemma**: How to maximize rewards while minimizing risks?", government leaders from Namibia, Uruguay and India stressed the benefits of AI such as in healthcare, education, digital literacy training, the need for ethical implementation and data privacy via a national AI strategy, but also the downsides such as the increase in cybercrime.

- "The critical conversation on **AI safety and risk**" saw researchers and company executives underscore the complexity and urgency of AI safety and risk management.

- The discussion in the session "**To share or not to share:** the dilemma of open-source vs proprietary large language models", hosting the Executive Director of the Linux Foundation and executives of Google, Meta, the Wikimedia Foundation and a policy researcher of the Future of Life Institute, emphasized that open source plays a critical role in fostering innovation, ensuring transparency and preventing market consolidation, while also recognizing the need for responsible governance to address risks.

- "Harmonizing high-tech: **The role of AI standards** as an implementation tool" welcomed the Executive Heads of the world's leading international standards organizations – ITU, the International Organization for Standardization (ISO), and the International Electrotechnical Commission (IEC). The three organizations have been coordinating and collaborating on standards development for decades. (They announced new initiatives on AI at the AI for Good Global Summit, following AI Governance Day.)

- In the closing session "**From principles to implementation – pathways forward**", the two co-chairs (from ITU and UNESCO) of the United Nations Interagency Working Group on AI, highlighted ongoing UN System-wide coordination efforts on AI, and the United Nation Systems White Paper on AI governance (May 2024).

We hope that the discussions and new connections formed will assist in moving AI governance forward from principles to implementation. We are looking forward to welcoming you at the next **AI for Good Global Summit (including AI Governance Day) in Geneva in the week of 7-11 July 2025**.

1.   Foreword by Doreen Bogdan-Martin, ITU Secretary-General

It was a tremendous pleasure to welcome a mullti-stakeholder group of senior leaders from governments, the private sector, international organizations, academia, the technical community, and civil society to the AI for Good Global Summit on 30-31 May 2024 and to AI Governance Day on 29 May 2024. This inclusivity was at the heart of our mission at the International Telecommunication Union (ITU) to create a neutral, global platform for artificial intelligence. AI Governance Day marks a truly global multistakeholder dialogue on AI governance, a topic that had gained significant importance over the past 12 months.

Discussions during AI Governance Day focused on three themes: the AI regulatory landscape, implementing AI governance frameworks, and ensuring inclusion and trust in these frameworks. The invitation-only morning sessions involved surveying the AI landscape to understand the current state of regulations and identify areas of common interest. An ITU survey revealed that 85% of our responding Member States had not yet established AI regulations, highlighting the urgent need for guidance and collaboration to avoid fragmented governance.

We also had a report from the United Nations High-Level Advisory Body on AI, with a live connection to co-chairs of the United Nations Secretary General's Advisory Body from their final in-person meeting in Singapore.

Discussions at AI Governance Day and at the [AI for Good Global Summit](#) set the stage for developing actionable, enforceable technical standards, essential for ensuring algorithm transparency, system safety, and security. The ITU had already published or was developing over 200 AI-related standards, but more are needed. Collaboration through mechanisms like the World Standards Cooperation, involving leaders from ISO and IEC, has been vital for advancing these standards globally.

Inclusion and trust are deeply interconnected and crucial for successful AI governance. Without trust, the adoption of AI would falter, potentially exacerbating digital divides. A survey by BCG highlighted that consumers in low and middle-income countries were more enthusiastic about AI than those in mature markets, viewing it as a means to leapfrog technological gaps and drive innovation in critical sectors like education and healthcare.

Reflecting on lessons from the World Summit on the Information Society (WSIS), we recognized the importance of a multi-stakeholder approach to governance. The WSIS process showed that governance could be iterative and inclusive, accommodating diverse perspectives and needs. This approach was deemed equally necessary for AI governance, ensuring that power was not concentrated in the hands of a few and that all stakeholders had a voice.

Our focus on human rights, inclusion, and capacity building was emphasized as paramount. It was essential that AI governance frameworks reflected core UN values and addressed the needs of all countries. The ITU's initiatives, in collaboration with UNDP and other partners, aimed to

upskill nations with low technological capabilities, helping them navigate the challenges and opportunities of AI.

I urge everyone to be ambitious and visionary. The goal is to make AI governance the cornerstone of a better digital future for all. Governance is a continuous process of listening, exchanging ideas, and implementing practical solutions. Together, we aimed to harness the power of AI for the greater good, ensuring it served as a force for positive global change.

I hope that the outcomes of the discussions held in Geneva will resonate beyond AI Governance Day and the AI for Good Global Summit.

Doreen Bogdan-Martin
ITU Secretary-General

Figure 1: Doreen Bogdan-Martin, Secretary-General, International Telecommunication Union (ITU)

2.   Preface by Prof. Robert Trager, moderator of AI Governance Day

At the 2023 edition of the AI for Good Global Summit, the ITU Secretary-General called for "guardrails to ensure AI is safe, responsible and inclusive."

As the speed of innovation races forward in AI, the transition from principles to actionable governance is a challenge. These principles, while vital, have predominantly remained at a high level, often too abstract in their application for direct application in the ever-evolving AI landscape.

Recently, a significant shift has been observed, as regulatory bodies worldwide have begun to codify these principles into concrete regulations, creating foreseeable regulatory pressure on the development of AI.

For example, China has instituted an Algorithm Registry, and in October 2023, the President of the United States issued an "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."

The European Union AI Act has perhaps gone furthest: it is set to enter into force in the second half of 2024 and to become progressively enforceable over a 36-month period.

What is AI Governance Day?

AI Governance Day spotlighted the status of worldwide efforts to shape the future of AI regulation – and hosted discussions on implementation strategies that are key to ensuring the AI race leaves no one behind.

It spanned a full day of interactive sessions, insightful discussions, and networking opportunities.

With a focus on fostering collaboration among prominent stakeholders, AI Governance Day provided a platform for government ministers, regulators, industry leaders, academicians, civil society, and UN representatives to engage in meaningful dialogue and chart the course for effective AI governance frameworks.

As the day opened, ministers and regulators tackled these questions:

- What is the landscape of AI governance, and how will it evolve?
- How do we implement AI governance frameworks?
- How do we ensure inclusion and trust?

Bringing AI governance discussions into focus

Afternoon sessions, open to the public, featured distinguished speakers and thought leaders, including high-level government officials to representatives from leading international organizations, reporting on the morning's outcomes and providing the state of play of major global AI governance processes.

Panel discussions included these topics:

- The government's AI dilemma: How to maximize rewards while minimizing risks?
- Leveraging the UN system to advance global AI governance efforts
- The critical conversation on AI safety and risk
- To share or not to share: the dilemma of open-source vs proprietary large language models
- Harmonizing high-tech: The role of AI standards as an implementation tool

We hope that the results of the discussions and the many new connections being made among the participants will assist in moving AI governance forward from principles to implementation.

Robert F. Trager, University of Oxford; Director, Oxford Martin AI governance initiative; senior research fellow, Blavatnik School of Government; International governance lead, centre for the governance of AI. Professor Trager moderated AI Governance Day.

Figure 2: Robert F. Trager, Professor, University of Oxford

3.   Format of AI Governance Day

3.1   Structure of AI Governance Day

AI Governance Day took place on Wednesday 29 May 2024 at the Geneva International Conference Center (CICG).

AI Governance Day was structured into a morning session from 9:00 to 13:00, followed by a networking lunch. The morning session was by invitation only and for those who purchased a Leaders Pass.

The morning session was focused, with 200 senior leaders engaging in the discussions. Chatham House Rule were used, i.e., participants we free to use the information received, but not the identity or the affiliation of the persons.

About two weeks prior to AI Governance Day, briefing notes prepared with the support of independent experts were distributed to the participants of the morning session of AI Governance Day. The briefing notes were structured according to the three themes discussed in the morning:

–   Theme 1: What is the landscape of AI governance and how will it evolve?
–   Theme 2: How do we implement AI governance frameworks?
–   Theme 3: How do we ensure inclusion and trust?

The text of the briefing notes has been incorporated into this report.

The afternoon session was public and live-streamed and featured panel discussions, available [here](#).

Master of Ceremony and Moderator of AI Governance Day was Professor Robert Trager, University of Oxford. He was assisted in the morning session by co-moderator Anka Reuel, a computer science PhD student at Stanford University whose research focuses on enabling organizations to develop and deploy responsible, socially beneficial AI systems.

Figure 3: Anka Reuel, PhD student of computer science, Stanford University, co-moderator

3.2   Run of show, morning session

– **Opening** by Professor Robert Trager, moderator
– **Welcome** by Doreen Bogdan-Martin, ITU Secretary-General
– **AI Governance pulse check – pre-discussion poll** (see chapter 4.7)
– **Multistakeholder panel: "How can we move AI governance from principles to implementation?"**
  Panelists:

  • H.E. Ms. Rose Pola Pricemou, Ministre, GUINEA (Ministère des Postes, des Télécommunications et de l'Economie Numérique)
  • H.E. Mr. Dohyun Kang, Vice Minister, Ministry of Science and ICT, Republic of Korea
  • Lane Dilg, Head of Strategic Partnerships, OpenAI
  • Stuart Russell, Professor of Computer Science, University of California, Berkeley
  • Rumman Chowdhury, CEO Humane Intelligence
  • Tristan Harris, Director, Executive Director of the Center for Humane Technology
  • Daisy McGregor, Deputy Director, International AI Policy, Department for Science, Innovation and Technology, United Kingdom

Moderator:

- Robert Trager, Professor, University of Oxford

– Update from the United Nations High-level Advisory Body on AI (HLAB)

- Carme Artigas, Co-chair of United Nations High-level Advisory Body on AI
- Amandeep Singh Gill, UN Secretary-General's Envoy on Technology

– **Breakouts.** Each table discussed one of three themes:

- Theme 1: What is the landscape of AI governance and how will it evolve?
- Theme 2: How do we implement AI governance frameworks?
- Theme 3: How do we ensure inclusion and trust?
- Format of breakout session:
    – 20 tables with 10 people each. Each table had
      - Ministers/Regulators
      - UN representatives
      - Industry leaders, Academia, Civil Society
      - Expert Advisor
      - Rapporteur (= Notetaker)
    – The ITU Secretariat made a seating plan which tried to balance geography and gender. Participants received their seating assignment as they came into the room.
    – Participants were kindly asked to write on cards at their tables what they consider noteworthy/useful/innovative. About 150 cards were received and incorporated into this report.

– Coffee break
– Insights from breakouts
  Moderators:

- Robert Trager, Professor, University of Oxford
- Anka Reuel, PhD student, Stanford University

– **AI Governance pulse check – post-discussion poll** (see Chapter 4.7) Same questions as in pre-discussion poll. Did opinions change?
– Closing
– Networking lunch

3.3   Run of show, afternoon session

– 14:00 – 14:10 Welcome address, Doreen Bogdan-Martin
– 14:10 – 14:30 Insights and key findings on AI Governance Implementation
– 14:30 – 15:15 State of play of major global AI Governance processes
– 15:15 – 15:45 Leveraging the UK system to advance global AI Governance efforts
– 16:00 – 16:30 The Government's AI dilemma: how to maximize rewards while minimizing risks?
– 16:30 – 17:00 The critical conversation on AI safety and risk
– 17:00 – 17:45 The dilemma of open source vs. proprietary Large Language Model
– 17:45 – 18:15 The role of AI standards as an implementation tool
– 18:15 – 18:25 From principles to implementation – pathways forward

Figure 4: Roundtable participants during the morning breakout session

Figure 5: Center Stage Welcome Address featuring Doreen Bogdan-Martin, Secretary-General, International Telecommunication Union (ITU)

Figure 6: The panel of speakers from a Center Stage discussion

4. Morning sessions

4.1 Welcome speech by Doreen Bogdan-Martin, ITU Secretary-General

It is a tremendous pleasure to welcome you all to our first ever Governance Day and a series of roundtables. This is kind of a different format for many of us. So, I am looking forward to having interactive, dynamic discussions today.

It is inspiring to see so many senior leaders with us this morning, early morning, and ready to roll up their sleeves. And of course, we have representatives from governments, the private sector, international organizations, academia, the technical community, and civil society, all here to discuss and help shape AI governance for the future.

I am particularly pleased to have so many representatives from both developed and developing countries. This representation aligns perfectly with ITU's ambition to provide the most inclusive, diverse, neutral, and global platform for artificial intelligence.

I am also pleased that this morning we will be linking up with Singapore. We will have a connection with the Secretary General's co-chairs of the United Nations High Level Advisory Body on AI (HLAB). They're meeting in Singapore this week, and we will hear from them live this morning. Their work, of course, is also very important as we look to the future and the global conversations around AI.

Excellencies, ladies and gentlemen, this morning is really a sort of ground zero for AI governance innovation. These discussions will help us to get a better grasp of the current AI landscape, who's

doing what, and where we are heading, how we can implement AI governance frameworks, and how we can ensure inclusion and trust are at the heart of all of our efforts.

These are the three topics that we're going to be discussing in the breakouts this morning, and I would like to briefly say a few words about each.

The AI regulatory landscape

First, from the AI regulatory landscape. We have seen quite a fast response, in particular, over the past year, with governments activating regional groups on the global level as well, starting to discuss this issue. But I think many are still contemplating what to do. The ITU actually conducted a landscaping survey among our 193 Member States, and we found that about 85% of our responding Member States hadn't yet put in place regulations or policies. Some are just beginning to think about these issues. But I think one important result of the survey was that it demonstrated that all countries are eager to learn. And so that's also part of the objective of the discussions this morning. Through dialogue, we can help to avoid fragmentation when it comes to AI governance.

Collaboration around areas of common interest and priorities comes next. A good example is standards, and we're joined this morning by many of our standards friends. Standards are definitely central to this debate and, of course, to the entire summit. It's critical for all AI governance initiatives.

How to implement AI governance frameworks?

That brings me to my second point, which is how to implement AI governance frameworks. From algorithm transparency to safety and security of systems, standards really serve as a prerequisite for the effective implementation of guardrails. Simply put – and I look to Gabriela [Ramos] from UNESCO – the best ethical or human rights guidelines would be incomplete without being translated into actionable, enforceable technical standards. ITU already has more than 200 AI-related standards, either published or currently under development. Our standardization process ensures that all voices are heard, including those from the developing world. But more standards are needed to address the pressing challenges around artificial intelligence. And, more importantly, we need to be developing them in a coordinated way, using established mechanisms like the World Standards Cooperation. I recognize the leaders of ISO and IEC who are with us this morning.

Inclusion and trust

That's a natural segue to the third topic this morning, which is inclusion and trust. These two elements are deeply interconnected. Without trust, people will hesitate to engage with AI, potentially creating yet another digital divide in an already unequal digital world. I think this risk is real. For the first time this year, adverse outcomes of AI actually entered into the top 10 rankings of the World Economic Forum Global Risks Perception Survey. It's important to note that not everyone feels the same way about AI. A recent survey carried out by BCG revealed that consumers in low and middle-income countries were actually much more excited than consumers in mature markets. Many see this as an opportunity, an opportunity to leapfrog technological gaps and accelerate innovation in vital areas such as education and healthcare, and I would say all the SDGs.

Excellencies, ladies and gentlemen, the outcomes of our discussions today will resonate throughout this summit and beyond. I ask you this morning and throughout the day, to be ambitious. Let's use AI to tackle our most pressing global challenges and achieve the SDGs. I would also ask you to be visionary. Let's make AI governance the cornerstone for a better digital future for all. But, above all, I dare you to have hope. These are not ordinary times. These are times when we need to show audacity, solidarity, and a shared sense of responsibility.

Thank you very much. It's a really a great pleasure to welcome you all to our governance day and a series of roundtables. This is kind of a different format for many of us. So looking forward to having lots of interactive dynamic discussions today.

4.2 How can we move AI governance from principles to implementation? – Multistakeholder panel

Panelists:

- H.E. Ms. Rose Pola Pricemou, Ministre, Guinea (Ministère des Postes, des Télécommunications et de l'Economie Numérique)
- H.E. Mr. Dohyun Kang, Vice Minister, Ministry of Science and ICT, Republic of Korea
- Lane Dilg, Head of Strategic Partnerships, OpenAI
- Stuart Russell, Professor of Computer Science, University of California, Berkeley
- Rumman Chowdhury, CEO Humane Intelligence
- Tristan Harris, Director, Executive Director of the Center for Humane Technology
- Daisy McGregor, Deputy Director, International AI Policy, Department for Science, Innovation and Technology, United Kingdom

Moderator:

- Robert Trager, Professor, University of Oxford

Figure 7: Speakers from the Multistakeholder Panel: How can we move AI Governance from Principles to Implementation?

Investment in AI science

- There is an urgent need for increased investment in AI science, surpassing the funding allocated to other scientific fields. The potential for AI to surpass human intelligence through Artificial General Intelligence requires preemptive measures to maintain control over these powerful entities before their widespread deployment.

Regulatory frameworks and safety measures

- Historical precedents, i.e the uncontrolled risks of nuclear power illustrated by the Chernobyl disaster, underscore the importance of rigorous safety measures and oversight in AI development.
- AI applications must be restricted to sectors where safety can be unequivocally ensured, similarly to the stringent safety requirements in the pharmaceutical and nuclear industries. The principle of proving safety before deployment should be paramount.
- Current international and domestic policy frameworks are inadequate to address the rapidly evolving AI landscape. There is a need for new institutions and regulatory tools to manage AI risks effectively.

International cooperation and summits

- Past and upcoming international summits on AI governance and safety in the UK (2023), Korea (2024), and France (2025) are important in establishing global standards and interoperability of AI safety measures.
- These summits aim to foster cooperation among heads of state, government leaders, UN agencies, and other stakeholders to address the AI divide, promote inclusivity, and ensure that safety measures are integrated into AI development.

Figure 8: H.E. Ms. Rose Pola Pricemou, Ministre, Guinea (Ministère des Postes, des Télécommunications et de l'Economie Numérique) during the Multistakeholder Panel

Figure 9: Rumman Chowdhury, CEO, Humane Intelligence, during the Multistakeholder Panel

Challenges in AI governance

- One of the primary challenges in AI governance is the nascent understanding of AI models and their implications. Balancing the benefits and risks of AI requires a deep understanding of these technologies and their potential impacts.
- Interoperability at the international level is crucial to enable small companies to scale across borders and participate in the global AI ecosystem. This requires detailed workstreams and ongoing collaboration among countries.

Risks of the AI race

- The competitive race among AI companies often prioritizes market dominance over safety, leading to the use of unlawful shortcuts. Negative consequences include the development of AI models that can hack into potential safety limits implemented.
- Governance frameworks must shift the incentives towards safety and responsible development to mitigate these risks. The historical failure to anticipate the externalities of social media serves as a cautionary example.

Preparedness and science-based governance

- Effective AI governance must be rooted in scientific principles, with preparedness work based on reliable scientific data. This includes leading in safety and capability while ensuring models are not prematurely released to the public.
- There is a need for standardized evaluations and methodologies to assess AI models, considering their probabilistic outputs and the challenges in creating consistent tests.

Inclusivity and capacity building

- AI governance must address the specific needs of developing countries, ensuring that AI solutions are tailored to local realities and cultural specifics. Inclusivity, for instance of young women in African societies, is vital for equitable AI development
- International cooperation on standards, pooling financing resources, and avoiding fragmented efforts are essential for effective capacity building and governance.

Education and transparency

- Transparency in AI interactions is critical, including the right to know whether one is interacting with a human or a machine. The burden of proving the safe and ethical use of AI should rest on developers.
- Education on AI technologies and their implications should be a priority to ensure that society is adequately prepared for the changes brought by AI.

Moving from summits to action

- Transitioning from discussions at summits to actionable plans through AI Safety Institutes and other mechanisms is crucial for implementing effective AI governance. Emphasizing the definition and regulation of AI safety is necessary to address the risks associated with AI technologies.
- Global digital compacts and resolutions, such as the WTSA Resolution for the ITU to coordinate standards efforts within the UN, are steps towards cohesive international governance frameworks.

Figure 10: Tristan Harris, Director, Executive Director of the Center for Humane Technology, speaking during the Multistakeholder Panel

Figure 11: Stuart Russell, Professor of Computer Science, University of California, Berkeley, speaking during the Multistakeholder Panel

Figure 12: Lane Dilg, Head of Strategic Partnerships, OpenAI, speaking during the Multistakeholder Panel

4.3 Theme 1: What is the landscape of AI governance and how will it evolve?

4.3.1 Examples of multilateral and national AI governance frameworks

Governments around the world are starting to establish rules with respect to AI. Here is a list of multilateral and national initiatives as of the end of May 2024 (the list is indicative and non-exhaustive).

Table 1: Examples of national AI Governance initiatives

Examples of National AI Governance Initiatives	
1. Brazil	AI Strategy (April 2021), Draft AI Bill (proceeding through Parliament May 2024)
2. China	Internet Information Service Algorithmic Recommendation Management Provisions (March 2022), Provisions on the Administration of Deep Synthesis Internet Information Services (November 2022), Measures for the Management of Generative AI Services (July 2023), Global AI Governance Initiative (Oct 2023)
3. Estonia	National AI Strategy (2022-23)
4. France	National AI Strategy (April 2024)
5. India	National Strategy for AI (June 2018) Principles for Responsible AI (Feb 2021)
6. Japan	AI Guidelines for Business (April 2024)
7. Kazakhstan	Digital Transformation Partnership (Oct 2023), AI in Government Services (Feb 2024)
8. Kenya	National AI Strategy (April 2024)
9. Republic of Korea	National Strategy for AI (Oct 2019)
10. Rwanda	National AI Policy (April 2023)
11. Singapore	Proposed Model AI Governance Framework for Generative AI (Jan 2024), AI Verify Platform (June 2023)
12. UAE	National Strategy for AI (Oct 2017)
13. UK	A Pro-Innovation Approach to AI (August 2023), Inspect Platform on AI Safety (May 2024)
14. USA	Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI (Oct 2023)

Table 2: Examples of multilateral AI Governance initiatives

Examples of Multilateral AI Governance Initiatives	
1. African Union	Multistakeholder consultations on a Continental Strategy on AI (April 2024), AUDA-NEPAD White Paper: Regulation and Responsible Adoption of AI in Africa (June 2023)
2. AI Safety Summit	The Bletchley Declaration on AI Safety (Nov 2023) (UK host with follow-up by the Republic of Korea and France), Seoul Declaration (May 2024), France-China Joint Declaration on AI Governance (May 2024)
3. ASEAN	ASEAN Guidance on AI Governance and Ethics (Feb 2024)
4. BRICS	Agreement to establish a BRICS AI Study Group (Aug 2023)
5. Council of Europe	Framework Convention on AI, Human Rights, Democracy and the Rule of Law (17 May 2024)
6. Digital Cooperation Organization	Generative AI Shaping the Digital Economy (Jan 2024)
7. European Union	EU AI ACT (April 2024 version of text)
8. EU-US Trade and Technology Council	Joint Statement (April 2024) Joint Roadmap (Dec 2022)
9. Latin America and the Caribbean	The Santiago Declaration to Promote Ethical AI (Oct 2023)
10. G7 Hiroshima Process	G7 International Guiding Principles and Code of Conduct for Organizations Developing Advanced AI Systems (Oct 2023)
11. G20	Endorsement of OECD Recommendations on AI (June 2019) and G20 AI principles (June 2019) CAIPD AI Statements (Dec 2023)
12. GCC	Ethical Framework for AI in the work of Attorneys-General and Public Prosecutors' (Oct 2023)
13. Global Partnership on AI	Multistakeholder Initiative hosted by OECD (June 2020) Ministerial declaration (December 2023)
14. ITU	AI for Good Summit 2024: AI Governance Day (May 2024)
15. ISO	Two key AI standards agreed. AI Management system (Dec 2023), AI Guidance On Risk Management (Feb 2023)
16. IEEE	Recommended Practice For Organizational Governance Of AI (Feb 2020), Algorithmic Bias Considerations (Sept 2023)
17. League of Arab States	Arab AI Working Group to Develop an Arab AI Strategy (Feb 2021)
18. Nordic Council of Ministers	Declaration on AI in the Nordic-Baltic Region (May 2018)
19. Southern Africa	Windhoek Statement on AI in Southern Africa (Sept 2022)
20. OECD	Recommendations on AI (May 2019), OECD AI Policy Observatory (Feb 2020), OECD Framework For The Classification Of AI Systems (Feb 2022), AI Language Models Technological, Socio-Economic And Policy Considerations (April 2023)
21. MERCOSUR	Declaration on Principles of Human Rights in the field of AI (Nov 2023)
22. Responsible AI in the Military Domain Summit (REAIM)	Call to Action (Netherlands and Republic of Korea led process) Political Declaration (US initiative) (Feb 2023)
23. UNESCO	UNESCO Recommendation on Ethics of AI (Nov 2021)
24. UN High-Level Advisory Board on AI	Interim Report on Governing AI for Humanity (Dec 2023)
25. UN Inter-Agency Working Group on AI	Terms of Reference (March 2021), UN System White Paper in AI Governance (May 2024)
26. UN General Assembly	Resolution A/78/256 Seizing the opportunities of safe, secure and trustworthy AI systems for sustainable development (March 2024)
27. UN Security Council	High-level debate on AI convened by UK (July 2023), Arria-formula meeting on AI, peace and security convened by UAE and Albania (Dec 2023)
28. UN Human Rights Council	Resolution 42/15 The Right to Privacy in the Digital Age (7 Oct 2019)
29. WHO	Ethics and governance of AI for health Guidance on large multi-modal models (Jan 2024)

4.3.2   The debate about national vs international AI governance

As AI regulations start to take shape, questions arise as to which areas of AI governance should be addressed at the national or regional level, and which at the international level.

Items to be regulated internationally might include:

- AI in warfare
- AI and Human Rights: with due consideration given to the guidance by UNESCO and the Human Rights Council in this regard.
- Data privacy and cross-border data flows
- Interoperability and standards: The International Telecommunication Union (ITU), International Organization for Standardization (ISO), and International Electrotechnical Commission (IEC) are collaborating through the World Standards Cooperation (WSC) framework to develop international AI standards. The ITU has published over 100 AI-related standards, with 120 more in development as of 2024, multiple of them in collaboration with other UN agencies. ISO and IEC have formed the joint subcommittee ISO/IEC JTC 1/SC 42 to advance AI standardization, developing foundational standards, reference architectures, frameworks, and guidelines for trustworthy AI systems.
- Shared research resources, in particular on AI Safety. Some countries have set up, or are considering setting up, an AI Safety Institute, for example:

  - United Kingdom: The UK has established the AI Safety Institute, which is the "first state-backed organisation focused on advanced AI safety for the public interest." The UK government published on 18 May 2024 an up-to-date, evidence-based International Scientific Report on the Safety of Advanced AI.
  - United States: The US has created an AI Safety Institute within the National Institute of Standards and Technology (NIST), following the Executive Order on AI.
  - Singapore: Singapore has established a "Generative AI Evaluation Sandbox" to bring together industry, academic, and non-profit actors to evaluate AI capabilities and risks.
  - Canada: The Canadian government has included in its 2024 budget funds to create an AI Safety Institute of Canada to ensure the safe development and deployment of AI.
  - Japan established an AI safety institute in February 2024.
  - The European AI office called for in the European AI Act will cover AI safety but has a wider scope than just safety and also include research, innovation, deployment aspects, and international engagement.

Items to be regulated nationally might include:

- Sector-specific AI applications: National regulations, reflecting local needs, values, and legal systems, can address the deployment of AI in sectors such as healthcare, transport, finance, manufacturing, human resources, critical infrastructure (gas, water, electricity), law enforcement, administration of justice, education, national employment.
- Consumer protection

There may also be a hybrid approach, where international guidelines provide a broad framework while allowing for national or regional specificity.

- Intellectual Property Rights (IPR) in AI
- Interoperability and standards
- Environmental sustainability

The United Nations System published on 2 May 2024 a <u>UN System White Paper in AI Governance: An analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance</u>.

<u>Potential discussion questions</u>

- What are the learnings that can be drawn multilateral and national AI governance efforts?
- Which AI standards and regulations need to be agreed upon at the international level?
- Within each category, with which priority should the various items be tackled?
- Which process should be used to determine the prioritization of the various items?
- What role should the UN take in international AI governance efforts?
- How should participation in international AI governance decision-making be structured?
- How do we address cross-border risks of AI systems?
- How can we safely share knowledge about what is working and what is not working well?
- How can regulation adapt and keep pace with technological advancements?

4.3.3   Insights from the breakout sessions: theme 1

The landscape of AI governance is intricate and evolving, with various approaches and development stages across countries and regions. Key themes include human-centric development, leveraging existing frameworks, inclusion, global coordination, private sector involvement, and balancing governance with regulation.

- **Human-centric AI development**: AI development should prioritize human welfare and societal betterment, focusing on ethical principles and social good.
- **Leveraging existing frameworks**: using existing regulatory frameworks from industries like automotive, pharmaceuticals, and cybersecurity can streamline AI governance. International organizations like the UN can provide unified principles to prevent fragmentation.
- **Inclusivity and capacity building**: addressing biases in AI and ensuring inclusive models are essential. Enhancing AI capabilities in underrepresented regions, particularly in the Global South, involves improving data collection and usability.
- **Multilateral and coordinated efforts**: global coordination is crucial to prevent big tech companies from setting standards. A multilateral approach is more effective than regional or national efforts alone. Reducing fragmentation and consolidating efforts within strong institutions is vital.
- **National and local applicability**: implementing international AI governance frameworks locally is challenging due to different adoption levels and needs. Smaller countries need support to build necessary institutions.
- **Private sector involvement**: private companies, especially in healthcare, must incorporate value-based AI governance in their operations. The private sector is key in ethical AI development.
- **Governance versus regulation**: governance encompasses broader objectives than regulation. Initiatives by UNESCO and ITU on ethics and standards are important. Regulations must adapt quickly to keep pace with AI innovations.
- **Addressing linguistic and cultural divides**: global frameworks must be balanced with local contextualization to ensure inclusive AI development. Resolving linguistic and cultural divides is crucial.
- **Data governance**: data governance is fundamental in AI governance to prevent digital colonization. Ensuring quality and accessible data and establishing cross-sector standards are necessary.

- **Cross-sector collaboration and standards**: interoperability is essential for AI governance, ensuring sectors can keep up with advancements. Regional digital regulatory sandboxes allow collaborative testing and refining of AI frameworks.

- **Current Landscape**

  - Early stage regulation: many countries are just beginning to develop AI governance frameworks, with regulation often driven by companies rather than comprehensive policies. Global leadership and inclusive representation are needed.
  - Country-specific approaches: countries take unique approaches based on their needs and development stages. The EU's AI Act sets a significant example. Balancing innovation with regulation is a critical challenge.
  - Inclusivity in regulation: every nation must contribute to inclusive AI regulations to prevent dominance by a few powerful countries or corporations. AI governance must consider cultural, ethical, and religious values.
  - Global and local balance: a combination of global principles and local adaptations is needed. Embedding technical language and ethical considerations into policies is essential.

- **Future evolution**

  - Global coordination: international bodies like the UN are expected to establish global AI governance documents, balancing standards with local regulations. Ensuring developing countries participate in AI advancements is critical.
  - Comprehensive frameworks: learning from existing frameworks, such as nuclear regulation, can help create robust AI governance structures. AI literacy should include understanding AI's implications, ethics, and governance.
  - Dynamic and adaptable regulations: regulations must be dynamic and adaptable to keep pace with AI innovations. A blend of global standards and local adaptations will ensure inclusive and equitable access to AI.

- **Intersection of civil society and industry**

  - Government lag and civil society's role: governments often lag in adapting to AI, with civil society remaining reactive. Fragmented approaches lack integration with data governance and cybersecurity.
  - Bottom-up vs. top-down approaches: bottom-up approaches risk duplicating efforts, while top-down approaches may lack detailed roadmaps. Establishing clear definitions for robustness and safety is crucial.
  - Geopolitical approaches: different regions have distinct AI governance approaches: the EU focuses on rights, China on economic development, and the US on maintaining leadership. Identifying applications needing strict regulations is essential.

- **Learnings from multilateral and national efforts**

  - Variance between countries: advanced countries have varying AI strategies, presenting challenges for Least Developed Countries (LDCs). Startups in developing countries often adopt AI rapidly without sufficient scrutiny. Standardized benchmarks can guide AI adoption.
  - Cross-border risks and regulation absence: lack of regulation across borders presents risks. Governance of high-risk AI applications, like Generative AI, is crucial. Existing standards from organizations like WHO and ISO can provide resources.

- **Governance models and international cooperation**

  - UN values and principles: existing legal instruments for AI regulation should be leveraged. Increasing global awareness and establishing regional AI innovation centers are essential.

- **Government roles and responsibilities**
    - Cabinet-level oversight: there is debate on whether AI should have dedicated oversight or be integrated into existing frameworks. Ensuring government responsibility for AI harms and mandating transparency is critical.
    - Cybersecurity and AI: the growth of cyberattacks underscores AI's role as a defense system. Addressing severe, state-sponsored attacks and determining responsibility is necessary.

4.4 Theme 2: how do we implement AI governance frameworks?

4.4.1 Lagging laws, lagging tech: the AI governance paradox

In the rapidly evolving landscape of AI, a paradox has emerged: while regulation is often seen as lagging behind technological advancements, there is an equally critical yet less recognized issue where technology fails to keep up with regulatory demands. While existing regulatory frameworks struggle to adapt to the pace of AI innovation, simultaneously, the current state of technology and tools available does not allow for monitoring, checking, and controlling AI systems. This gap poses risks and underscores the need for advancing tools capable of ensuring effective governance.

AI governance includes the governance of data, the governance of algorithms, and the governance of computing resources (compute).

- **Governance of data**: The governance of data refers to the policies, procedures, and standards necessary to manage the lifecycle of data within AI systems. Data governance ensures data quality, integrity, and security, which are essential for the reliable operation of AI technologies. This includes establishing robust protocols for data collection, storage, and sharing, and implementing privacy and security measures to protect sensitive information. Data governance frameworks must address issues such as consent, transparency, and accountability to maintain public trust and comply with regulatory requirements. Moreover, data governance plays a crucial role in mitigating biases in AI systems by promoting diverse and representative datasets, thereby enhancing the fairness and accuracy of AI outcomes.
- **Governance of algorithms**: The governance of algorithms focuses on the ethical and responsible development, deployment, and oversight of AI models and their decision-making processes. This aspect of AI governance aims to ensure that algorithms operate transparently, fairly, and without discrimination. It involves creating standards and guidelines for algorithmic accountability, which include regular audits, performance evaluations, and the ability to explain AI decisions to stakeholders. Algorithmic governance also emphasizes the importance of ethical considerations, such as avoiding unintended harmful consequences and ensuring that AI applications align with societal values.
- **Governance of compute**: The governance of compute addresses the management and oversight of computational resources required to develop, train, and deploy AI systems. As AI models become increasingly complex and resource-intensive, the need for sustainable and equitable access to computing power grows. Compute governance involves setting policies for the efficient and fair allocation of computational resources, ensuring that these resources are used responsibly and do not disproportionately favor well-resourced entities over smaller or less funded organizations. Additionally, it includes considerations for environmental sustainability, as the energy consumption of AI training and operations has a significant ecological impact. By implementing strategies to optimize energy use and reduce carbon footprints, compute governance aims to balance the advancement of AI with the imperative of environmental stewardship.

Where technology lags behind AI governance

In the realm of AI governance, several areas reveal a significant lag in technological advancements needed to meet regulatory requirements. These gaps span across data management, computational oversight, model integrity, and deployment practices, highlighting the urgent need for innovation and robust methodologies to support effective governance.

- **Data**
  - Methods are lacking to clearly define and implement fairness in AI systems.
  - There are no robust ways to fully safeguard privacy, especially in large language models.
  - It is difficult for users to know if their data was used to train a model and its influence on model behavior.

- **Compute**
  - There are only nascent methods for verifying if a large model is being trained in order to enforce compute-based reporting requirements.

- **Model**
  - Comprehensive and reliable methods to evaluate AI systems for safety, ethics, and reliability are lacking.
  - Securing model weights from unauthorized access is challenging.

- **Deployment**
  - Approaches to constrain model outputs to certain values are imperfect and can be circumvented.
  - Detecting AI-generated content, especially non-watermarked content, remains difficult.

4.4.2 Compute governance

"Compute" or "computing power" (often used interchangeably) refers to the necessary computational resources required to run software. Compute governance is arguably an area where measurements can be more straightforward and quantifiable.

The training process of AI models requires intensive compute, i.e. computing resources. and training a leading AI model tends to take months on a complex computing infrastructure involving specialized computer chips. In contrast, running a single inference query (e.g. having an AI model respond to a single question) requires much less compute, but the total amount of compute used for inference is still very large, since large AI companies need to run millions of user queries per day.

Compute is essential not only for training AI models, but also for deploying (i.e. operating) them. Just as operating expenses outpace initial fixed costs for many large-scale projects, the majority of available AI compute resources are used for operating AI models rather than training them. Frontier AI models are so large that they cannot be efficiently operated at large-scale with household amounts of typical consumer hardware. Instead, for models in high demand, inference requires thousands of AI chips housed in specialized data centers to adequately serve the needs of thousands of users. The wider the deployment of AI systems (which requires more compute), the more impact they will likely have (both beneficial and harmful).

"Compute governance", the setting of rules on computing resources to achieve governance, can be an attractive tool for AI governance. This is because compute is detectable and quantifiable, allowing for effective monitoring and control. For example, energy-intensive, specialized data center infrastructure is an indicator of compute activity. In contrast, while data and algorithms are also essential ingredients of AI, it is much more challenging for governments to quantify them.

Many are of the opinion that using compute providers (e.g. Microsoft Azure, Amazon Web Services (AWS), Apple, Bytedance, Meta, Oracle, Tencent, and Google Cloud) as intermediary regulators would be most effective in addressing risks associated with large-scale AI training to prevent bad actors from training advanced AI models, rather than addressing all AI-related risks. This is because non-compute-intensive AI models are often feasible to train and run on widely available customer hardware, so cloud providers have less ability to oversee such activities.

Compute providers can therefore play an essential role in AI governance via four key functions:

- Securers: protecting AI systems and critical infrastructure
- Record keepers: improving transparency for regulators
- Verifiers: monitoring customer activities
- Enforcers: taking actions against breaches of rules

International cooperation is essential to handle cross-border supervision and data challenges (e.g. ensuring that personal data is protected according to different regional standards and regulations), as it reduces the risk of compute providers and AI developers moving to jurisdictions with less regulatory oversight.

In addition to its potential role in regulation, compute has the potential to advance international cooperation on AI, by enabling states and companies to demonstrate their adherence to their commitments without leaking sensitive data. States may be able to show that approximately all of their AI compute was used consistently with their commitments, meaning significant compute would not have been available for other purposes. These approaches could leverage Privacy-Enhancing Technologies (PETs) to enable assurance while preserving confidential data.

Potential discussion questions

- How does compute governance differ from data governance and algorithm governance?
- Are there real-world examples of effective compute governance?
- How can compute resources be effectively monitored and controlled to ensure compliance with governance policies?
- How can compute providers improve transparency for regulators and stakeholders?
- What are the potential frameworks for international cooperation on compute governance?
- What are the potential risks of over-regulation, and how can they be mitigated?
- How might compute governance evolve with advancements in AI and computing technologies?

4.4.3   Insights from the breakout sessions: theme 2

Implementing an AI governance framework involves addressing challenges across data, compute, models, and deployment. Here is a direct and concise approach summarizing the key points and elements discussed to implement an adequate framework.

- **Inclusive and Collaborative Governance**

  – Stakeholder Involvement: Include government, private sector, NGOs, scientists, and academia; create an international AI Governance Agency with an advisory board of experts and citizens.
  – Education and Legislation: Integrate AI education into schools; develop AI laws and policies at national and international levels.

- **Regulation and Enforcement**

  – Global Standards: Cooperate internationally to set common standards; establish national institutes for ongoing regulatory discussions.
  – Decentralized Decision-Making: Empower specific units for quicker regulatory decisions.

- **Data Governance**

  – Fairness: Develop and integrate standardized fairness metrics.
  – Privacy: Implement differential privacy and federated learning; set regulatory standards for privacy.
  – Transparency: Mandate clear data usage policies.

- **Compute Governance**

  – Verification**:** Create reporting standards for compute usage.
  – Monitoring**:** Use compute providers for real-time monitoring and reporting.

- **Model Governance**

  – Evaluation**:** Standardize safety, ethics, and reliability evaluations; establish third-party certification bodies.
  – Security**:** Implement encryption and secure access protocols for models.

- **Deployment Governance**

  – Output Control: Develop systems to monitor and constrain harmful outputs.
  – Detection: Invest in tools to detect AI-generated content.

- **Interoperability and Common Definitions**

  – Technical Standards: Ensure systems are interoperable globally; establish common definitions for software, hardware, data, and resources.
  – Enforcement Models: Create audit models for balance between innovation and control.

- **Practical Implementation and Testing**

  – Testing Tools: Develop practical testing tools for AI systems; propose certification frameworks.
  – Continuous Testing: Emphasize ongoing testing for AI safety and efficacy.

- **Monitoring and Enforcement**

  – Frameworks: Develop interoperable assessment frameworks considering diverse regulatory capacities and application levels.

- **Standards and Global Processes**

  – Standards Organizations: Leverage existing standards organizations like IEC, ISO, and ITU; use established processes to develop AI standards.

- **Balancing Risks and Opportunities**

  – Risk Management: Balance AI risks and opportunities, increase investment in AI safety research.
  – Proactive Measures: Ensure AI safety before deployment; involve government to balance safety and competition.

- **Education**

  – Investment: Invest in AI education and mental health; promote public understanding and ethical awareness of AI.

- **Governance Pace**

  – Acceleration: Speed up AI governance to match technological advances; establish liability frameworks and encourage whistleblowers.
  – Adaptability: Develop adaptable legislative frameworks; encourage open-source tools for AI testing and safety.

- **Transparency and Accountability**

  – User Awareness: Ensure users know when they interact with AI; place proof of safety on AI developers.
  – Incentives: Link AI governance to incentives; include diverse perspectives in AI development.

- **AI Safety Institute**

  – Establishment: Create an AI Safety Institute for enforcing safety and transparency standards; invest in interoperability to aid integration.

- **Flexible and Responsive Legislation**

  – Adaptability: Develop adaptable legislative frameworks; encourage open-source tools for AI testing and safety.
  – Standards and Certification: Implement ubiquitous standards; offer third-party and self-certification.

- **Risk Management**

  – Representative Data: Ensure unbiased data sets; invest in computing power, including quantum computing.
  – Mitigating Impact: Create indemnity and risk mitigation strategies.

4.5   Theme 3: how do we ensure inclusion and trust?

4.5.1   Inclusion – connecting the unconnected

ITU has been providing statistics on global connectivity, particularly in terms of Internet access, for many years.

Figure 13: Individuals using the Internet / Percentage of population vs time
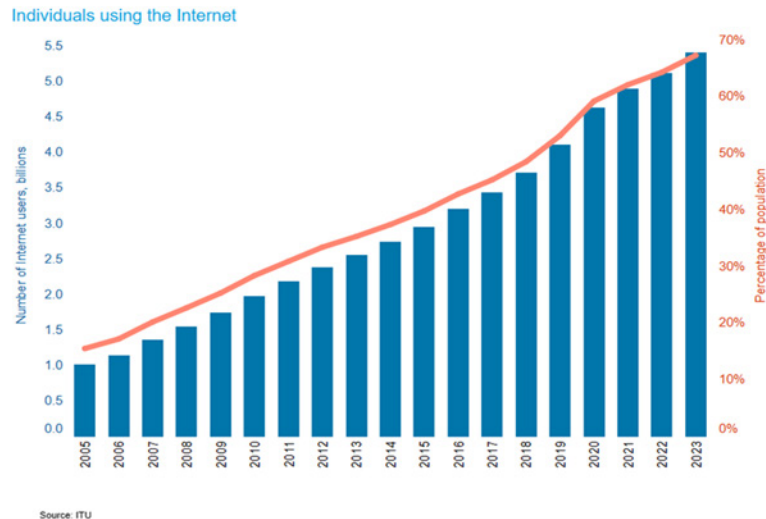
For the year 2023, global Internet usage is up to approximately 67% of the world's population (8+ billion people), or 5.4 billion people, marking a 4.7% increase since 2022, up from the previous year's growth rate of 3.5%. The offline population in 2023 decreased to an estimated 2.6 billion, about 33% of the global total.

70% of men worldwide use the Internet, compared to 65% of women. This results in 244 million more men than women using the Internet as of 2023. Gender disparity is still a distant prospect in regions with low Internet use.

4.5.2 Giving voice for the global majority

Many people and many countries that are being impacted by AI are not integrated in ongoing discussions to set AI policies.

Many share the concern that foreign companies supplying AI systems might capture a significant part of their economic value and knowledge. Since AI shapes our perception of the world, how can cultures be preserved if tools are built on foreign references? This issue is even more pressing for less digitized societies and non-English-speaking countries, as generative AI models are mainly trained on English data and perform best in that language.

Potential discussion questions

- How do we reach inclusive participation from individuals representing diverse backgrounds, cultures, and regions in AI research, development, and decision-making processes?
- How do we enhance access to AI resources to achieve sustainable development goals?
- How do we address bias and discrimination in AI algorithms and systems to prevent inequalities and to ensure fair outcomes for all?
- How do we provide capacity building, education, and skill development in AI among not represented, or underrepresented, communities to enhance their participation in and contribution to the field?

- How do we uphold human rights and ethical considerations in AI development such as transparency, accountability, privacy and others for all individuals globally?
- How do we have the global majority take part in shaping the future of AI technologies and AI governance?

4.5.3   Open vs closed sourcing of (generative) AI models

The debate surrounding the open- vs closed-sourcing of increasingly capable AI models highlights the tension between the benefits of transparency, external oversight and rapid innovation against the potential risks of misuses and unintended consequences. "Open-sourcing AI models" for our purposes means that the AI model architecture and the associated weights are freely and publicly available to anyone to use or modify. Closed, or proprietary, refers to AI systems, particularly foundation models, where the underlying algorithms, model architecture, datasets, and training methodologies are kept confidential by the entity that developed them.

Arguments in favor of open sourcing of (generative) AI models:

- Promotes innovation: By making the models public, a broader range of developers can contribute to and enhance the technology.
- Increases transparency: Open-sourcing allows for community auditing of the models, which can lead to improvements in model safety and ethics. More eyes spot more bugs.
- Fosters collaboration: Open-source AI systems encourage collaboration among researchers and developers, leading to faster innovation and improvement of technology.
- Increases access: Open-source AI systems are more accessible to a wider range of users, including those from resource-constrained environments, promoting inclusivity and access to AI technology. This can reduce the knowledge and resource gap between large corporations and smaller entities, counteracting the centralization of power in AI companies.
- Prevents vendor lock-in: Users of closed-source AI systems may become dependent on a single vendor, making it difficult and costly to switch to alternative solutions in the future.
- Reduces monopolistic practices: Provides opportunities for smaller entities to participate in and benefit from advanced AI without the prohibitive costs of developing proprietary models.

Arguments against open sourcing of (generative) AI models:

- Security risks: There is an increased risk of misuse as more actors can access powerful AI tools, potentially leading to harmful applications.
- Quality control: It may be challenging to maintain high standards of quality and reliability when control over modifications is decentralized.
- Irreversibility: If a model has been released with a flaw that would allow grave misuse, or inherent safety risks of the model, there is no straightforward way to prevent someone from continuing to use the model or to ensure that users install patches to fix the model.
- Regulatory challenges: Open-source models could complicate efforts to enforce compliance with ethical standards and legal regulations due to their widespread and uncontrolled distribution.

Potential Discussion Questions

- How has open-sourcing of AI models benefited startups, academic institutions, and developers in resource-constrained environments?
- Are we seeing the advantages/disadvantages from already open- or closed-sourced models?

- Are there other ways of pursuing open-source objectives while keeping models closed-source?
- How do we decide if models should be open- or closed-sourced?

4.5.4   Insights from the breakout sessions: theme 3

The following outlines key strategies and actions essential for ensuring inclusion and trust in AI.

- **Differentiation and guidelines**

    - Develop guidelines to distinguish between AGI and specific AI models.
    - Tailored oversight and regulation are essential for managing each AI type effectively.

- **Incentives and policies**

    - Implement compute taxes and investment requirements to incentivize businesses to prioritize inclusivity and trustworthiness in AI development.
    - Transparency and Oversight
    - Establish sandbox environments for AI testing with oversight from civil society, academia, and public/private sectors.
    - Ensure testing processes and outcomes are transparent to build community trust.

- **Certification and ethical standards**

    - Create sector-specific certifications with diverse global input.
    - Facilitate international dialogue to establish ethical guidelines that respect local regulations and cultural differences.

- **Education and connectivity**

    - Promote digital literacy and expand connectivity to underserved communities.
    - Support global participation in AI by improving digital skills and access to AI resources.

- **Governance and regulatory frameworks**

    - Develop clear ethical guidelines and regulatory frameworks with global consensus.
    - Ensure governance processes are open and inclusive, allowing for public input and scrutiny.

- **Addressing bias**

    - Urgently address biases in AI algorithms to ensure fairness and prevent inequalities.
    - Open Ecosystem and Collaboration
    - Encourage an open AI ecosystem to foster innovation and trust.
    - Promote collaboration between developers, governments, and NGOs.

- **Data hubs and connectivity**

    - Develop data hubs and improve connectivity, especially in developing regions, to support AI infrastructure and access.

- **Social dialogue and diverse perspectives**

    - Facilitate regular discussions between stakeholders to understand and adopt AI.
    - Ensure diverse perspectives are included in AI development and decision-making processes.

- **Transparency and accountability**
    - Make AI systems transparent and accountable.
    - Users should be informed when AI is involved in decisions.

- **Cybersecurity and surveillance**
    - Protect AI systems from cyber threats.
    - Balance AI surveillance benefits with privacy and civil liberties.

4.6 Update from United Nations high-level advisory body on AI

Reporting live from Singapore where the United Nations High-level Advisory Body on AI (HLAB) met for its third and final meeting in Singapore, 28-29 May 2024, were:

- Carme Artigas, Co-Chair of United Nations High-level Advisory Body on Artificial Intelligence
- Amandeep Singh Gill, UN Secretary-General's Envoy on Technology

Carme Artigas, one of the two co-chairs of the United Nations High-level Advisory Body on AI, recalled that the AI Advisory Body has been working since last October to address the global governance deficit, especially concerning inclusivity and diversity. The Advisory Body published an interim report in December 2023, and its final report was aimed to be published in the early summer of 2024. She noted that the Advisory Body had conducted more than 40 consultations involving more than 1 000 experts and some 20 deep dives into thematic areas.

Carme Artigas highlighted the importance of increasing developing countries' capacities to benefit from AI and the need for reliable, timely scientific information for policymakers.

Ms Artigas outlined six key conclusions from recent meetings:

1) providing reliable, scientific and timely information for policy makers on AI capabilities, opportunities and risks;
2) linking existing governance efforts through regular dialogue;
3) promoting international cooperation around standards;
4) supporting the global South to avoid an AI divide;
5) pooling financial resources, expertise, data and compute to leverage AI for the Sustainable Development Goals;
6) establishing a coordinating mechanism to avoid fragmented efforts and support a safe, inclusive AI future.

> *"I want to share what we're feeling here: there is a momentum, and the momentum is now. This is the moment when we need to connect all global efforts and take bold decisions because what we are seeing when we're talking about governance is not a hindrance, it is a capacity, an enabler for AI development for good." (Carme Artigas)*

Amandeep Singh Gill, the UN Secretary-General's Envoy on Technology, added that the upcoming United Nations Summit of the Future (22-23 September 2024) offers a chance to include AI governance in a global digital compact.

"There is a concentration of AI capacity in some areas, and we need to address this urgently so that some of the harms that were highlighted during the panel discussion can be addressed, and we can direct the arc of investments, political interest, and talent in the direction of AI for the SDGs." (Amandeep Gill)

Mr Gill said that the United Nations were ready to assist the international community in harnessing AI's benefits while tackling its risks, ready to help building capacity to reinforce the vital work on standards, to help producing assessments that assist policymakers in the public and private sectors, and to assist researchers and scientists in making the right decisions to improve the quality of their decision-making.

Figure 14: Carme Artigas, Co-Chair of United Nations High-level Advisory Body on Artificial Intelligence, reporting live from Singapore

4.7   Governance pulse check

A real-time poll was taken among the participants at the beginning of the morning session (pre-pulse check) and at the end of the morning session (post-pulse check). Five statements related to AI in general and five on AI governance were presented to the participants. Participants were asked: "Do you agree with these statements?". Answers could be given on a scale from "strongly disagree" to "strongly agree". Please note also the distribution curves overlaid for each of the ten statements.

Figure 15: Pre-pulse check on whether participants strongly disagreed (value of 1) or strongly agreed (value of 5) with statements regarding AI in general. The scale was from 1 to 5.

Pre-pulse check	Post-pulse check
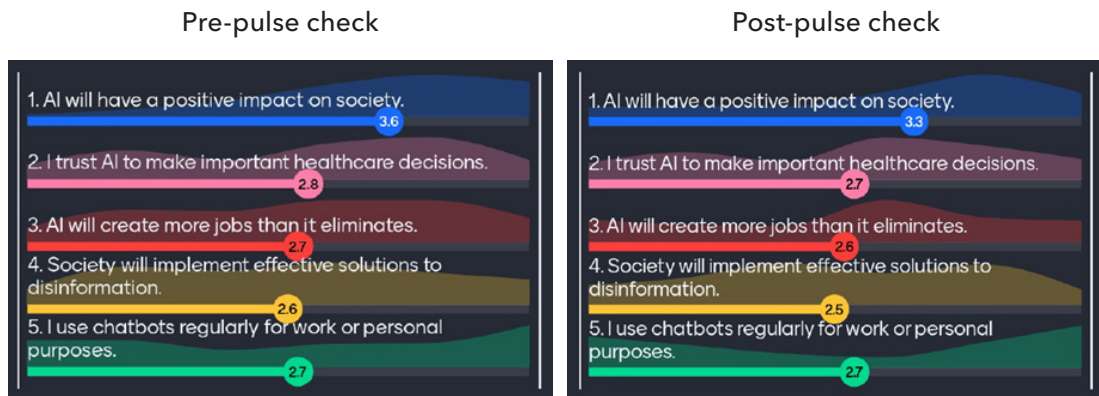
Figure 16: Pre-pulse check on whether participants strongly disagreed (value of 0) or strongly agreed (value of 5) with statements regarding AI governance. The scale was from 0 to 5.

Pre-pulse check	Post-pulse check
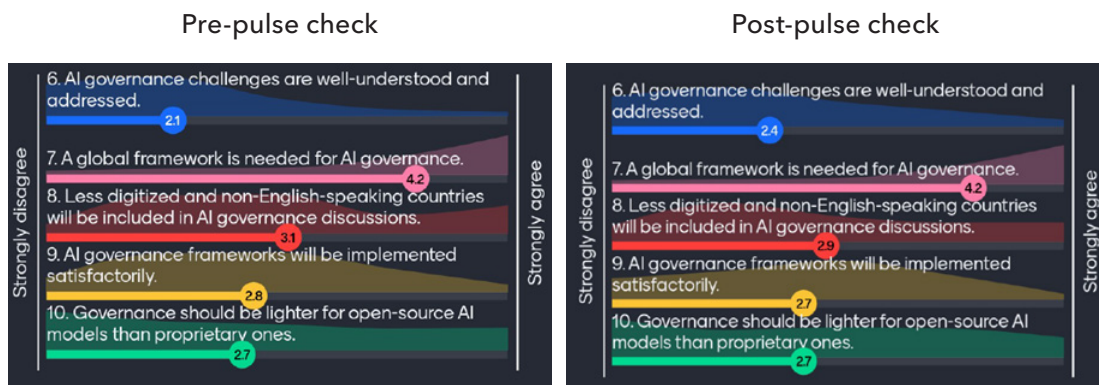
5.   Public afternoon session

5.1   Welcome speech, Doreen Bogdan-Martin, ITU Secretary-General

Good afternoon, everyone. Welcome to Day Zero of the AI for Good Global Summit. Our eagerly anticipated Governance Day is off to a running start. We've already put our AI experts and government leaders to work this morning. We've spent the entire morning exchanging ideas on three critical topics: surveying the AI landscape and understanding how it might evolve, looking at how to implement AI governance frameworks, and, perhaps most importantly, discussing how we can ensure inclusion and trust as we implement those frameworks.

This morning, we heard about various governance efforts, the areas they have in common, as well as some of their differences. Crucially, we learned from developing countries because we want to ensure that they are not left out of the process. This challenges the argument that governments lack initiative when it comes to tech regulation. In just a few moments, you'll be hearing from some of our amazing roundtable participants who will be sharing the outcomes of their work.

But first, let me tell you why we're doing this. Why are we here today? What is AI Governance Day all about, and why are we at the ITU going to keep doing it?

As many of you know, ITU is the UN agency for digital technologies, and we have been working to harness AI for good for the past seven years. We've been convening the UN system around AI, and we've been co-leading an interagency coordination mechanism with UNESCO since 2021. Through our AI for Good platform, a multi-stakeholder community of 28 000 people from over 180 countries, our focus has been on putting artificial intelligence at the service of the Sustainable Development Goals. That's been our compass.

What's new is this much sharper, stronger focus on governance. It's not the benefits but the risks of artificial intelligence that keep us all awake at night. Much has been said about AI governance in the media, academic circles, startups, tech giants, and from local governments all the way to the United Nations, which recently adopted a historic resolution recognizing AI's potential to advance the SDGs.

Governance and technology – we have been here before

Ladies and gentlemen, at the heart of all of this is a conundrum: how do we govern a technology if we don't yet know its full potential? There is no one answer to that question, but we do know that we have been here before. Twenty years ago, the Internet was met with a similar mix of shock, awe, and skepticism. It raised the same questions about how our economies, societies, and environment would transform for better and for worse. We're still grappling with those questions two decades later. In fact, we still don't know the full potential of the internet because a third of humanity has never connected. But before we could even realize the potential, generative AI came along.

Yet even with the convergence of these world-changing, interdependent technologies, governance efforts have emerged. They may not be perfect, but we're not starting from scratch. The Internet Governance Forum and the WSIS Forum were born out of the World Summit on the Information Society. Some of you, like me, were there when this all happened 20 years

ago. I remember how then, as now, we didn't even have the vocabulary to describe what we were dealing with, but that didn't stop us from moving forward.

What we've learned from the WSIS (World Summit on the Information Society) process is that we can take steps toward governance even if we're building the plane as we fly it. We can come together as a community, share experiences, practices, lessons learned, barriers, and challenges, knowing that once more, there is no one-size-fits-all approach to balancing benefits and reconciling regulatory risks. We must look at governance from many different angles and the only way forward is through a multi-stakeholder approach.

That's why I'm so glad that today, gathered in this room, we have our WSIS community with us. Welcome to the WSIS community. We hope that you will help guide us through these many complex questions and challenges.

After listening closely to this morning's discussions, I believe there are three key pieces that must be part of any AI governance effort.

Development of technical standards

The first piece, and obviously very relevant to the ITU, is the development of technical standards. As we heard this morning, those working on AI governance already recognize how technical standards can help implement effective guardrails and support interoperability. This is where ITU has such a key role to play as an international standards development organization. We already have over 200 AI-related standards that we have either developed or are in the process of developing. As part of the World Standards Cooperation, a high-level collaboration between IEC, ISO, and ITU, we are advancing the development of global standards that can make AI systems more transparent, explainable, reliable, and secure. This provides certainty in the market and eases innovation for both large and small industry players everywhere, including in developing countries.

Putting core UN values at the heart of AI governance

The second element is putting human rights, inclusion, and other core UN values at the heart of AI governance. All stakeholders deserve a voice in shaping AI's present and future. But who can afford the compute resources that go into producing AI applications? Who is on the teams that design the foundational models? Right now, the power of AI is concentrated in the hands of a few. This is risky and ethically precarious for humanity. We must work towards an inclusive environment where diverse perspectives, including those on gender, are reflected in policies that align with UN values. International AI governance efforts must account for the needs of all countries, and that's why the United Nations, together with governments, companies, academics, civil society, and the technical community, must play a key role in ensuring that power is distributed equitably. This will not happen automatically.

Development through capacity building

The third element is inclusive development through capacity building. ITU has a long history of bringing the voices of the global south to the emerging technology table. Part of this means making sure that every workforce in the world can deal with the challenges and risks brought about by artificial intelligence. That's why we've been integrating AI capacity support in our digital transformation offerings. We'll continue to roll out these initiatives with many of our UN

partners, including UNDP, focusing on countries with low technological capabilities to help upskill them no matter where they are in their AI journey.

Are you AI ready?

Ladies and gentlemen, governance is not a given. An AI readiness survey that ITU recently conducted among its 193 member states demonstrated that a majority of our responding Member States, actually 85%, don't have any AI regulations or policies in place. Today, some might at least start thinking about the policy elements and what to do next. This makes the work we're doing today and beyond absolutely fundamental and essential.

All good governance starts with listening – listening to experts, exchanging ideas and experiences with peers, identifying gaps, and building on potential areas of convergence. Governance is never a one-and-done process; it is an iterative, sometimes frustratingly slow, but ultimately necessary multi-stakeholder process. Taking stock of the landscape and facilitating deep discussions, as we did this morning, is the first step in transforming principles into practical implementation. Implementation, ladies and gentlemen, is what today is all about.

I know everyone in this room has a stake in seeing AI used as a force for good in this world. As we heard from the UN Secretary-General's high-level advisory body, we need to take bold decisions and view governance not as an inhibitor but as an enabler – an enabler for AI for Good.

That's why today, I'm calling on all of you to get involved, take action, and participate actively in the AI governance activities happening here at the ITU. Let's harness the power of this AI community to govern AI with and for the world. Let's show them what it looks like, how it's done, and let's do it together.

Thank you very much.

5.2   Leaders speak: Insights and key findings on AI governance implementation

Panelists:

– H.E. Ms. Emma Inamutila Theofelus, Minister of Information and Communication Technology, Namibia
– H.E. Mr. Mauricio Lizcano, Minister of Information Technologies and Communications, Colombia
– H.E. Mr. Zunaid Ahmed Palak, Minister of State for the Ministry of Posts, Telecommunications and Information Technology, Bangladesh

Moderator:

– Robert F. Trager, Professor at the University of Oxford

The three ministers summarized the discussions in the morning with 200 attendees. The panel discussion underscored the importance of a coordinated, inclusive, and human-centric approach to AI governance. The ministers highlighted both the opportunities and challenges that lie ahead in the realm of AI. As Minister Palak aptly quoted his Prime Minister, "If you want to go fast and far, innovate together." This sentiment encapsulates the spirit of global cooperation

needed to navigate the complexities of AI governance and ensure that its benefits are shared equitably across the world.

Figure 17: The panelists and moderator from the Center Stage session "Leaders speak: Insights and key findings on AI governance implementation"

Minister Theofelus, Namibia

Minister Emma Theofelus kicked off the discussion by highlighting the productive nature of the morning's conversations. She outlined several critical areas of focus

– **Global coordination**: The necessity of coordinating diverse efforts at a global level to reduce fragmentation among UN agencies, governments, and regional bodies.

> *"Governance is surely broader than regulation, and therefore we need to reduce the fragmentation […] to ensure that we have standards that we can all comply [with] around AI governance." (Emma Inamutila Theofelus)*

– **Human-centric AI**: Ensuring that AI remains focused on human rights and maintaining a human element in its applications.
– **Leveraging existing instruments**: Building on existing frameworks and regulations, rather than reinventing the wheel.

> *"We don't necessarily need to create […] or build new institutions, we can already build on existing capacities and existing institutions." (Emma Inamutila Theofelus)*

– **Inclusive data governance**: Addressing data governance as a fundamental step to inclusive AI governance.

> *"We need everybody around the table, whether it's global north or […] south, to ensure that nobody is left behind." (Emma Inamutila Theofelus)*

- **Local applicability**: Considering national capacities, cultural, and linguistic differences to avoid biases in AI systems.
- **Keeping up with technology**: Involving AI developers in governance discussions to keep pace with the fast-evolving nature of AI.
- **Balancing regulation and innovation**: Ensuring that regulation does not stifle innovation but instead sparks it.

> *"We must strike a balance between regulation and innovation and ensure that regulation actually sparks innovation" (Emma Inamutila Theofelus)*

- **Parallel discussions**: Promoting continuous dialogue among all stakeholders to avoid isolated discussions between producers and users of AI.

Minister Lizcano, Colombia

Minister Mauricio Lizcano of Colombia provided a reflection on the risks and opportunities AI presents. He stressed the importance of human-centric AI and the need for global cooperation. Minister Lizcano highlighted two major risks:

- **Human-serving AI**: Ensuring that AI serves humanity and not the other way around.
- **Competition**: Managing the competition between companies and countries to prioritize safety over winning the AI race.

He called for a democratization of AI, ensuring that it is not limited to a select few but is accessible and understandable to all.

> *"AI doesn't have to be a discussion [among] elite people […] we need real people to understand what is happening with AI." (Mauricio Lizcano)*

Mr. Minister Lizcano also shared Colombia's ambitious initiatives to democratize AI through education and capacity building.

> *"In Colombia, we are training secondary [students] in machine learning and python [coding] […] creating the first faculty of AI in Latin America[1]." (Mauricio Lizcano)*
>
> ---
>
> [1]   [Colombia's government to contribute 50 billion pesos to new AI center in Bogotá](#) (The Bogota Post Article).

Figure 18: H.E. Ms. Emma Inamutila Theofelus, Minister of Information and Communication Technology, Namibia

Figure 19: H.E. Mr. Mauricio Lizcano, Minister of Information Technologies and Communications, Colombia

Mr. Minister Palak, Bangladesh

Minister Zunaid Ahmed Palak of Bangladesh introduced the development of "G Brain," a government AI aimed at enhancing service delivery.

"Our most ambitious initiative in Bangladesh is the development of government brain, in short G Brain." (Zunaid Ahmed Palak)

Minister Palak recounted a recent demonstration where a startup created his digital twin using AI, showcasing AI's potential in education and training.

"This experience was both fascinating and eye-opening." (Zunaid Ahmed Palak)

Figure 20: H.E. Mr. Zunaid Ahmed Palak, Minister of State for the Ministry of Posts, Telecommunications and Information Technology, Bangladesh

However, Minister Palak also warned of the dangers of AI-generated misinformation, citing a recent incident involving a fake photograph of a cyclone's aftermath.

"This incident highlights a serious risk associated with AI defects, spray of fake information and the consequences of which can be dire, affecting social harmony, public safety and the credibility of our institutions" (Zunaid Ahmed Palak)

He called for robust measures to counteract the misuse of AI, emphasizing the need for ethical guidelines and regulatory frameworks.

> *"We must also develop robust measures to safeguard against this misuse [through] ethical guidelines and regulatory frameworks to ensure AI is used responsibly"*
> *(Zunaid Ahmed Palak)*

5.3   State of play of major global AI governance processes

Panelists:

– H.E. Mr. Hiroshi Yoshida, Vice-Minister for Policy Coordination, Ministry of Internal Affairs and Communications, Japan

– Thomas Schneider, Ambassador and Director of International Affairs, Swiss Federal Office of Communications, and Chair of the Council of Europe Committee on AI

– H.E. Mr. Dohyun Kang, Vice Minister, Ministry of Science and ICT, Republic of Korea

– SHAN Zhongde, Vice Minister, Ministry of Industry and Information Technology, People's Republic of China

– Alan Davidson, Assistant Secretary of Commerce for Communications and Information, Head of the National Telecommunications and Information Administration (NTIA)

– Juha Heikkilä, Adviser for Artificial Intelligence, European Commission.

Moderator:

– Ebtesam Almazrouei, Founder and CEO of AI E3, United Arab Emirates

Figure 21: The panelists and moderator of the Center Stage session: State of play of major global AI Governance processes

Mr. Ambassador Thomas Schneider on the Council of Europe treaty

Mr. Ambassador Thomas Schneider highlighted the complexity of AI governance and the necessity for a context-based approach. He drew parallels to engine regulation, emphasizing that AI should not be regulated as a monolith but in terms of its functions and impacts.

> *"We regulate the people that are driving the engines. We regulate the infrastructure. We regulate or protect people affected. It's all context-based. It's not the engine; it's the function of the engine, the effect of the tool. [...] The same logic should be applied to AI." (Thomas Schneider)*

He also spoke about the Council of Europe treaty, designed to ensure that existing human rights protections apply to AI.

> *"The Council of Europe treaty was drafted not in a spirit to create new human rights, not to reinvent the wheel, but actually to make sure that the existing human rights and protections for democracy and the rule of law are applied in the context where AI is used." (Thomas Schneider)*

Figure 22: Thomas Schneider, Ambassador and Director of International Affairs, Swiss Federal Office of Communications, and Chair of the Council of Europe Committee on AI

Mr. Juha Heikkilä on the European AI act

Mr. Juha Heikkilä provided an overview of the European AI Act, described as "the first comprehensive, horizontal, and legally binding regulation globally." He detailed the phased implementation of the Act and its risk-based approach.

> *"The implementation itself is based on a strong pre and post market system of enforcement and supervision. It's a decentralized system of national notified bodies checking compliance with the AI Act requirements before high-risk systems can be placed on the EU market." (Juha Heikkilä)*

Mr. Juha Heikkilä highlighted also the importance of the European AI Office, the scope of which is much broader that the role of safety institutes set up in some countries as it also deals with research, innovation and deployment aspects, as well as international engagement.

Mr. Heikkilä emphasized the importance of trust in AI systems.

> *"Trust is important because trust is the sine qua non for uptake […] and uptake is the sine qua non for benefits to materialize." (Juha Heikkilä)*

Figure 23: Juha Heikkilä, Adviser for Artificial Intelligence, European Commission

Mr. Alan Davidson on the U.S. executive order

Mr. Alan Davidson discussed the U.S. approach to AI governance, starting with voluntary commitments from leading AI companies and the issuance of an executive order by President Biden. He highlighted the establishment of the U.S. AI Safety Institute and the need for international collaboration.

> "Just as an example, I will talk about the sustainable development goals, which, as many of you know, we are on track to achieve just 12% of our goals in that space. In these benchmarks, we've plateaued on many of them, and on some of them, we're actually regressing. But studies suggest that AI could accelerate progress on 80% of the SDGs, in part by automating the work and improving decision-making." (Alan Davidson)

Figure 24: Alan Davidson, Assistant Secretary of Commerce for Communications and Information, Head of the National Telecommunications and Information Administration (NTIA)

Mr. Shan Zhongde on China's AI governance

Vice Minister Shan Zhongde emphasized China's commitment to AI ethics and the practical steps taken to implement AI governance. He spoke about China's efforts to prevent risks, promote transparency, and ensure equity and justice in AI applications.

> "We are drafting and creating different strategies and methods, and through research, we are progressing. We are working on different algorithms, and in that way, we are engaging in in-depth work with AI to adopt a number of policies and regulations. We are also [...] dividing products in different categories: finance, health, transport systems. [...] In all these fields we have specific standards." (Shan Zhongde)

Figure 25: Shan Zhongde, Vice Minister, Ministry of Industry and Information Technology, People's Republic of China

Figure 26: H.E. Mr. Hiroshi Yoshida, Vice-Minister for Policy Coordination, Ministry of Internal Affairs and Communications, Japan

Mr. Hiroshi Yoshida on the G7 Hiroshima process

H.E. Mr. Hiroshi Yoshida discussed Japan's international efforts in AI governance, particularly through the Hiroshima AI Process.

> *"[Instead of hesitating] to get the best of AI because of the risks […] we should mitigate those risks and get the best use" (Hiroshi Yoshida)*

He highlighted the importance of interoperability in AI governance frameworks.

> *"The concept of the Hiroshima AI Process is that we need some kind of governance framework, but it should be interoperable. [However], interoperable [does not mean that] every country has to take the same action." (Hiroshi Yoshida)*

Mr. Dohyun Kang on Korea's AI safety initiatives

H.E. Mr. Dohyun Kang shared insights from Korea's AI Safety Summit which dealt with safety, innovation and inclusivity. He emphasized the country's focus on international collaboration for AI safety standards and the importance of taking practical steps to enhance AI governance.

> *"The Seoul Declaration emphasized the importance of testing and measuring AI safety and addressing all the various side impacts of AI." (Dohyun Kang)*

Figure 27: H.E. Mr. Dohyun Kang, Vice Minister, Ministry of Science and ICT, Republic of Korea

The path forward

The panelists agreed on the necessity of international cooperation and multi-stakeholder discussions to develop effective AI governance frameworks. They emphasized the need for transparency, accountability, and interoperability to ensure that AI technologies are developed and deployed responsibly.

Mr. Thomas Schneider underscored the importance of a global governance framework that reflects diverse cultural and situational needs.

> *"We should cooperate together, and not just governments, but all stakeholders, to develop a global governance and cooperation framework that allows us to do the same in different ways that reflect our situations, cultures, and needs." (Thomas Schneider)*

The discussion concluded with a call to action for continuous collaboration and the sharing of best practices to harness AI for the benefit of all humanity. Ms. Ebtesam Almazrouei summarized the sentiment of the panel.

> *"AI for good should be our compass, and how we can harness the power of AI across all the 17 SDGs is a crucial step that we all should agree on as government, industry leaders, NGOs, and academic institutions." (Ebtesam Almazrouei)*

Figure 28: Ebtesam Almazrouei, Founder and CEO of AI E3, United Arab Emirates

Figure 29: The panelists and moderator from the Center Stage session: Leveraging the UN system to advance global AI Governance efforts
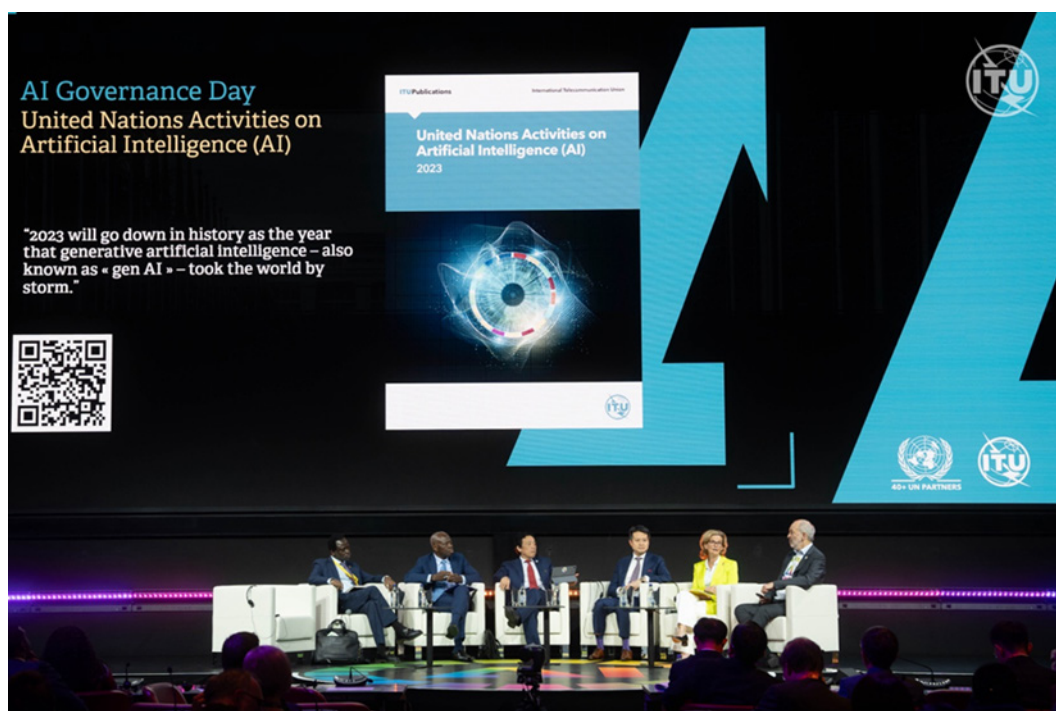
Looking ahead: the future of AI governance

As countries continue to develop and refine their AI governance frameworks, the insights shared during this panel provide a valuable roadmap. The emphasis on inclusivity, international collaboration, and the adaptation of regulatory approaches to specific cultural and situational contexts will be key to ensuring that AI technologies are both innovative and responsibly managed.

The establishment of AI safety institutes in various countries, the implementation of comprehensive regulatory frameworks like the European AI Act, and the continued dialogue through international forums and summits all contribute to a robust global AI governance ecosystem.

The panelists' collective vision of a future where AI technologies are harnessed for the greater good, while mitigating risks and ensuring equitable access, underscores the importance of their ongoing efforts. As the field of AI continues to evolve, the lessons learned and shared during this panel will undoubtedly shape the future of AI governance worldwide.

5.4 Leveraging the UN system to advance global AI Governance efforts

Panelists:

- Doreen Bogdan-Martin, Secretary-General, International Telecommunication Union (ITU)
- Tshilidzi Marwala, Rector, United Nations University (UNU); Under-Secretary-General of the United Nations
- Daren Tang, Director General, World Intellectual Property Organization (WIPO)

- Gilbert Houngbo, Director-General, International Labour Organization (ILO)
- Dongyu Qu, Director-General, Food and Agriculture Organization (FAO)

Moderator:

- Reinhard Scholl, Programme Chair, AI for Good

The panel focused on two goals: how AI can help achieve the United Nations Sustainable Development Goals (SDGs), and how the UN system can support AI governance efforts and international collaboration.

Achieving the the United Nations sustainable development goals with AI

The discussion began with a question posed by Programme Chair Mr. Reinhard Scholl to Secretary-General Doreen Bogdan-Martin regarding the role of the UN system in ensuring that AI supports the Sustainable Development Goals. Secretary-General Bogdan-Martin emphasized the importance of collaboration and leveraging existing platforms.

> *"The proof is actually right here on this amazing panel […] what countries need, what countries want to see, is actually the UN working as one on the ground." (Doreen Bogdan-Martin)*

Secretary-General Bogdan-Martin highlighted the fundamental pillars of the 2030 Agenda: people, planet, prosperity, peace, and partnership, noting that AI is central to all these pillars. She pointed out the AI for Good platform, which brings together over 40 UN agencies and 27 000 experts from 180 countries, the just now launched UN AI Activities Report with some 400 initiatives and projects within the UN system, ITU's technical work with WHO and WIPO on health, with FAO on agriculture, as well as the UN's interagency mechanism on AI which ITU leads with UNESCO.

Addressing the skills gap

Mr. Tshilidzi Marwala, Rector of the United Nations University and UN Under-Secretary-General, was asked about the UN's role in closing the skills gap and ensuring equitable benefits from an AI-dominated economy. Under-Secretary-General Marwala outlined four critical areas: access to data, access to expertise, access to computing resources, and access to good applications. He noted the high cost of cloud computing and the lack of high-performance computing resources in the global South.

> *"We need to create a platform where people from the global South and people from the global North can come together and co-create." (Tshilidzi Marwala)*

He also stressed the importance of changing human behavior, creating incentives, and establishing proper policy frameworks to ensure effective utilization of AI.

Figure 30: Doreen Bogdan-Martin, Secretary-General, International Telecommunication Union (ITU) speaking alongside Daren Tang (left), Director General, World Intellectual Property Organization (WIPO) and Reinhard Scholl, Program Chair, AI for Good

Figure 31: Tshilidzi Marwala, Rector, United Nations University; Under-Secretary-General of the United Nations

Balancing intellectual property and innovation

Mr. Daren Tang, Director General of the World Intellectual Property Organization (WIPO), addressed the balance between intellectual property rights and promoting innovation and equitable access in AI governance. Director General Tang highlighted that intellectual property (IP) should be viewed as a means to an end, rather than an end in itself.

"We have to look at IP as a means to an end […] creating jobs, supporting entrepreneurs, supporting people who have got the hunger to pick up skills and need those skills to bring their ideas to the market. […] Ultimately, it is a catalyst for growth and development." (Daren Tang)

Mr. Tang emphasized the need for an ecosystem approach to ensure that technology can be meaningfully absorbed, particularly in the global South. Tang also underscored the human-centered nature of the IP system, which places the human creator at its core.

"We need to make sure that AI and technology enhance and supplement our humanity." (Daren Tang)

Figure 32: Daren Tang, Director General, World Intellectual Property Organization (WIPO)

Impact of AI on jobs

Mr. Gilbert Houngbo, Director-General of the International Labour Organization (ILO), addressed concerns about the impact of AI on jobs. Director-General Houngbo acknowledged that jobs are already being lost due to technological advancements, but he also noted that millions of new jobs are being created.

"The key thing is scaling, reskilling, reconversion, and upskilling." (Gilbert Houngbo)

Mr. Houngbo pointed out the importance of lifelong learning and the need for comprehensive data on the impact of AI on informal economies, particularly in low-income countries.

"The UN system will be best placed to ensure that nobody is left behind" (Gilbert Houngbo)

Figure 33: Gilbert Houngbo, Director-General, International Labour Organization (ILO)

Improving food security with AI

Mr. Dongyu Qu, Director-General of the FAO, spoke about the potential of AI to improve food security. He highlighted the need for increased food production to meet the demands of a growing global population and urbanization.

"We need the produce more with less. More quantity, high quality, more food diversity [...] with less input, less negative impact on [the] environment, less social injustice, that's the agricultural food system and rural development [common vision] we share." (Dongyu Qu)

He emphasized the role of AI in improving efficiency, productivity, and value chain management in agriculture. Director-General Qu also underscored the importance of cooperation between biologists and IT experts to leverage AI effectively, stressing the need for a holistic approach to solving food security issues.

Figure 34: Dongyu Qu, Director-General, Food and Agriculture Organization (FAO)

Recommendations for AI governance

In the latter part of the discussion, Programme Chair Scholl asked each panelist for their recommendations on AI governance and international collaboration.

Director General Daren Tang stressed the importance of inclusivity and involving the global South in AI discussions, noting the positive attitudes towards AI in many developing countries.

"We need to be the platform where we are a big tent and we're inclusive." (Daren Tang)

Under-Secretary-General Tshilidzi Marwala echoed the need for a platform for co-creation between the global North and South.

"We need to create a platform where people from the global South and people from the global North can come together and co-create" (Tshilidzi Marwala)

Director-General Dongyu Qu emphasized the importance of a positive, cooperative approach and solving problems holistically.

"Be positive, be cooperative, and solve the problem holistically." (Dongyu Qu)

Director-General Gilbert Houngbo called for UN leadership in coordinating efforts within the UN system and with other multilateral and non-multilateral stakeholders. He also stressed the need to ensure that AI does not exacerbate inequalities.

"The UN system will be best placed to ensure that nobody is left behind" (Gilbert Houngbo)

Secretary-General Doreen Bogdan-Martin concluded with a call for better coordination and inclusivity in AI efforts.

"Let's build on our respective and collective assets. […] Let's make sure we're inclusive and that we don't further exacerbate the divides." (Doreen Bogdan-Martin)

The panel's insights underscored the critical role of the UN system in advancing global AI governance efforts and achieving sustainable development goals. Through collaboration, inclusivity, and a focus on equitable access, the UN can help ensure that AI benefits all of humanity.

5.5   The Government's AI dilemma: how to maximize rewards while minimizing risks?

Panelists:

–   H.E. Ms. Emma Inamutila Theofelus Minister Ministry of Information and Communication Technology, Namibia
–   Mercedes Aramendia Falco, President, Directorio Unidad Reguladora de Servicios de Comunicaciones (URSEC), Uruguay
–   Niraj Verma, Additional Secretary Department of Telecommunications, Government of India

Moderator:

–   Robert F. Trager, Professor, University of Oxford.

This panel discussion underscored the multifaceted nature of AI governance. While the benefits of AI in sectors like healthcare, education, and finance are immense, the associated risks, particularly in cybersecurity and data privacy, require careful management. A collaborative approach, rooted in international standards and adapted to local contexts, appears to be the most promising path forward. As nations like Namibia, Uruguay, and India continue to navigate this complex landscape, their experiences and strategies offer lessons for the global community.

Namibia: leveraging AI for development and addressing cyberthreats

Her Excellency Ms. Minister Emma Inamutila Theofelus emphasized the dual nature of AI as both a potential boon and a threat.

Namibia's challenge is providing services across vast distances in a sparsely populated country – only 3 million people spread over 825 000 square kilometers [Editor: about twice the area of California].

Ms. Theofelus pointed to healthcare as a sector where AI can make a substantial impact. By analyzing data from censuses and routine data collection, AI can help prioritize the placement of healthcare facilities, such as dialysis centers and cancer wards, where they are most needed. This data-driven approach can save lives, optimize investments, and ensure the healthcare system operates efficiently. In education, AI could bridge the gap between current skills and future industry demands by forecasting which skills will be needed and training students accordingly.

However, Ms. Theofelus also noted the risks associated with AI, particularly in the realm of cybersecurity. Namibia experiences around 2.7 million cyber attacks, especially in the financial sector. She stressed the need for increased digital literacy among citizens to protect themselves online.

"We cannot reach every single person because there are barriers in terms of distance [and] barriers in terms of rolling out facilities for them to get trained." (Emma Inamutila Theofelus)

AI could help overcome these barriers by providing personalized training programs in local languages.

On the importance of international governance frameworks and collaborative efforts to address the global nature of AI challenges, Ms. Theofelus advocated for a unified approach, emphasizing that disparate frameworks can be confusing and counterproductive. She stressed the need for consensus on common principles and values that can be adapted to the specific contexts of different countries.

Figure 35: Reinhard Scholl, Programme Chair, AI for Good

Figure 36: The panel and moderator from the Center Stage session: The Government's AI dilemma: how to maximize rewards while minimizing risks?

"We want to see there to be some consensus around a governance framework that would work for the majority […] there must be a way in which we can all come to common grounds and once we take that approach I think it becomes easier for national states and national governments to look at the governance framework and tailor it to their own country." (Emma Inamutila Theofelus)

Uruguay: national strategies and ethical frameworks

Ms. Mercedes Aramendia Falco discussed Uruguay's comprehensive approach to AI governance, rooted in national and international standards. Uruguay's National AI Strategy, initially established in 2020, was updated in 2021 to align with UNESCO's recommendations, ensuring transparency, accountability, and ethical implementation.

"We have a national strategy for AI which dates from 2020. In 2021, following the recommendations of UNESCO, we understood the need to update that strategy on AI." (Mercedes Aramendia Falco)

Uruguay's approach involves collaboration across the entire ecosystem, including public and private sectors, to develop an inclusive and effective strategy.

One of Uruguay's key initiatives is the Ceibal Plan, launched in 2007, which aims to universalize access to education by providing internet and laptops to all public school students. This plan has evolved to incorporate AI, ensuring that students not only have access to technology but also understand how to use it and the associated risks.

"We cannot just have one concrete action, we need many parallel activities to ensure that AI has a positive impact and to ensure that students can access internet and be connected. We are analyzing data to do so [and] we are using AI [for it]." (Mercedes Aramendia Falco)

Despite the benefits, Ms. Falco acknowledged the risks, particularly concerning data privacy and security. She emphasized the importance of maintaining transparency and control over algorithms to prevent misuse of data. In sectors like finance and healthcare, where AI can drive significant improvements, safeguarding sensitive information and ensuring ethical use are paramount.

On the importance of collaboration towards international frameworks, Ms. Falco noted that Uruguay's updated national strategy on AI is aligned with international standards to ensure a cohesive approach. She emphasized the need for ongoing assessment and collaboration with various stakeholders, including the private sector, to develop talent and build capacity, highlighting Uruguay's efforts to incorporate diverse perspectives in shaping AI governance.

Figure 37: H.E. Ms. Emma Inamutila Theofelus, Minister of Information and Communication Technology, Namibia

"Following UNESCO's recommendations, we have started a brand new process: we have approved a national law in which a national agency […] has been set up, which has created a new strategy which incorporates the international standards. The public and private sectors participate in this in this strategy" (Mercedes Aramendia Falco)

India: harnessing AI for economic growth and combating cybercrime

Mr. Niraj Verma highlighted AI's potential to boost India's GDP significantly. A recent study suggests that AI could contribute between $360 billion and $460 billion to India's GDP by 2030. Additional Secretary Verma pointed to the success of India's Unified Payments Interface (UPI) system, which processed approximately 300 billion transactions worth $2 trillion last year.

However, Mr. Verma also highlighted the dark side of AI, particularly the rise in cyber attacks and cyber crimes. The Department of Telecommunications has taken proactive steps, such as launching the Sanchar Saathi portal to combat these threats.

"In the last 3 to 4 months, we have disconnected 6 million SIM cards because of cyberattacks." (Niraj Verma)

Figure 38: Mercedes Aramendia Falco, President, Directorio Unidad Reguladora de Servicios de Comunicaciones (URSEC), Uruguay

Figure 39: Niraj Verma, Additional Secretary Department of Telecommunications, Government of India

The role of AI safety institutes

The idea of AI safety institutes was another key topic of discussion. These institutes aim to create a network for monitoring and mitigating the risks associated with AI deployment. While the concept is appealing, the implementation may vary based on a country's resources and existing infrastructure.

Ms. Falco highlighted the importance of standards and guidelines in ensuring the safe and ethical use of AI.

> *"Standards help us to be able to regulate, that's the work of policy makers in general, and companies when they develop their own standards, […] have references, guidelines which help them go down the right line. That just simplifies everyone's work […] and then we can bring in the participation of tech technicians experts and other stakeholders in the ecosystem so that we can best […] create clear rules."*
> *(Mercedes Aramendia Falco)*

Uruguay's approach involves incorporating these standards into national policies and collaborating with international bodies to maintain a balance between innovation and risk management.

Ms. Theofelus offered a thought-provoking perspective on the terminology used, suggesting a more positive framing might be beneficial. For countries like Namibia, with limited resources, integrating AI safety functions into existing institutions could be a more feasible approach.

> *"Just even the wording 'AI safety institute' already presumes that there's some threat, as opposed to looking at the positive side of things." (Emma Inamutila Theofelus)*

Mr. Verma supported the idea of AI safety institutes, particularly for a large and diverse country like India. He emphasized the role these institutes could play in providing safeguards and setting standards that startups and other entities could follow.

> *"It is a good idea to have an AI safety institute […]. In my country there are a lot of startups and there would be some guidelines [and standards] to follow." (Niraj Verma)*

5.6   The critical conversation on AI safety and risk

Panelists:

- Professor Stuart Russell: Professor of Computer Science at the University of California, Berkeley

Figure 40: Robert F. Trager, Professor, University of Oxford

- Rumman Chowdhury: CEO of Humane Intelligence and USA Science Envoy for Artificial Intelligence
- Lane Dilg: Head of Strategic Partnerships at OpenAI
- Hakim Hacid: Acting Chief Researcher, Technology Innovation Institute (TII)

Moderator:

- Amir Banifatemi: Co-Founder and Director of AI Commons

The panel discussion underscores the complexity and urgency of AI safety and risk management. A multifaceted approach involving rigorous standards, institutional safeguards, and continuous research is essential. As AI technology continues to evolve, the dialogue on safety and risk must remain a priority, ensuring that AI advancements benefit humanity while minimizing potential harms.

Setting the stage: Amir Banifatemi's opening remarks

Mr. Amir Banifatemi began by highlighting the significance of AI safety in the context of rapid technological advancements. He emphasized the need for trustworthiness, reliability, and scalability in AI systems.

> *"AI safety is important because it helps us anchor notions of trustworthiness, reliability, and scalability, and as we go forward with launching systems, we need to think about how safety can be put into place. It is not just about talking about safety, it is also understanding how we can put in place AI safety mechanisms, governance regulation, learning from different Industries." (Amir Banifatemi)*

Figure 41: Panel and moderator from the Center Stage session: The critical conversation on AI safety and risk

Figure 42: Amir Banifatemi: Co-Founder and Director of AI Commons

Professor Stuart Russell on learning from other industries

Mr. Stuart Russell, a leading voice in AI safety, drew parallels between AI and other high-stakes industries like aviation, pharmaceutics and nuclear power. He highlighted the rigorous safety standards in these fields and the need for similar measures in AI.

Figure 43: Stuart Russell, Professor of Computer Science at the University of California, Berkeley

"With aircrafts, there has to be an airworthiness certificate before the airplane can be sold. With medicines, another area that is now safe bud did not use to be safe, the medicine has to go through extensive clinical trials before it is allowed to be sold."
(Stuart Russell)

He pointed out the challenges of applying similar safety standards to AI, particularly due to the opaque nature of deep learning and transformer models, which are often seen as "black boxes."

Mr. Russell also warned about the potential consequences of insufficient safety measures, citing historical examples like the Chernobyl disaster, drawing a stark comparison to the potential risks of AI.

"Despite all that effort, we had Chernobyl, and Chernobyl ended up wiping out the global nuclear industry." (Stuart Russell)

Lane Dilg on balancing innovation and safety

Ms. Lane Dilg of OpenAI discussed the organization's approach to balancing innovation with safety. She emphasized that safety and innovation are inextricably linked and that OpenAI is committed to both.

Figure 44: Lane Dilg, Head of Strategic Partnerships at OpenAI speaking alongside Professor Stuart Russell, Professor of Computer Science at the University of California, Berkeley; Rumman Chowdhury, CEO of Humane Intelligence and USA Science Envoy for Artificial Intelligence and Hakim Hacid, Acting Chief Researcher, Technology Innovation Institute (TII)

"We do consider innovation and safety inextricably intertwined [...] such that we are never looking at only one of those pieces. A couple of ways in which you will see us doing that work: […] being in spaces and in conversations like this [one]; trying to be sure that we are aware of risks that are being raised by civil society and in governance conversations, [ensuring] we are responsive to those" (Lane Dilg)

She highlighted OpenAI's iterative deployment strategy, which involves releasing models in stages to gather feedback and ensure preparedness.

Ms. Lane Dilg also mentioned OpenAI's work on preparedness frameworks and their focus on technical tools and evaluations.

"We are very focused on the technical tools and evaluations that will enable this kind of assessment and this kind of scientific assessment and real judging of capabilities and risks." (Lane Dilg)

Figure 45: Rumman Chowdhury, CEO of Humane Intelligence and USA Science Envoy for Artificial Intelligence

Rumman Chowdhury on identifying risks systematically

Ms. Rumman Chowdhury, a data scientist and ethicist, provided insights into how organizations can systematically identify and manage risks associated with AI. She stressed the importance of evidence-based approaches and the use of established risk management frameworks.

> *"Think through the applications use cases and ensure that what you're doing is evidence-based. We now have a plethora of different risk management frameworks in the US." (Rumman Chowdhury)*

She pointed to frameworks like the NIST Risk Management Framework (RMF) and UNESCO's guidelines as valuable tools for assessing societal impacts.

Ms. Rumman Chowdhury also highlighted the need for a comprehensive view of AI systems, considering not just the models but also the broader socio-technical context.

> *"When we think about identifying risks but also thinking through safeguards, don't just think about the AI models [...] to think about it. Some of the protections you are making are institutional and regulatory." (Rumman Chowdhury)*

Figure 46: Hakim Hacid: Acting Chief Researcher, Technology Innovation Institute (TII)

Acting Chief Researcher Hakim Hacid on human alignment and AI safety

Mr. Hakim Hacid focused on the importance of aligning AI systems with human values. He stressed the need for transparency, control, and verification mechanisms to ensure AI systems are beneficial to humans.

> *"At the end of the day, if you want to make a system safe, it has to be mapped to some human values, to some expectations. The issue here is that it is difficult to define these human values at the end of the day. " (Hakim Hacid)*

He acknowledged the challenges in defining these values and emphasized the importance of continuous control and verification.

Mr. Hakim Hacid also called for patience and collaboration in the pursuit of AI safety.

> *"We need clearly a lot of work to be done on the safety side, but we need also to be patient and work together to get this safety a little bit more mature." (Hakim Hacid)*

Lane Dilg on addressing major safety issues

Ms. Lane Dilg provided specific examples of how OpenAI has addressed major safety issues. OpenAI has been integrating the standard of Coalition on Content Provenance and Authenticity (C2PA) to ensure the provenance of digital content.

Figure 47: Lane Dilg, Head of Strategic Partnerships at OpenAI

"That is a standard that we have integrated in our image generation capabilities and that we also have committed to integrating into our video generation capabilities before deployment." (Lane Dilg)

Head of Strategic Partnerships Dilg also discussed OpenAI's response to cyber risks, highlighting the publication of six critical measures for AI security. Additionally, she mentioned the establishment of a Safety and Security Committee within OpenAI to oversee safety measures and ensure accountability.

Rumman Chowdhury on effective regulation

Ms. Rumman Chowdhury addressed the effectiveness of current regulations in mitigating AI risks. She acknowledged the challenges of evaluating AI models, given their probabilistic nature, and called for more robust benchmarks and evaluation methods.

Ms. Rumman Chowdhury highlighted the role of bias bounty programs and red teaming in identifying and mitigating risks, underscoring the importance of independent scrutiny.

"Red teaming is the practice of bringing in external individuals to stress test the negative capabilities of AI models. Again, it's an inexact science. How many people should be red teaming? How do you know you're done red teaming? Figuring some of these things out will only happen as we perform more of these tests." (Rumman Chowdhury)

Figure 48: Rumman Chowdhury: CEO of Humane Intelligence and USA Science Envoy for Artificial Intelligence

Professor Stuart Russell on promising areas of research

Stuart Russell emphasized that the private sector needed to ramp up its safety research; the resources of academia and government are a drop in the bucket. He stressed the importance of getting the the incentives right that we are training the AI systems to achieve.

Stuart Russell warned that we were hopeless to write down objectives for an AI system completely and correctly, but that what we were doing with large language models was even worse because we are simply training them to imitate human beings.

One area of research that Professor Russell has been working is about so-called "assistance games" where the AI agent is deliberately kept in the dark about the preferences and interests of humans.

> *"I am cautiously optimistic, but it does feel as if we're in a race that we shouldn't have to be in between when we figure out how to control AI systems and when we figure out how to produce AGI" (Stuart Russell)*

5.7 To share or not to share: the dilemma of open source vs. proprietary large language models

Panelists:

- Jim Zemlin, Executive Director of the Linux Foundation
- Melike Yetken Krilla, Head of International Organizations at Google

Figure 49: Panelists and moderator during the Center Stage session: To share or not to share: the dilemma of open source vs. proprietary Large Language Models

– Melinda Claybaugh, Director of Privacy Policy at Meta
– Chris Albon, Director of Machine Learning at the Wikimedia Foundation
– Isabella Hampton, Policy Researcher at the Future of Life Institute.

Moderator:

– Bilel Jamoussi, Deputy to the Director and Chief of Telecommunication Standardization Policy Department at the International Telecommunication Union (ITU)

The open source philosophy

Mr. Jim Zemlin, representing the Linux Foundation, emphasized the foundational role of open source in modern technology.

> *"Open source has been a fundamental building block for all modern technology systems." (Jim Zemlin)*

He highlighted that 80% to 90% of the code in any modern computing system is open source[1]. Mr. Jim Zemlin pointed out that large language models would not exist without open source tools like PyTorch and other components.

However, Jim Zemlin acknowledged the challenges, particularly market consolidation. He proposed the need for standards to define what constitutes an open large language model.

---

[1] Linux Foundation and Harvard's Census II Study; Boston Consulting Group (BCG) Report; PhoenixNAP's Software Composition Analysis

Figure 50: Jim Zemlin, Executive Director of the Linux Foundation

Meta's approach to open source

Meta, a company known for its dual contributions to both open source and proprietary AI, was represented by Ms. Melinda Claybaugh. She underscored Meta's commitment to open source while recognizing the need for a nuanced approach.

> *"What we really want to convey is that this is not binary [...] there's actually a real spectrum." (Melinda Claybaugh)*

Meta's approach includes releasing model weights while keeping training data proprietary. Ms. Claybaugh emphasized Meta's commitment to responsible open sourcing, including rigorous testing and the release of responsible user guides for developers.

> *"For us, a responsible open approach is all the kind of testing that are done from the data collection stage, filtering data, doing risk assessments and mitigations along the way." (Melinda Claybaugh)*

Ethical considerations

From an ethical standpoint, Isabella Hampton from the Future of Life Institute discussed the implications of keeping LLMs proprietary versus open source. She argued that open source should be viewed as a means to an end, not the end itself.

Figure 51: Melinda Claybaugh, Director of Privacy Policy at Meta

"Open source is a tool that we can leverage to accomplish our goals." (Isabella Hampton)

Ms. Hampton highlighted the importance of maintaining a focus on transparency, competition, and safety in the development of these models.

Google's view

Ms. Melike Yetken Krilla of Google recognized the benefits and risks associated with open source models. She shared Google's history of open source contributions, such as the Transformer architecture and the AlphaFold protein structure prediction.

"There is a balance needed in the regulatory action between embracing and allowing some of this innovation while ensuring competition and doing so together" (Melike Yetken Krilla)

Ms. Krilla advocated for a thoughtful and gradual approach to releasing models, with safety testing and commitments to avoid harm.

Figure 52: Isabella Hampton, Policy Researcher at the Future of Life Institute

Figure 53: Melike Yetken Krilla, Head of International Organizations at Google

"First, we're looking at safety testing in advance […] and then identifying how and at what level to release." (Melike Yetken Krilla)

Figure 54: Chris Albon, Director of Machine Learning at the Wikimedia Foundation

Wikipedia's open content model

Mr. Chris Albon from the Wikimedia Foundation highlighted the role of open content in broadening access to knowledge. He underscored the importance of transparency and the community-driven model of Wikipedia.

> *"Wikipedia is one of the best things the internet ever created, a huge pool of information created by humans" (Chris Albon)*

Mr. Albon noted that the integration of open source models into platforms like Wikipedia enhances the value proposition by providing tools for better content moderation and accuracy. However, he stressed the need for credit to the original sources.

Governance and regulation

In addressing the potential for both positive and negative impacts of open source LLMs, the panelists discussed necessary governance frameworks and policies. Jim Zemlin emphasized the importance of placing the regulatory burden on those best equipped to handle it.

> *"Put the regulatory burden on those who are most equipped to handle it Upstream."*
> *(Jim Zemlin)*

Melinda Claybaugh called for a nuanced approach that recognizes the reality of the open source ecosystem.

> *"I think we really need to avoid a kind of blanket approach to regulation." (Melinda Claybaugh)*

Ms. Isabella Hampton expressed optimism about initiatives like the National AI Research Resource (NAIRR), which aims to provide resources for safety research.

Krilla highlighted the importance of collaboration on standardization, involving governments, civil society, and businesses.

> *"We're thinking very thoughtfully about how we are looking at releasing these models and to whom." (Isabella Hampton)*

A consensus was reached on the need for a balanced and nuanced approach to the open source versus proprietary debate. The leaders emphasized that open source and proprietary models each have their place, depending on the specific context and goals. The discussion underscored the critical role of open source in fostering innovation, ensuring transparency, and preventing market consolidation, while also recognizing the need for responsible governance and collaboration to address potential risks.

5.8 Harmonizing high-tech: the role of AI standards as an implementation tool

Panelists:

- Seizo Onoe, Director of the Telecommunication Standardization Bureau (TSB), International Telecommunication Union (ITU)
- Philippe Metzger, Secretary-General & CEO, International Electrotechnical Commission (IEC)
- Sergio Mujica, Secretary-General, International Organization for Standardization (ISO)

Moderator:

- Bilel Jamoussi, Deputy to the Director and Chief of Telecommunication Standardization Policy Department, International Telecommunication Union (ITU)

The harmonization of AI standards is essential for addressing the complex challenges and opportunities presented by AI technologies. The collaborative efforts of ITU, IEC, and ISO are providing a unified framework for AI standards, ensuring trust, safety, and interoperability. Through coordinated action and public-private collaboration, these organizations are well-equipped to lead the way in AI standardization, fostering a sustainable and inclusive digital future.

Figure 55: Panelists and moderator during the Center Stage session: Harmonizing High-Tech: The role of AI standards as an implementation tool

The unique position of IEC, ISO, and ITU in standardization

IEC Secretary-General Philippe Metzger set the stage by emphasizing the historical significance and complementary nature of the three organizations.

> "We certainly consider ourselves leading standards development organizations. ITU is probably the oldest global international organization founded in 1865, ISO in 1947, IEC in 1906. I think we have a long track record and consider ourselves sister organizations because we are quite complementary in what we are doing." (Philippe Metzger)

Mr. Philippe Metzger highlighted that these organizations operate on principles of transparency, diversity, inclusivity, and consensus. He also noted the adaptability of these organizations to emerging technologies like AI.

> "We are consensus based, the standards that we are producing are really reflecting global involvement of […] communities. […] Our systems have proven to be quite adaptable despite the long age of our organizations. AI is the next big dimension, stemming from the foundations we have already been working on." (Philippe Metzger)

Figure 56: Philippe Metzger, Secretary-General & CEO, International Electrotechnical Commission (IEC)

The importance of coordinated action

ISO Secretary-General Sergio Mujica underscored the necessity of coordinated action among the three organizations to provide clarity and consistency in AI standards.

> *"It is our responsibility as leading organizations in the area of standardizations to provide something that makes sense to everyone. People don't care whether it is an ITU, IEC, or ISO standard; they want to know what it is and how it addresses AI-related challenges." (Sergio Mujica)*

He elaborated on the World Standards Cooperation (WSC), which facilitates this collaboration.

> *"The WSC focuses on two main areas: technical collaboration and promoting a positive message about standardization. We coordinate our work programs and create tools for operationalizing our collaboration." (Sergio Mujica)*

Mr. Sergio Mujica highlighted a concrete example of successful collaboration: the development of high-efficiency video coding. This example illustrates how coordinated efforts can lead to significant advancements and global recognition.

Figure 57: Sergio Mujica, Secretary-General, International Organization for Standardization (ISO)

"A wonderful example when we talk about the high efficiency video coding, is the fancy solution that we have created to compress ultra-high-definition TV videos, recognized with an Emmy award." (Sergio Mujica)

Public-private collaboration and policy objectives

Mr. Seizo Onoe, Director of ITU's Telecommunication Standardization Bureau, discussed the role of standards in supporting public-private collaboration and achieving policy objectives.

"Standardization itself is an outcome of public-private collaboration. In the process of standardization, we have involvement from many stakeholders, including governments, regulators, and companies." (Seizo Onoe)

Mr. Seizo Onoe provided examples of how voluntary standards can aid policy objectives, particularly in areas like service quality and sustainability.

"Regulators promoting good service quality work with companies specializing in service quality KPIs and monitoring tools to develop standards that capture the common understanding between companies and regulators." (Seizo Onoe)

Figure 58: Seizo Onoe, Director of the Telecommunication Standardization Bureau (TSB), International Telecommunication Union (ITU)

He also emphasized the role of standards in supporting policy and regulation.

"Policy and regulation can establish rules, and technical standards can provide practical tools to uphold these rules. For example, security controls set by standards are practical tools to ensure safety and privacy." (Seizo Onoe)

Addressing AI-specific issues

When it comes to AI-specific standards, Metzger highlighted four key areas of focus: trustworthiness, sustainability, functional safety, and data quality.

"We have guidance on addressing societal concerns, ethical considerations, and explainability of machine learning models. We are also working on the treatment of unwanted bias and human oversight of AI systems." (Philippe Metzger)

IEC Secretary-General Metzger pointed out the importance of sustainability in AI. Functional safety is another critical area, as well as data quality is equally essential.

Figure 59: Panelists and moderator during the Center Stage session: Harmonizing High-Tech: The role of AI standards as an implementation tool

"We are looking at how AI systems can be created and operated in a sustainable way, including their carbon impact. […] In the context of AI, we need to ensure that electronically controlled products and systems remain reliable and trustworthy. […] We are focusing on the characteristics, properties, and quality of data used in AI processes to ensure meaningful analytic results." (Philippe Metzger)

ISO Secretary-General Mujica echoed these sentiments and highlighted the foundational standards developed by the joint technical committee on AI.

"One significant contribution is agreeing on definitions and terminologies. For example, we talk about AI, advanced AI, and risk, but we don't always mean the same thing." (Sergio Mujica)

He also mentioned the recent launch of a management standard on AI, providing practical guidance for organizations.

"It offers guidance on the implementation of AI, addressing risks, and defining responsibilities within organizations." (Sergio Mujica)

Figure 60: Bilel Jamoussi, Deputy to the Director and Chief of Telecommunication Standardization Policy Department, International Telecommunication Union (ITU)

Practical applications and future directions

ITU's Seizo Onoe provided insights into practical applications of AI standards, citing various collaborative efforts.

> *"The AI for Health initiative with WHO, AI for Agriculture with FAO, and AI for Natural Disaster Management with WMO and UNEP are examples of how we are leveraging AI for global benefits." (Seizo Onoe)*

He also addressed the need to mitigate the negative aspects of emerging technologies.

> *"AI technology is promising for accelerating sustainable development, but we must mitigate negative impacts like deep fakes and misinformation. We are working on standards to verify the authenticity and provenance of multimedia content." (Seizo Onoe)*

In a final note, Mr. Bilel Jamoussi asked the panel if the World Standards Cooperation is effectively coordinating the AI standards landscape. The response was unanimously positive. Panelists affirmed their dedication to ongoing collaboration and inclusivity in AI standardization efforts.

Figure 61: Gabriela Ramos (UNESCO) and Tomas Lamanauskas (ITU) care closing AI Governance Day with the session "From principles to implementation – pathways forward"

Figure 62: Tomas Lamanauskas, Deputy Secretary-General, International Telecommunication Union (ITU), co-chair of the United Nations Interagency Working Group on AI

5.9 From principles to implementation – pathways forward

Panelists:

– Gabriela Ramos, Assistant Director-General for Social and Human Sciences, United Nations Educational, Scientific and Cultural Organization (UNESCO), co-chair of the United Nations Interagency Working Group on AI

– Tomas Lamanauskas, Deputy Secretary-General, International Telecommunication Union (ITU), co-chair of the United Nations Interagency Working Group on AI

The session "From Principles to Implementation – Pathways Forward" featured speakers Mr. Tomas Lamanauskas and Ms. Gabriela Ramos discussing the significant rise in generative AI and the accompanying regulatory and governance challenges faced by policymakers worldwide.

Mr. Tomas Lamanauskas emphasized the UN's response to AI advancements, noting that ITU has a long history of developing AI standards in various sectors such as disaster management, health, autonomous driving, and agriculture, many of which have been developed in collaboration with UN partner agencies.

"We provide the platform for all of us to come together, and this AI for Good Summit is an example of that. This Summit didn't start last year; it started seven years ago, and it's always useful to remember that. Seven years ago, it started as a solution Summit, it started to think, discuss, and agree on how we use AI to progress forward, to propel it. Of course, since then we've gotten a little bit more fearful of AI; we are now much more afraid of it than we were then. So now we discuss how to really put the right guardrails, how to safeguard it. But even in that regard, we also have work that has been done." (Tomas Lamanauskas)

A comprehensive White Paper, the "United Nations System White Paper on AI Governance: An analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance", was published in May 2024. The White Paper was prepared by the Inter-Agency Working Group on Artificial Intelligence, under the leadership of UNESCO and ITU, and endorsed by the Chief Executive Board for Coordination (CEB) of the United Nations, comprising the Executive Heads of the United Nations.

Ms. Gabriela Ramos discussed the importance of a multi-stakeholder approach and the integration of AI across various sectors. She praised the efforts of WHO, UNICEF, UNESCO, and FAO in using AI for health, education, children's well-being, and food systems. She highlighted UNESCO's Readiness Assessment Methodology (RAM), which helps countries advance their AI governance and has been implemented in around 50 countries. Additionally, UNESCO achieved a global consensus on AI ethics, signed by 194 countries, focusing on human rights and dignity.

The speakers stressed the practical implementation and capacity development needed for effective AI governance. Mr. Tomas noted that ITU has 220 technical standards for AI and is working on interoperability, regulatory approaches, and combating deep fakes and misinformation. New initiatives like the AI for Good Impact and a flagship report aim to share knowledge and assist developing countries.

Figure 63: Gabriela Ramos, Assistant Director-General for Social and Human Sciences, United Nations Educational, Scientific and Cultural Organization (UNESCO), co-chair of the United Nations Interagency Working Group on AI

Mr. Tomas Lamanauskas and Ms. Gabriela Ramos underscored the importance of creating and implementing responsible AI frameworks and ensuring global governance keeps pace with technology. They encouraged continued multi-stakeholder discussions and the use of the UN's AI governance paper as a resource. The AI for Good Summit reflects the best of UN traditions and innovation, involving a diverse global community ready to tackle AI challenges and opportunities together.

"AI for Good [...] brings us together with all the international institutions and adopts a multi-stakeholder approach, which is very important." (Gabriela Ramos)

Figure 64: Cover page of the United Nations system White Paper on AI Governance, produced by the Inter-Agency Working Group on AI, available at https://unsceb.org/united-nations-system-white-paper-ai-governance

United Nations System   **02/05/2024**

                                                    Advance unedited version

CEB
Chief Executives Board
for Coordination

High-Level Committee on Programmes (HLCP)

Inter-Agency Working Group on Artificial Intelligence (IAWG-AI)

United Nations System White Paper on AI Governance:
An analysis of the UN system's institutional models, functions, and existing
international normative frameworks applicable to AI governance

Appendix 1: Essential vocabulary for AI governance

"**AI lifecycle**" refers to the entire process of AI development and deployment, broken down into distinct stages. The stages typically include:

- **Design**: This initial phase involves conceptualizing and designing the AI model or system based on specific needs and objectives.
- **Training**: During training, the designed models learn from vast amounts of data to develop the ability to perform tasks such as recognizing patterns or making decisions.
- **Enhancement**: After training, AI systems may undergo further refinements and enhancements to improve their accuracy, efficiency, and performance.
- **Deployment**: In the final stage, the AI system is deployed in a real-world environment to perform the tasks it was designed for.

A machine learning or AI model, particularly a neural network, can have billions or even trillions of **parameters**. The number of parameters is often added to the name of the model, e.g. "<name> 670B" means that the model has 670 billion parameters. The most advanced AI models are called "**frontier AI models**."

The number of parameters is an important factor in a model's performance, but it is not the only factor. Other factors include the quality of the training data and the model architecture.

Each parameter has a numerical value. Parameters are also often referred to as **weights**. During the **training process**, the machine learning algorithm adjusts these parameters to minimize the difference between its predictions and the actual outcomes. This process is repeated many times and requires processing enormous amounts of data.

The final set of weights, obtained after training, is what gives the machine learning model its predictive power.

Once a model has been trained on a dataset, it is ready for deployment in the real world: it is ready for **inference**, i.e. to infer/deduce new content. The trained model applies what it has learned to make predictions on new, unseen data. For example, each time you enter a prompt into a chatbot, the chatbot generates a response based on its training – this is an inference. Every prompt leads to another inference.

The training process requires intensive **compute**, i.e. computing resources, and for frontier AI models tend to take months on a complex computing infrastructure involving specialized computer chips. In contrast, running a single inference query (e.g. having an AI model respond to a single question) requires much less compute, but the total amount of compute used for inference is still very large, since large AI companies need to run millions of user queries per day.

"**Open source**" commonly refers to software that is made available with its source code accessible to anyone, allowing anyone to inspect, modify and distribute the software. There is not yet a common terminology as to what "open source" means in the context of AI models: companies might release just the source code for training their AI model, or include the weights, or even provide the training data, or they may have restrictions attached to their release. Open-sourcing most typically, though imprecisely, refers to publication of model weights.

Appendix 2: List of participants of the morning session

For the list of participants of the morning session, please refer to the website: https://aiforgood .itu.int/summit24/programme/#day0.

Appendix 3: Acknowledgments and sources

The ITU Secretariat would like to express its sincere gratitude to several experts whose contributions have greatly contributed to AI Governance Day and its curation. A big thank you goes to Robert Trager, Professor at the University of Oxford; Director, Oxford Martin AI Governance Initiative; Senior Research Fellow, Blavatnik School of Government; International Governance Lead, Centre for the Governance of AI, and Anka Reuel, Computer Science Ph.D. student at Stanford University. They both agreed to moderate AI Governance Day. They also pointed out relevant research papers, shared their unpublished research with us, and made their network of experts available to us. Robert Trager has further provided practical suggestions, generous advice, and expert guidance that has been paramount in the planning of AI Governance Day.

We are also deeply appreciative of Sam Daws, Senior Advisor to the Oxford University Martin School AI Governance Initiative, and Director of Multilateral.AI, and to Marta Ziosi, Postdoctoral Researcher at the Oxford Martin AI Governance Initiative, who compiled the examples of multilateral and national AI initiatives listed in chapter 4.3.1.

We are very grateful to Lennart Heim, Lead on AI & Compute at RAND's Technology and Security Policy Center researcher at RAND, and Mauricio Baker, Technology and Security Policy Fellow at RAND, for their generous comments and explanations on compute governance and the review of the vocabulary.

Last but not list we thank all the participants who made their way to Geneva for AI Governance Day. We hope you found it a fruitful experience.

It goes without saying that there are lots of elves from the ITU Secretariat working in the background to make this all happen – thank you all!

Research papers:

- "**Governing Through the Cloud: The Intermediary Role of Compute Providers in AI Regulation**", Lennart Heim, Tim Fist, Janet Egan, Sihao Huang, Stephen Zekany, Robert Trager, Michael a. Osborne & Noa Zilberman; [Whitepaper](#), 13 March 2024, by AI Governance Institute, Oxford Martin School, University of Oxford.
- "**Computing Power and the Governance of Artificial Intelligence**", Girish Sastry, Lennart Heim, Haydn Belfield, Markus Anderljung, Miles Brundage, Julian Hazell, Cullen O'Keefe, Gillian K. Hadfield, Richard Ngo, Konstantin Pilz, George Gor, Emma Bluemke, Sarah Shoker, Janet Egan, Robert F. Trager, Shahar Avin, Adrian Weller, Yoshua Bengio, Diane Coyle, 13 February 2024, URL http://arxiv.org/abs/2402.08797 [cs].
- "**Position Paper: On the Importance of Technical Research and Talent for AI Governance**", Anka Reuel, Lisa Söder, Ben Bucknall, Trond Arne Undheim. March 2024. Accepted to the 2024 International Conference on Machine Learning (ICML).
- "**Open-Sourcing Highly Capable Foundation Models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives**", Elizabeth Seger, Noemi Dreksler, Richard Moulange, Emily Dardaman, Jonas Schuett, K. Wei, Christoph Winter, Mackenzie Arnold, Seán Ó hÉigeartaigh, Anton Korinek, Markus Anderljung, Ben Bucknall, Alan Chan, Eoghan Stafford, Leonie Koessler, Aviv Ovadya, Ben Garfinkel, Emma Bluemke, Michael Aird, Patrick Levermore, Julian Hazell, Abhishek Gupta, 29 September 2023, [2311.09227 (arxiv.org)](#).
- "**What should be Internationalised in Frontier AI Governance?**", March 2024, Draft.

- "**Increased Compute Efficiency and the Diffusion of AI Capabilities**", Konstantin Pilz, Lennart Heim, Nicholas Brown, 13 February 2024, arxiv.org.2311.153772v2.
- "**International Scientific Report on the Safety of Advanced AI**", Interim Report, 17 May 2024.
- **"Facts and Figures 2023 - Internet use (itu.int)"**, International Telecommunication Union.
- **UN System White Paper in AI Governance**: An analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance, United Nations System, 2 May 2024.

Link to the afternoon live recording available here.