



Gobernanza de la Transición hacia la Inteligencia Artificial General (IAG)

Consideraciones Urgentes para el Informe de la Asamblea General de la ONU para el Consejo de Presidentes de la Asamblea General de las Naciones Unidas (UNCPGA).

Resumen Ejecutivo

Los sistemas de IA avanzan rápidamente hacia la inteligencia artificial general (IAG), caracterizada por sistemas capaces de igualar o superar la inteligencia humana en diversas tareas cognitivas. Con las mayores inversiones financieras de la historia impulsando esfuerzos de I+D sin precedentes, los líderes y expertos de la industria anticipan que la IAG podría surgir en esta década, generando beneficios extraordinarios para la humanidad. Entre estos beneficios, la IAG podría acelerar los descubrimientos científicos relacionados con la salud pública, transformar muchas industrias, aumentar la productividad y contribuir a la consecución de los Objetivos de Desarrollo Sostenible.

Sin embargo, la IAG también podría generar riesgos únicos y potencialmente catastróficos. A diferencia de la IA tradicional, la IAG podría ejecutar de forma autónoma acciones dañinas más allá de la supervisión humana, lo que resultaría en impactos irreversibles, amenazas de sistemas de armas avanzados y vulnerabilidades en infraestructuras críticas. Debemos asegurarnos de que estos riesgos se mitiguen si queremos cosechar los extraordinarios beneficios de la IAG.

Para abordar eficazmente estos desafíos globales, es esencial una acción internacional inmediata y coordinada, con el apoyo de las Naciones Unidas. Dichas acciones deberían ser impulsadas por una Asamblea General especial de la ONU, específicamente sobre la IAG, para debatir los beneficios y riesgos de la IAG y el posible establecimiento de un observatorio global de la IAG, un sistema de certificación para una IAG segura y fiable, una Convención de la ONU sobre la IAG y una agencia internacional de la IAG. Sin una gestión global proactiva, la competencia entre naciones y corporaciones acelerará el desarrollo arriesgado de la IAG, socavarán los protocolos de seguridad y exacerbarán las tensiones geopolíticas. La acción internacional coordinada puede prevenir estos resultados, promoviendo el desarrollo y el uso seguros de la IAG, la distribución equitativa de los beneficios y la estabilidad global.

¹ METR: Medición de la capacidad de la IA para completar tareas largas
<https://arxiv.org/abs/2503.14499>



Introducción

El progreso en IA ha sido rápido en los últimos años y meses² y podría acelerarse aún más, en parte porque las empresas de IA están invirtiendo grandes sumas en la creación de agentes de IA más capaces y autónomos, y por el creciente uso de los modelos de IA más potentes para impulsar la propia investigación en IA.³ Se espera ampliamente que estas mejoras en las capacidades de IA conduzcan a la "Inteligencia Artificial General" (IAG): sistemas de IA que igualan o superan el rendimiento humano en la mayoría de las tareas cognitivas.

Si bien existe desacuerdo sobre cuándo se espera la IAG, todos los expertos de este Panel creen que es posible que se desarrolle en esta década. Las empresas de IA están destinando cientos de miles de millones de dólares para lograr la IAG muy pronto, lo que convierte este esfuerzo en el mayor de I+D en la historia de la humanidad. El sector privado tiene la responsabilidad de desarrollar tecnología mucho más segura y debería tener incentivos para hacerlo; pero la carrera competitiva para lograr la IAG primero lo impulsa a centrar todos sus esfuerzos en las capacidades, en lugar de la seguridad, para así "ganar la carrera".

Los riesgos actuales asociados con la IA se derivan principalmente del mal uso de la tecnología por parte de los humanos. Sin embargo, la IA general también presenta un riesgo fundamentalmente diferente, ya que sus amenazas potenciales van más allá del uso indebido por parte de los humanos. La IA general podría generar y ejecutar planes de forma autónoma con resultados catastróficos, superando la capacidad humana para reconocer, analizar y responder a amenazas emergentes y disrupciones sin precedentes^{4,5}. Sumado a la tendencia de autoconservación observada recientemente de las IA avanzadas, esto podría conducir a situaciones en las que la IA general se vuelva incontrolable.

Esto debería ser una preocupación global compartida. Los riesgos relacionados con la IA general no se limitan a industrias o sociedades específicas, sino que tienen implicaciones globales, independientemente de su origen. Garantizar la integración segura y armoniosa de la IA general requiere no solo esfuerzos nacionales o corporativos, sino una gobernanza internacional proactiva, liderada por las Naciones Unidas. Las Naciones Unidas están excepcionalmente cualificadas para facilitar un acuerdo científico sobre riesgos y estrategias de mitigación, generar consenso político en torno a un enfoque compartido para la mitigación de riesgos, coordinar políticas, promover normas o salvaguardas,

² See the International AI Safety Report, Bengio et al 2025.

³ "Claude 3.7 (o1) Knows When it is in Alignment Evaluations"

<https://www.apolloresearch.ai/blog/claude-sonnet-37-o1-knows-when-its-in-alignment-evaluations>

⁴ Véase Meinke et al., 2024, Los modelos de Frontera son capaces de esquematizar en contexto, <https://arxiv.org/abs/2412.04984>

⁵ «Claude 3.7 (o1) sabe cuándo está en evaluación de alineación»

<https://www.apolloresearch.ai/blog/claude-sonnet-37-o1-knows-when-its-in-alignment-evaluations>



responder a emergencias y, potencialmente, realizar o coordinar investigaciones conjuntas sobre seguridad.

Sin una gobernanza global, el potencial transformador de la IA general para abordar los desafíos globales podría verse infrautilizado o mal dirigido. Además, la coordinación global será esencial para gestionar las amenazas catastróficas globales que se prevé que plantee la IAG.



I. Urgencia para que la Asamblea General de la ONU actúe sobre la gobernanza de la IAG y posibles consecuencias si no se actúa

En medio del complejo entorno geopolítico y en ausencia de normas internacionales cohesivas y vinculantes, la prisa competitiva por desarrollar la IAG sin las medidas de seguridad adecuadas está aumentando el riesgo de accidentes o mal uso, militarización y fracasos existenciales⁶. Las naciones y las corporaciones están priorizando la velocidad sobre la seguridad, socavando los marcos de gobierno nacionales y haciendo que los protocolos de seguridad sean secundarios a las ventajas económicas o militares. Dado que muchas formas de IAG de gobiernos y corporaciones podrían surgir antes de que finalice esta década, y dado que establecer sistemas de gobernanza nacionales e internacionales tomará años, es urgente comenzar los procedimientos necesarios para prevenir los siguientes resultados:

1. **Consecuencias irreversibles:** una vez que se logra la IAG, su impacto puede ser irreversible. Con muchas formas fronterizas de IA que ya muestran un comportamiento engañoso y de autopreservación⁷, y el impulso hacia IA más autónomas, interactivas y de automejora integradas con infraestructuras, los impactos y la trayectoria de la IAG pueden plausiblemente terminar siendo incontrolables. Si eso sucede, puede que no haya manera de volver a un estado de supervisión humana confiable. La gobernanza proactiva es esencial para garantizar que la IAG no cruce nuestras líneas rojas⁸, lo que lleva a sistemas incontrolables sin una forma clara de volver al control humano..
2. **Armas de Destrucción Masiva:** La IAG podría permitir que algunos estados y actores no estatales maliciosos construyan armas químicas, biológicas, radiológicas y nucleares. Además, grandes enjambres de armas autónomas letales controlados por la IAG podrían constituir en sí mismos una nueva categoría de armas de destrucción masiva.
3. **Vulnerabilidades Críticas de la Infraestructura:** Los sistemas nacionales críticos (por ejemplo, redes eléctricas, sistemas financieros, redes de transporte, infraestructura de comunicaciones y sistemas de salud) podrían estar sujetos a poderosos ciberataques lanzados por o con la ayuda de la IAG. Sin disuasión nacional y coordinación internacional, actores no estatales maliciosos, desde terroristas hasta el crimen organizado transnacional, podrían llevar a cabo ataques a gran escala.
4. **Concentración de poder, desigualdad global e inestabilidad:** El desarrollo y uso descontrolados de la IAG podrían exacerbar las disparidades de riqueza y poder a

⁶ Respuesta de OpenAI al Plan de acción sobre IA de la Oficina de Política Científica y Tecnológica de EE. UU. <https://cdn.openai.com/global-affairs/ostp-rfi/ec680b75-d539-4653-b297->

⁷ Diálogos internacionales sobre seguridad de la IA (2024): <https://idaia.ai/dialogue/idaia-beijing/>

⁸ Cass-Beggs, Duncan, Stephen Clare, Dawn Dimowo, and Zaheed Kara. 2024. "Framework Convenon on Global AI Challenges." Center for Internaonal Governance Innovaon. <https://www.cigionline.org/publicaons/framework-convenon-on-global-ai-challenges/>.



una escala sin precedentes. Si la IAG permanece en manos de unas pocas naciones, corporaciones o grupos de élite, podría consolidar el dominio económico y crear monopolios globales sobre la inteligencia, la innovación y la producción industrial. Esto podría provocar desempleo masivo, un desempoderamiento generalizado que afecte las bases legales, la pérdida de privacidad y el colapso de la confianza en las instituciones, el conocimiento científico y la gobernanza. Podría socavar las instituciones democráticas mediante la persuasión, la manipulación y la propaganda generada por la IA, y aumentar la inestabilidad geopolítica, incrementando las vulnerabilidades sistémicas. La falta de coordinación podría generar conflictos por los recursos, las capacidades o el control de la IAG, que podrían derivar en una guerra. La IAG pondrá énfasis en los marcos legales existentes: numerosos problemas nuevos y complejos relacionados con la propiedad intelectual, la responsabilidad, los derechos humanos y la soberanía podrían saturar los sistemas jurídicos nacionales e internacionales.

5. **Riesgos existenciales:** La IAG podría utilizarse indebidamente para causar daños masivos o desarrollarse de maneras contrarias a los valores humanos; incluso podría actuar de forma autónoma, sin supervisión humana, desarrollando sus propios objetivos de acuerdo con los objetivos de autopreservación ya observados en las IA de vanguardia actuales. La IAG también podría buscar el poder para garantizar que pueda ejecutar los objetivos que determine, independientemente de la intervención humana. Gobiernos nacionales, expertos destacados y las empresas que desarrollan la IAG han declarado que estas tendencias podrían conducir a escenarios en los que los sistemas de IAG busquen dominar a los humanos. Estas no son hipótesis descabelladas de ciencia ficción sobre un futuro lejano: muchos expertos destacados consideran que todos estos riesgos podrían materializarse en esta década, y sus precursores ya están ocurriendo.² Además, los principales desarrolladores de IA no tienen hasta el momento una propuesta viable para prevenir estos riesgos con un alto grado de certeza.
6. **Pérdida de extraordinarios beneficios futuros para toda la humanidad:** una gestión adecuada. La IA general promete mejoras en todos los campos, para todas las personas, desde la medicina personalizada, la cura del cáncer y la regeneración celular, hasta los sistemas de aprendizaje individualizados, el fin de la pobreza, la lucha contra el cambio climático y la aceleración de los descubrimientos científicos con beneficios inimaginables. Garantizar un futuro tan magnífico para todos requiere una gobernanza global, que comienza con una mayor concienciación global tanto de los riesgos como de los beneficios. Las Naciones Unidas son fundamentales para esta misión.

II. Propósito de la Gobernanza de la ONU para la Transición a la IAG

Dado que la IAG podría desarrollarse en esta década, es imperativo, tanto desde el punto de vista científico como ético, que construyamos estructuras de gobernanza sólidas para prepararnos tanto para los extraordinarios beneficios como para los extraordinarios riesgos que podría conllevar.



El propósito de la gobernanza de la ONU en la transición a la IAG es garantizar que el desarrollo y el uso de la IAG estén alineados con los valores humanos, la seguridad y el desarrollo globales. Esto implica: 1) Impulsar la investigación sobre la alineación y el control de la IA para identificar métodos técnicos para dirigir o controlar sistemas de IA cada vez más capaces; 2) Orientar el desarrollo de la IAG: establecer marcos que garanticen que se desarrolle de forma responsable, con sólidas medidas de seguridad, transparencia y en consonancia con los valores humanos; 3) Desarrollar marcos de gobernanza para la implementación y el uso de la IAG: prevenir el uso indebido, garantizar el acceso equitativo y maximizar sus beneficios para la humanidad, minimizando al mismo tiempo los riesgos. 4) Fomentar visiones futuras de una IAG beneficiosa: nuevos marcos para el desarrollo social, ambiental y económico; y 5) Proporcionar una plataforma neutral e inclusiva para la cooperación internacional: establecer estándares globales, construir un marco legal internacional y crear incentivos para su cumplimiento; así, fomentar la confianza entre las naciones para garantizar el acceso global a los beneficios de la IAG.

III. Sesión de la Asamblea General de la ONU sobre consideraciones clave de la IAG

Uno de los mayores desafíos de la gobernanza de la IAG es la incertidumbre en torno a su futuro desarrollo tecnológico. Esto dificulta predecir con precisión los posibles beneficios y riesgos. En consecuencia, se debe establecer un marco de respuesta amplio e integral para anticipar y mitigar las posibles amenazas, a la vez que se refuerzan los posibles beneficios. Las Naciones Unidas pueden proporcionar una coordinación internacional crucial para el desarrollo y uso de la IAG. Es particularmente importante que todas las naciones estén representadas en este proceso y que se reduzcan las divisiones geopolíticas; actualmente, solo la ONU parece estar bien posicionada para desempeñar este papel. Los siguientes puntos deben considerarse durante una sesión de la Asamblea General de la ONU específicamente sobre la IAG:

A. Observatorio Global de la IAG

Se necesita un Observatorio Global de la IAG para monitorear el progreso de la investigación y el desarrollo relacionados con la IAG y proporcionar alertas tempranas sobre la seguridad de la IA a los Estados Miembros. Este Observatorio debería aprovechar la experiencia de otras iniciativas de la ONU, como el Panel Científico Internacional Independiente sobre IA, creado por el Pacto Digital Mundial, y la Metodología de Evaluación de la Preparación de la UNESCO.

B. Sistema Internacional de Mejores Prácticas y Certificación para una IAG Segura y Confiable

Se necesita un sistema internacional para identificar las mejores prácticas y certificar la seguridad, el desarrollo y el uso de la IAG. La verificación de la alineación de la IAG con los valores humanos, el comportamiento controlado y no engañoso, y el desarrollo seguro es esencial para la confianza internacional.



C. Convención Marco de las Naciones Unidas sobre la IAG

Se necesita una Convención Marco sobre la IAG⁹ para establecer objetivos compartidos y protocolos flexibles para gestionar los riesgos de la IAG y garantizar una distribución equitativa de los beneficios a nivel mundial. Debería definir niveles de riesgo claros que requieran una acción internacional proporcionada, desde el establecimiento de normas y regímenes de licencias hasta centros de investigación conjuntos para la IAG de mayor riesgo, así como líneas rojas o trampas⁹ en el desarrollo de la IAG. Una Convención proporcionaría la base institucional adaptable esencial para una gobernanza globalmente legítima, inclusiva y eficaz de la IAG, minimizando los riesgos globales y maximizando la prosperidad global derivada de ella.

D. Estudio de viabilidad de una agencia de la ONU para la IAG

Dada la amplitud de las medidas necesarias para prepararse para la IAG y la urgencia del problema, es necesario investigar la viabilidad de una agencia de la ONU sobre la IAG, idealmente mediante un proceso acelerado. Se ha sugerido un modelo similar al OIEA, entendiendo que la gobernanza de la IAG es mucho más compleja que la de la energía nuclear y, por lo tanto, requiere consideraciones únicas en dicho estudio de viabilidad.

IV. Estas recomendaciones contribuyen a la implementación del Pacto para el Futuro de las Naciones Unidas y otras iniciativas de la ONU

Múltiples iniciativas de la ONU exigen el desarrollo de una IA segura y fiable. Entre ellas, las resoluciones de la Asamblea General de las Naciones Unidas sobre IA (A/78/L.49, A/78/L.86 y A/C.1/79/L.43), junto con el Pacto para el Futuro de las Naciones Unidas, el Pacto Digital Mundial y la Recomendación de la UNESCO sobre la Ética de la IA, instan a la cooperación internacional para desarrollar una IA beneficiosa para toda la humanidad, gestionando al mismo tiempo de forma proactiva los riesgos globales.

Estas iniciativas han atraído la atención mundial sobre las formas actuales de IA. Este informe se basa en estas iniciativas de la ONU al abordar específicamente el desarrollo de la IAG en el futuro próximo.

⁹ Russell, Stuart, Edson Prestes, Mohan Kankanhalli, Jibu Elias, Constanza Gómez Mont, Vilas Dhar, Adrian Weller, Pascale Fung, and Karim Beguir, "AI red lines: The opportunities and challenges of setting limits." World Economic Forum, 11 March 2025. <https://www.weforum.org/stories/2025/03/ai-red-lines-uses-behaviours/> Karnofsky, Holden. 2024. "A Sketch of Potential Tripwire Capabilities for AI." Carnegie Endowment for International Peace. December 10, 2024. <https://carnegieendowment.org/research/2024/12/a-sketch-of-potential-tripwire-capabilities-for-ai?lang=en>.



Los compromisos asumidos por el Pacto para el Futuro se promueven de diversas maneras en este informe. Una sesión de la Asamblea General de las Naciones Unidas centrada en la IAG responde al compromiso del Pacto para el Futuro con el diálogo global sobre la gobernanza de la IA. Recomendaciones sobre una Convención Marco de las Naciones Unidas sobre la IAG y un estudio de viabilidad para una agencia de la ONU dedicada a la IAG. El Observatorio que hemos propuesto apoyaría la labor del próximo Panel Científico Internacional Independiente sobre IA, uno de los resultados clave del Pacto Digital Mundial. Finalmente, el Sistema Internacional de Mejores Prácticas y Certificación para una IAG Segura y Confiable contribuiría a la confianza y la transparencia, como lo exigen las Resoluciones de la Asamblea General de las Naciones Unidas, la UNESCO y el Pacto para el Futuro.

V. Conclusión

Es urgente concienciar a los líderes nacionales e internacionales sobre los beneficios y riesgos de la futura IAG, a diferencia de las formas actuales de IA. La Asamblea General de las Naciones Unidas es el foro adecuado para iniciar dicho debate global.

Se requerirá la coordinación internacional del desarrollo y el uso de la IAG para aprovechar sus extraordinarios beneficios, a la vez que se salvaguardan los derechos humanos y la seguridad. Este Panel de la IAG recomienda firmemente que la Asamblea General de la ONU actúe con urgencia para abordar estas cuestiones durante una sesión de la Asamblea General, específicamente para establecer un marco de gobernanza global para la IAG. Sin estas medidas, los riesgos de un desarrollo y uso incontrolados de la IAG —que abarcan desde un aumento drástico de la desigualdad global hasta amenazas existenciales— son inmensos. Este enfoque liderado por la ONU, que incluye un observatorio global, una certificación internacional, una Convención de la ONU sobre la IAG y una agencia dedicada a la IAG, aumenta la probabilidad de que el desarrollo y el uso de la IAG beneficien a toda la humanidad, minimizando al mismo tiempo los riesgos. Este marco debe ser inclusivo, transparente y aplicable para fomentar la confianza y la cooperación entre las naciones.



Apéndice

Términos de Referencia: Panel de Alto Nivel sobre Inteligencia Artificial General (IAG) para el Consejo de Presidentes de la Asamblea General de las Naciones Unidas (UNCPGA)

Contexto

La Declaración de Seúl 2024 de la UNCPGA solicita la creación de un panel de expertos en inteligencia artificial general (IAG) para proporcionar un marco y directrices que la Asamblea General de las Naciones Unidas considere al abordar las cuestiones urgentes de la transición a la IA.

Este trabajo debe aprovechar, y evitar la duplicación, los amplios esfuerzos sobre los valores y principios de la IA realizados por la UNESCO, la OCDE, el G20, el G7, la Alianza Mundial sobre IA y la Declaración de Bletchley, así como las recomendaciones del Órgano Asesor de Alto Nivel sobre IA del Secretario General de las Naciones Unidas, el Pacto Digital Mundial de las Naciones Unidas, la Red Internacional de Institutos de Seguridad de la IA, la Convención Marco del Consejo Europeo sobre IA y las dos Resoluciones de la Asamblea General de las Naciones Unidas sobre IA. Estas se han centrado más en formas más específicas de IA. Actualmente, no se presta la misma atención a la IAG.

La IA es bien conocida en todo el mundo y se utiliza con frecuencia, pero la IAG no existe ni existe todavía. Muchos expertos en IAG creen que podría lograrse en un plazo de uno a cinco años y que, con el tiempo, podría evolucionar hacia una superinteligencia artificial que escapa a nuestro control. No existe una definición universalmente aceptada de IAG, pero la mayoría de los expertos coinciden en que se trata de una IA de propósito general capaz de aprender, editar su código y actuar de forma autónoma para abordar numerosos problemas novedosos con soluciones innovadoras similares o superiores a las capacidades humanas. La IA actual no posee estas capacidades, pero la trayectoria de los avances técnicos apunta claramente en esa dirección.

El Pacto Digital Mundial de las Naciones Unidas insta a un Diálogo Mundial sobre la gobernanza de la IA en el seno de las Naciones Unidas. Expertos del sector privado en IAG han destacado la urgente necesidad de un diálogo global para comprender mejor las oportunidades y los riesgos de la IAG. Una Sesión Especial de la Asamblea General de las Naciones Unidas sobre IAG es probablemente la forma más rápida, rentable y con menor impacto para impulsar dicho diálogo.

Propósito

En respuesta a la Declaración de Seúl 2024 de la UNCPGA, elaborar un informe inicial para el Presidente de la UNCPGA y sus miembros para la reunión de la UNCPGA del 8 al 10 de abril de 2025 en Bratislava.



El informe debe identificar los riesgos, las amenazas y las oportunidades de la IAG. Debe centrarse en concienciar sobre la necesidad de movilizar a la Asamblea General de la ONU para abordar la gobernanza de la IAG de manera más sistemática. Se centra en la IAG que aún no se ha logrado, en lugar de en las formas actuales de sistemas de IA más limitados. Debe enfatizar la urgencia de abordar los problemas de la IAG lo antes posible, dado el rápido desarrollo de la IAG, que puede presentar graves riesgos para la humanidad, así como extraordinarios beneficios para ella.

El informe también debe incluir acuerdos multilaterales y acciones del sector privado para abordar estos desafíos sin precedentes. Debe responder a los líderes del sector privado en materia de IAG que piden coordinación internacional y acción multilateral ante lo que podría ser el desafío de gestión más difícil que la humanidad haya enfrentado jamás.

Procedimientos

- Convocar un panel de alto nivel (de 5 a 8 miembros) de expertos internacionales en IAG sobre las posibles amenazas de la IAG para la humanidad y las oportunidades que esta podría beneficiar a la humanidad, así como sobre cuestiones políticas relacionadas.
- El panel sobre IAG se reunirá virtualmente de forma regular a partir de enero de 2025 y elaborará el informe inicial para la próxima reunión de la UNCPGA en Bratislava, en la primavera de 2025.
- Con base en los comentarios sobre el informe inicial durante la reunión de la UNCPGA en Bratislava, el panel finalizará el informe y lo presentará al Secretario General de la UNCPGA. Si el Presidente de la UNCPGA lo acepta, se remitirá al Presidente de la Asamblea General de la ONU, tentativamente, antes del 1 de mayo de 2025.



Miembros del Panel Independiente de Alto Nivel sobre IAG para el Consejo de Presidentes de la Asamblea General de la ONU

Jerome Glenn (EE. UU.), Presidente

Miembro con derecho a voto del IEEE sobre Gobernanza Organizacional de la IA; autor del documento Horizonte 2025-27 de la Unión Europea sobre IAG: Problemas y Oportunidades; Director Ejecutivo del Proyecto del Milenio y autor de sus artículos "Problemas de Gobernanza Internacional en la Transición de la Inteligencia Artificial Estrecha a la IAG", "Requisitos para la Gobernanza Global de la IAG" y "Trabajo/Tecnología 2050: Escenarios y Acciones". Autor de "Mente Futura: Inteligencia Artificial" (1989).

Renan Araujo (Brasil)

Director de Investigación del Instituto de Política y Estrategia de IA, especializado en la gestión de riesgos relacionados con el desarrollo de la IAG. Actualmente, dirige el trabajo del IAPS sobre gobernanza internacional de la IAG. Es becario del Oxford China Policy Lab, abogado, cofundador de la Iniciativa Cóndor (que conecta a estudiantes brasileños con oportunidades de primer nivel para influir en la investigación y las políticas de IA) y trabajó en programas de gobernanza de IA en Rethink Priorities y el Instituto de Derecho e IA.

Yoshua Bengio (Canadá)

Profesor de informática en la Universidad de Montreal; Presidente del Grupo Asesor de Seguridad de la IA del gobierno canadiense; Presidente del Informe Internacional sobre Seguridad de la IA encargado por 30 países, además de la ONU, la OCDE y la UE. Director científico de Mila, el Instituto de IA de Quebec; Miembro del Consejo Asesor Científico del Secretario General de la ONU para Avances en Ciencia y Tecnología; galardonado con el Premio Turing y actualmente el informático más citado a nivel mundial.

Joon Ho Kwak (República de Corea)

Asesor técnico del Instituto Coreano de Seguridad de la IA; desempeñó un papel destacado en el desarrollo de las Directrices de la OCDE para el Desarrollo de una IA Confiable. Participante en el Proceso de Hiroshima del G7, en los preparativos de la Cumbre de Acción sobre IA de París, en el Grupo de Trabajo de IA Corea-EE. UU. y miembro de la delegación coreana ante las Redes Internacionales de Institutos de Seguridad de IA.



Lan Xue (China)

Presidente del Comité Nacional de Expertos en Gobernanza de IA; Decano del Instituto de Gobernanza Internacional de IA de la Universidad de Tsinghua; miembro del Grupo Asesor de la Dirección de Ciencia, Tecnología e Innovación de la OCDE; asesor del Instituto de Seguridad de IA de China; Copresidente del Consejo de Liderazgo de la Red de Soluciones para el Desarrollo Sostenible de las Naciones Unidas (UNSDSN); galardonado con el Premio Fudan a la Contribución Distinguida en Ciencias de la Gestión y con el Premio a la Contribución Distinguida de la Asociación China de Ciencias de la Ciencia y la Política Científica y Tecnológica.

Stuart Russell (Reino Unido y EE. UU.)

Profesor Distinguido de Ciencias de la Computación y Director del Centro de IA Compatible con los Humanos de la Universidad de California, Berkeley. Autor de "Inteligencia Artificial: Un Enfoque Moderno", el libro de texto estándar sobre IA utilizado en 1500 universidades de 135 países y citado más de 74 000 veces; Copresidente del grupo de expertos de la OCDE sobre el futuro de la IA y del Consejo Global de IA del Foro Económico Mundial.

Jaan Tallinn (Estonia)

Miembro del Órgano Asesor de IA de la ONU; formó parte del Grupo de Expertos de Alto Nivel sobre IA de la Comisión Europea; Cofundador del Centro para el Estudio del Riesgo Existencial de la Universidad de Cambridge y del Instituto para el Futuro de la Vida (ambas instituciones son líderes en temas de IA general); Miembro de la Junta Directiva del Centro para la Seguridad de la IA; Inversionista estonia en seguridad de la IA general; Ingeniera fundadora de Skype y FastTrack/Kazaa; y directora inversora fundadora de DeepMind Google.

Mariana Todorova (Bulgaria)

Representante búlgara en el Grupo Intergubernamental de la UNESCO sobre Marcos Éticos de IA; portavoz principal sobre IA general en los medios de comunicación búlgaros. Autor y conferenciante de renombre internacional sobre las dimensiones éticas y tecnológicas de la IA y la IAG; exdiputado y asesor del Presidente de la República de Bulgaria.

José Jaime Villalobos (Costa Rica)

Responsable de Gobernanza Multilateral en el Future of Life Institute; Investigador Asociado Sénior del Centro para la Innovación en Gobernanza Internacional; Investigador Afiliado de la Iniciativa de Gobernanza de IA Oxford Martin; Investigador Afiliado del Instituto de Derecho e IA; Doctor en Derecho Internacional; y coautor de libros y artículos destacados sobre gobernanza internacional de la IA.