SEOUL DECLARATION FOR SAFE, INNOVATIVE AND INCLUSIVE AI BY PARTICIPANTS ATTENDING THE LEADERS' SESSION OF THE AI SEOUL SUMMIT, 21st MAY 2024

1. We, world leaders representing Australia, Canada, the European Union, France, Germany, Italy, Japan, the Republic of Korea, the Republic of Singapore, the United Kingdom, and the United States of America, gathered at the AI Seoul Summit on 21st May 2024, affirm our common dedication to fostering international cooperation and dialogue on artificial intelligence (AI) in the face of its unprecedented advancements and the impact on our economies and societies.

2. Building on the work laid out at the AI Safety Summit held at Bletchley Park in the United Kingdom on November 2023, we recognize that AI safety, innovation, and inclusivity are inter-related goals and that it is important to encompass these priorities in international discussions on AI governance to address the broad spectrum of opportunities and challenges that the design, development, deployment, and use of AI presents and may present.

3. We recognize the importance of interoperability between AI governance frameworks in line with a risk-based approach to maximize the benefits and address the broad range of risks from AI, to ensure the safe, secure, and trustworthy design, development, deployment, and use of AI. We continue to focus on supporting the operationalisation of the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems. We recognize the particular responsibility of organizations developing and deploying frontier AI, and, in this regard, note the *Frontier AI Safety Commitments*.

4. We support existing and ongoing efforts of the participants to this Declaration to create or expand AI safety institutes, research programmes and/or other relevant institutions including supervisory bodies, and we strive to promote cooperation on safety research and to share best practices by nurturing networks between these organizations. In this regard, we welcome the *Seoul Statement of Intent toward International Cooperation on AI Safety Science,* which is annexed to this Declaration.

5. We call for enhanced international cooperation to advance AI safety, innovation and inclusivity to harness human-centric AI to address the world's greatest challenges, to protect and promote democratic values, the rule of law and human rights, fundamental freedoms and privacy, to bridge AI and digital divides between and within countries, thereby contributing to the advancement of human well-being, and to support practical applications of AI including to advance the UN Sustainable Development Goals.

6. We advocate for policy and governance frameworks, including risk-based approaches, that foster safe, innovative and inclusive AI ecosystems. Frameworks should facilitate a virtuous cycle between human creativity and the development and use of AI, promote socio-cultural, linguistic, and gender diversity, and promote environmentally sustainable development and use of technology and infrastructure throughout the life-cycle of commercially and publicly available AI systems.

7. We affirm the importance of active multi-stakeholder collaboration, including governments, the private sector, academia, and civil society to cultivate safe, innovative and inclusive AI ecosystems, and the importance of cross-border and cross-disciplinary collaboration. Recognizing that all states will be affected by the benefits and risks of AI, we will actively include a wide range of international stakeholders in conversations around AI governance.

8. We aim to strengthen international cooperation on AI governance through engagement with other international initiatives at the UN and its bodies, G7, G20, the Organization for Economic Co-operation and Development (OECD), the Council of Europe, and the Global Partnership on AI (GPAI). In this light, we acknowledge the Hiroshima AI Process Friends Group, welcome the recently updated OECD AI principles, and the recent adoption by consensus of the United Nations General Assembly resolution "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development" that solidified the global understanding on the need for safeguards for AI systems and the imperative to develop, deploy, and use AI for good, and welcome discussions on the Global Digital Compact in advance of the Summit of the Future in September 2024 and look forward to the final report of the UN Secretary-General's High-level Advisory Body on AI (HLAB).

9. Acknowledging the value of AI Summit dialogues as a high-level forum to advance discussion on AI governance which facilitates AI safety, innovation and inclusivity, we look forward to our third gathering at the upcoming AI Action Summit to be held in France.

Annex: Seoul Statement of Intent toward International Cooperation on AI Safety Science

Annex

Seoul Statement of Intent toward International Cooperation on AI Safety Science

1. Gathered at the AI Seoul Summit on 21st May 2024, and following on from the AI Safety Summit in Bletchley Park on 2nd November 2023 and acknowledging the Safety Testing Chair's Statement of Session Outcomes from the Bletchley Leaders' Session, world leaders representing Australia, Canada, the European Union, France, Germany, Italy, Japan, the Republic of Korea, the Republic of Singapore, the United Kingdom, and the United States of America affirm the importance of international coordination and collaboration, based in openness, transparency, and reciprocity, to advance the science of AI safety. We affirm that safety is a key element in furtherance of responsible AI innovation.

2. We commend the collective work to create or expand public and/or government-backed institutions, including AI Safety Institutes, that facilitate AI safety research, testing, and/or developing guidance to advance AI safety for commercially and publicly available AI systems.

    2.1 We acknowledge the need for a reliable, interdisciplinary, and reproducible body of evidence to inform policy efforts related to AI safety. We recognize the role of scientific inquiry and the benefits of international coordination for the advancement of such inquiry, so that ultimately the benefits of AI development and deployment are shared equitably around the globe.

    2.2 We affirm our intention to leverage and promote common scientific understandings through assessments such as the *International AI Safety Report,* to guide and align our respective policies, where appropriate, and to enable safe, secure, and trustworthy AI innovation, in line with our governance frameworks.

    2.3 We express our shared intent to take steps toward fostering common international scientific understanding on aspects of AI safety, including by endeavoring to promote complementarity and interoperability in our technical methodologies and overall approaches.

2.4 These steps may include taking advantage of existing initiatives; the mutual strengthening of research, testing, and guidance capacities; the sharing of information about models, including their capabilities, limitations, and risks as appropriate; the monitoring of AI harms and safety incidents; the exchange or joint creation of evaluations, data sets and associated criteria, where appropriate; the establishment of shared technical resources for purposes of advancing the science of AI safety; and the promotion of appropriate research security practices in the field.

2.5 We intend to coordinate our efforts to maximize efficiency, define priorities, report progress, enhance our outputs' scientific rigor and robustness, promote the development and adoption of international standards, and accelerate the advancement of evidence-based approaches to AI safety.

3. We articulate our shared ambition to develop an international network among key partners to accelerate the advancement of the science of AI safety. We look forward to close future collaboration, dialogue, and partnership on these and related endeavors.