IMPULSANDO LA

SEGURIDAD DE LA IA

DESDE LA PERSPECTIVA DE AMÉRICA LATINA Y EL CARIBE

AISI | AI SECURITY INSTITUTE

EON INSTITUTE

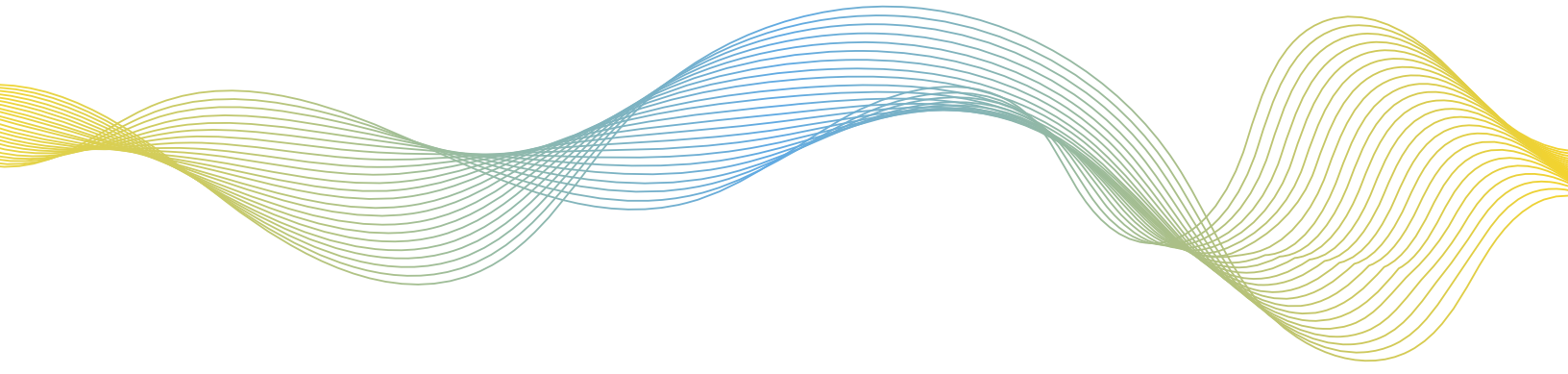
IMPULSANDO LA SEGURIDAD DE LA IA DESDE LA PERSPECTIVA DE AMÉRICA LATINA Y EL CARIBE

Enero 2026

Autoras
Daniela Rojas Arroyo, Directora Adjunta de Programas y Claudia May Del Pozo, Directora y Fundadora, Eon Institute

Coordinadoras del evento
Daniela Rojas Arroyo, Directora Adjunta de Programas y Julieta Sierra Banegas, Asistente de Programas Jr, Eon Institute

Diseño
Nicté Cabañas Segura, Asistente de Proyectos



ÍNDICE

01
INTRODUCCIÓN

01 INTRODUCCIÓN

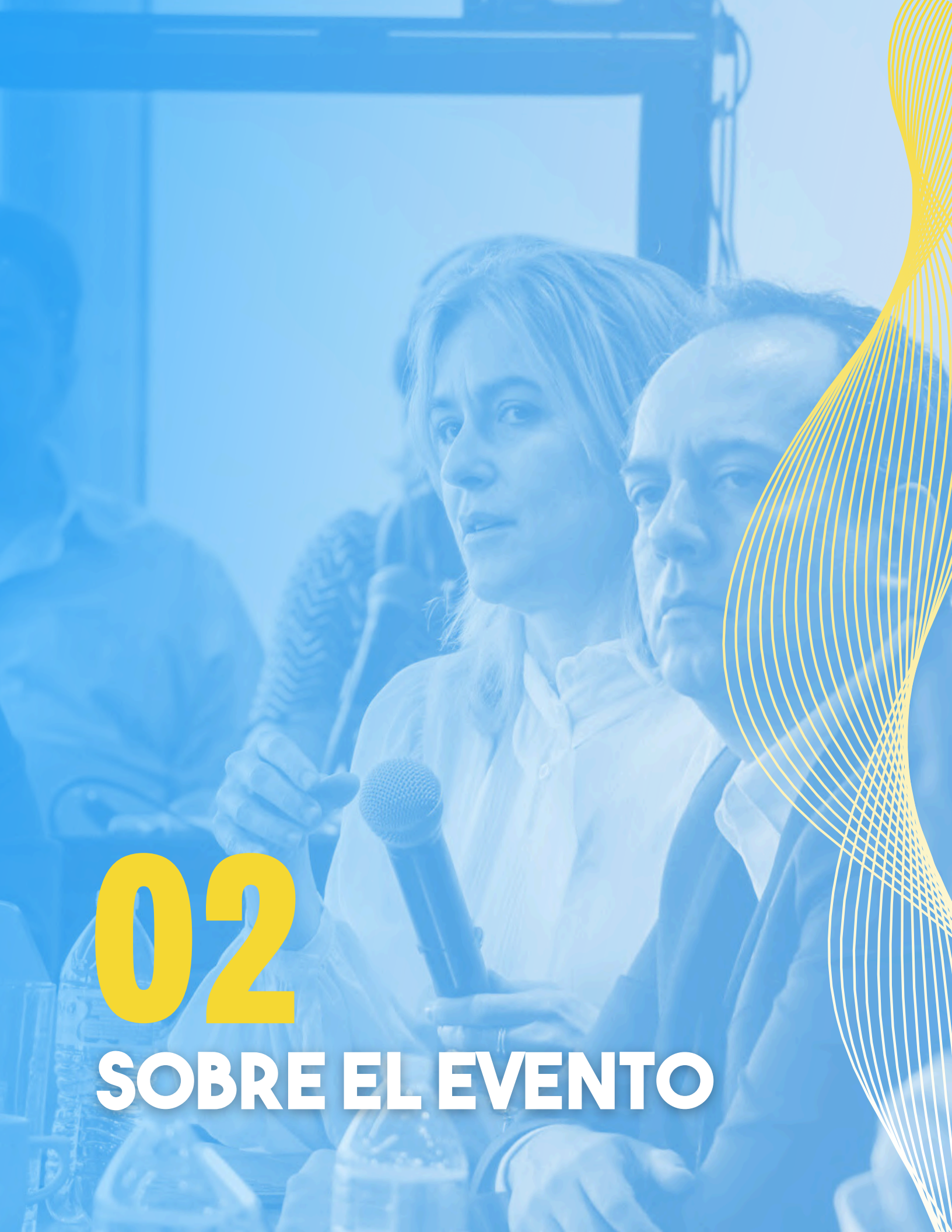
El Informe internacional sobre la seguridad de la IA de 2025, publicado en enero del mismo año, representa una referencia en la evaluación científica sobre los riesgos asociados a la Inteligencia Artificial (IA) avanzada, en particular la IA de propósito general (es decir, aquella capaz de realizar una amplia variedad de tareas), y sobre los métodos disponibles para mitigarlos. Además de ser un documento técnico, el Informe funciona como una valiosa fuente de información para las personas formuladoras de políticas públicas, ya que muestra las lagunas de evidencia y los principales desafíos que requieren nuestra atención urgente.

En 2024, el 40% de las empresas de América Latina y el Caribe (ALC) adoptaron algún sistema de IA (Borgeaud, 2025), pero, al mismo tiempo, ocupó el séptimo lugar entre nueve regiones globales en el Índice de Preparación Gubernamental para la IA 2024 de Oxford Insights. Este contraste evidencia una brecha cada vez mayor en la implementación: mientras que la adopción se acelera, las capacidades institucionales y de gobernanza no han evolucionado al mismo ritmo.

En este contexto, si bien el Informe internacional sobre la seguridad de la IA ofrece una visión global completa del estado actual de la IA de propósito general, también es importante complementarlo con perspectivas regionales que reflejen cómo se experimentan estas realidades en lugares como ALC. El contexto social, político y económico distintivo de la región, caracterizado por la desigualdad, la fragilidad institucional y el acceso desigual a la innovación (UNDP, 2021), da forma a un panorama diferente de oportunidades y riesgos con la IA de propósito general. El avance de la seguridad de la IA en ALC no sólo requiere crear conciencia sobre los últimos avances en los riesgos de la IA, sino también comprender cómo los contextos locales específicos de la región interactúan con estos desafíos globales, con el objetivo de también incluir estas realidades en las conversaciones globales.
La Secretaría del Informe y Eon Institute, un think tank tecno social mexicano, con el apoyo de la Embajada Británica en la

Ciudad de México, organizaron el primer evento en el extranjero dedicado a fortalecer las perspectivas de ALC sobre el debate alrededor de la seguridad de la IA. Celebrado en la Ciudad de México en septiembre de 2025, el evento reunió a 22 especialistas de todos los sectores de 6 países de ALC. En conjunto, las personas participantes no solo reflexionaron sobre los hallazgos del Informe internacional sobre la seguridad de la IA de 2025, sino que también compartieron sus experiencias, contextos y conocimientos para enriquecer la conversación.

Alineado con el objetivo del Informe de ayudar a evaluar, anticipar y gestionar los riesgos asociados a los sistemas de IA de propósito general, este documento cumple una doble función. Por un lado, recoge las discusiones de las personas participantes, cuyos resultados informarán futuras ediciones del Informe. Por otro lado, contribuye a la conversación global al amplificar las perspectivas y experiencias de una región que sigue estando en gran medida subrepresentada: América Latina y el Caribe.



02

SOBRE EL EVENTO

02

SOBRE EL EVENTO

Dado el objetivo de reflejar la diversidad de perspectivas de la región de ALC, el evento reunió a 22 especialistas de IA (de los cuales 40% fueron mujeres) provenientes de ocho países: Argentina, Chile, Colombia, Costa Rica, México, Perú, República Dominicana y Uruguay[1], representando gobierno, academia, sociedad civil, y el sector privado.

El evento se estructuró tanto para presentar como para analizar críticamente el Informe internacional sobre la seguridad de la IA de 2025. En una primera fase, las personas participantes recibieron una presentación de los hallazgos globales del Informe a cargo de Daniel Privitera, autor principal del Informe, y de Lucía Velasco, integrante del equipo de redacción. Las presentaciones se centraron en las tres categorías principales de riesgos de la IA de propósito general identificadas en el Informe: riesgos del uso malicioso, riesgos derivados del mal funcionamiento y riesgos sistémicos

los principales aprendizajes, lo que fomentó un diálogo más amplio entre todas las personas participantes. El resultado es un conjunto de hallazgos que están basadas tanto en la comprensión internacional, como en las realidades regionales, que subrayan en qué puntos los enfoques globales se alinean, o divergen, respecto a las necesidades de ALC.

Las conversaciones que tuvieron lugar durante el evento constituyen la base del análisis y de las conclusiones que se presentan en las siguientes secciones.

[1] TLos países participantes fueron seleccionados en función de su clasificación en el Índice de Preparación Gubernamental para la IA 2024 de Oxford Insight. Brasil no fue incluido en este evento por barreras lingüísticas.

03

PANORAMA DE LA IA DE PROPÓSITO GENERAL EN ALC

03

PANORAMA DE LA IA DE PROPÓSITO GENERAL EN ALC

En lo que respecta a la IA, América Latina y el Caribe presenta un panorama muy heterogéneo. En ella se encuentran algunos de los países más avanzados en la materia, como Brasil, Chile y México, que se sitúan entre los 50 primeros a nivel mundial (Oxford Insights, 2025), junto con otros que se encuentran en las primeras fases de preparación de la IA. Esta sección explora la adopción, la infraestructura, la innovación, la demanda, la aceptación, y las habilidades para brindar un panorama completo de la IA en ALC.

A pesar de que la adopción de la IA por parte de las empresas de ALC alcanza casi un 40% (Borgeaud, 2025), la tasa de adopción entre la población en general aún es limitada. Según un estudio reciente del Instituto de la Economía de la IA de Microsoft, la mayoría de los países de ALC reportaron tasas de adopción de entre el 10% y el 20% en la población en edad laboral, con algunos países (Colombia, Costa Rica y Uruguay) llegando a tasas de adopción de hasta el 30%, y otros reportando menos del 10% (Guatemala y Venezuela). En contraste, países como los Emiratos Árabes Unidos y Singapur reportaron hasta un 50% de adopción. Esto se debe en parte a las considerables disparidades de acceso en la región. Para contextualizar, menos de la mitad de la población cuenta con banda ancha fija, sólo el 9.9% cuenta con internet de alta velocidad en sus hogares (World Bank, 2021), y únicamente 4 de cada 10 zonas rurales en ALC tienen acceso a internet básico (UNDP, 2024, lo cual no es suficiente para interactuar con sistemas de IA públicos. La brecha de adopción no solo está relacionada con el acceso a la tecnología, sino que también refleja estructuras de inequidad regionales más profundas vinculadas con la educación, la innovación, y la infraestructura digital.

[2] Dado que los datos sobre la IA de propósito general siguen siendo limitados, este análisis también se basa en los datos disponibles relacionados con los sistemas tradicionales de IA.

En cuanto a infraestructura, la diferencia es aún más marcada. La región se queda muy atrás, ya que solo atrajo 2.6 miles de millones de dólares en el 2023, lo que equivale únicamente al 1.56% de la demanda mundial (Jung and Kats, 2025). En términos de innovación, los datos relacionados con la inversión internacional muestran que ALC solo recibe 1.12% de esta, a pesar de que la región es responsable del 6.6% del PIB mundial (ECLAC, 2025). Esto plantea un desafío importante a la hora de poner en marcha y hacer crecer iniciativas productivas, tecnológicas e innovadoras.

En relación con la capacidad de la región para desarrollar IA, según la tercera edición del Índice Latinoamericano de Inteligencia Artificial muestra que la formación avanzada sobre IA permanecen insuficientes y están concentradas en un pequeño número de países. Desde 2022, la brecha de talento ha aumentado en relación con la media mundial, lo que se asocia con la acelerada fuga de cerebros de especialistas. Otro de los hallazgos del índice es que la alfabetización en IA ha duplicado la disponibilidad de formación profesional y cuadruplicado el talento especializado. Sin embargo, a pesar del crecimiento de los programas escolares y los cursos de posgrado, la falta de formación avanzada sigue siendo un obstáculo para la capacidad de la región de desarrollar sus propias soluciones.

Curiosamente, a pesar de la imagen desafiante anterior, la demanda por herramientas de IA en ALC está despegando rápidamente. La región representa el 14% de las visitas globales de las soluciones de IA, una cifra notable al considerar que solo solo representa el 11% de las personas usuarias de internet en todo el mundo. Esta diferencia indica un gran interés por las soluciones de IA, con algunos países como Brasil, Chile y México que cuentan con tasas de aceptación superiores al 65% (Ipsos, 2025).

Más allá de las cifras, el panorama de cómo la IA se está adoptando e impactando en toda la región sigue estando incompleto, sobre todo en lo que respecta a la IA de propósito general. Aún persisten importantes lagunas de conocimiento sobre quién está adoptando y desarrollando estos sistemas, en qué sectores se están implementando, qué tipos de valor crean, qué retos específicos afronta ALC en comparación con otras regiones, y qué lecciones se pueden extraer de estas experiencias. Este informe contribuye a visibilizar estas brechas, ofreciendo un primer paso hacia una comprensión más clara del panorama de la IA de propósito general en la región.

[3] Un índice que muestra el progreso de la IA en términos de preparación, adopción y gobernanza en 19 países de América Latina y el Caribe, liderado por el Centro Nacional de Inteligencia Artificial (CENIA) de Chile. El director del CENIA, Álvaro Soto, también formó parte del grupo de personas expertas del Informe Internacional sobre la seguridad de la IA de 2025.

04

RIESGOS Y DESAFÍOS DE LA IA DE PROPÓSITO GENERAL EN ALC

04

RIESGOS Y DESAFÍOS DE LA IA DE PROPÓSITO GENERAL EN ALC

Los riesgos presentados a continuación surgieron de los debates entre las personas participantes, durante las diferentes actividades del evento, y se categorizaron según la clasificación del informe original.

4.1 RIESGOS DEL USO MALICIOSO

El Informe internacional de seguridad de la IA define a los riesgos de uso malicioso como aquellos que "actores maliciosos pueden usar la IA de propósito general para causar daño a individuos, organizaciones o a la sociedad". El Informe identifica cuatro áreas principales de preocupación:

1. Daño a individuos a través de contenido falso
2. Manipulación de la opinión pública
3. Delitos cibernéticos
4. Ataques biológicos y químicos

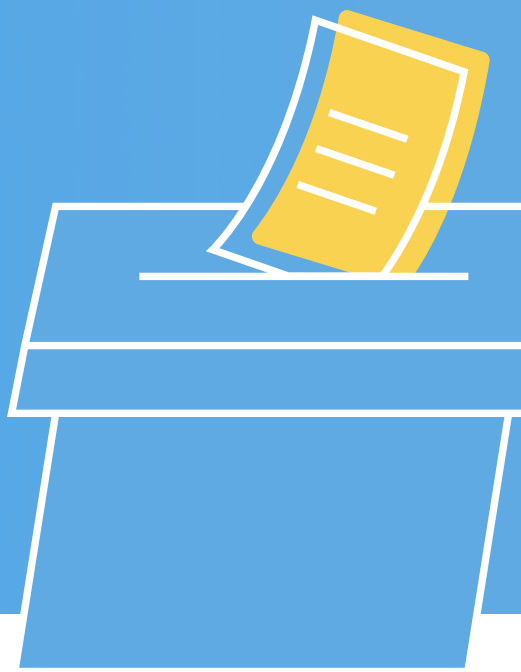
Información falsa, desinformación y democracia:

La preocupación compartida más frecuentemente entre las personas participantes en la discusión de ALC fue el aumento en el uso de la IA de propósito general para generar y difundir información falsa y desinformación, especialmente durante las elecciones y otros procesos democráticos. Dado que ocho países de la región han llevado o llevarán a cabo elecciones presidenciales en el ciclo electoral de 2025-2026, no es de extrañar que este riesgo se encuentre entre los principales. Las personas participantes señalaron que el uso de los deepfakes o ultrafalsos para crear texto debilita la confianza pública en los sistemas de información, y aunque las plataformas de redes sociales han mejorado sus filtros de contenido, sigue siendo complicado controlar material íntimo o político sensible. Además, enfatizaron que este desafío se ve agravado por las vulnerabilidad

preexistentes en ALC, donde se mencionaron los escándalos electorales, la polarización y el monopolio de los medios de comunicación, la escasa capacidad institucional para detectar o contrarrestar los contenidos sintéticos y los bajos niveles de alfabetización digital entre la población en general. El incidente de la campaña para la alcaldía de Bogotá de 2023, que se resume a continuación, ilustra los riesgos reales y crecientes que plantea el uso malicioso de la IA de propósito general en América Latina y el Caribe.

ELECCIONES DE BOGOTÁ 2023

En la recta final de las elecciones locales de 2023 en Colombia, la campaña por la alcaldía de Bogotá se vio perturbada por la circulación en redes sociales de grabaciones de audio falsas creadas con IA. Estas mostraban supuestas confesiones y acuerdos ilícitos que involucraban varios candidatos. Aunque algunas de estas grabaciones de audio se identificaron fácilmente como falsificaciones baratas (cheapfakes) o falsificaciones superficiales (shallowfakes) debido a su baja calidad, estas se difundieron ampliamente en plataformas como WhatsApp, TikTok y Twitter, generando confusión a la opinión pública. Ante la dificultad para verificar su autenticidad, varios políticos comenzaron posteriormente a utilizar a la IA como excusa para desacreditar las acusaciones en su contra, alegando que sus voces habían sido clonadas. Este fenómeno puso en evidencia los riesgos de la desinformación generada por la IA en los procesos electorales, así como la falta de herramientas eficaces para detectar este tipo de falsificaciones.



Violencia política y por razón de género en línea:

Muchas de las personas advirtieron que la información falsa y la desinformación generada por IA podría intensificar la violencia política y por razón de género en línea. Este tipo de violencia ha incrementado a medida que ha aumentado el número de candidatas mujeres en varios de los países de ALC. Una de las manifestaciones más dañinas que se discutió por las personas participantes fue alrededor de la creación y distribución de material no consensuado de deepfakes. Se estima que el 98 % de estos casos involucra imágenes íntimas sintéticas, y que el 99 % de las víctimas son mujeres y personas jóvenes. A pesar de que el número es cada vez mayor de países en la región, incluidos Argentina, Bolivia, Colombia, Guatemala, Honduras, México y Panamá, que han adoptado, o están en proceso de adopción de un marco legal (Ley Olimpia) para hacer frente a la violencia digital por razón de género, a menudo estas leyes están fragmentadas, se implementan incorrectamente o rara vez se aplican a casos que involucran material sintético o las tecnologías de IA de propósito general. El siguiente cuadro resume dicho marco legal.

LEY OLIMPIA

La "Ley Olimpia" es un conjunto de reformas legislativas orientadas a combatir la violencia digital, en particular aquellos delitos que vulneran la privacidad sexual de las personas a través de plataformas en línea, comúnmente denominados ciber-violencia. Estas reformas tipifican como delito la difusión no consentida de contenido sexual, al considerarla una violación a la privacidad. Desde la adopción en México, la iniciativa ha sido implementada en las 31 entidades federativas del país y ha inspirado legislaciones similares en otros países. Argentina, Colombia y Panamá, han aprobado recientemente sus propias versiones de la Ley Olimpia, mientras que movimientos feministas en Bolivia, Ecuador, Guatemala y Honduras impulsan activamente reformas comparables en sus respectivos países.

Ciberataques, fraudes y estafas:
Muchas Otra de las preocupaciones recurrentes entre las personas participantes fue el potencial de actores maliciosos de aprovechar los sistemas de IA para automatizar los ciberataques, diseñar esquemas de phishing sofisticado, y crear deepfakes para estafas. Por ejemplo, el reporte del Panorama de Amenazas de Kaspersky de 2024 revela que las estafas realizadas con IA han aumentado un 140 % en toda la región (OECD, n.d.). Estas prácticas difuminan aún más la línea entre el contenido legítimo y el fraudulento, lo que dificulta a la ciudadanía discernir la información confiable. Las personas participantes también enfatizaron la creciente amenaza de ciberataques a gran escala dirigidos contra infraestructuras críticas, que, en ausencia de medidas sólidas de ciberseguridad, pueden afectar gravemente la estabilidad de los países, sus economías y el funcionamiento de servicios públicos esenciales. De hecho, en 2024, ALC registró la mayor tasa de informes de ciberincidentes, con un aumento promedio del 25 % anual durante la última década (Diao & Cobos, 2024).

Las personas participantes destacaron varias causas de fondo y factores habilitantes, en particular marcos regulatorios débiles, una coordinación regional limitada para responder a amenazas cibernéticas transfronterizas y una inversión insuficiente en iniciativas locales de verificación de información (fact-checking) y tecnologías de detección. Hubo un amplio consenso en que superar estos desafíos requiere tanto salvaguardas técnicas como estrategias de largo plazo, incluidas la educación pública, una colaboración más estrecha con la sociedad civil y una mayor transparencia por parte de las grandes empresas tecnológicas que operan en la región.

4.2 RIESGOS POR FALLOS

El Informe internacional de seguridad de la IA define a los riesgos por fallos como los daños no intencionados realizados por la IA de propósito general. "Incluso cuando los usuarios no tienen intención de causar daño, pueden surgir riesgos graves debido al mal funcionamiento de la IA de propósito general". El Informe identifica cuatro áreas principales de preocupación:

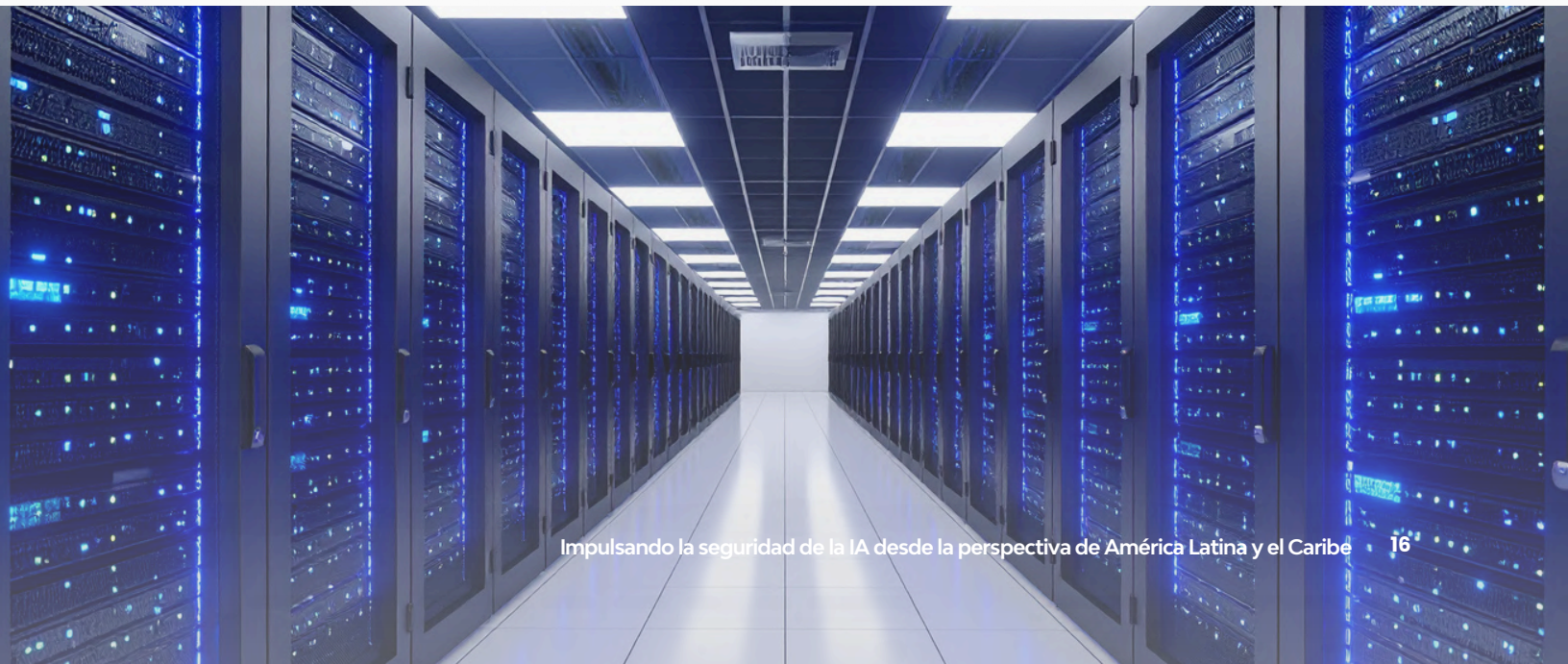
1. Problemas de fiabilidad
2. Sesgos
3. Pérdida de control

Contexto:

Uno de los principales desafíos de la IA de propósito general en ALC es que la mayoría de los modelos base no están diseñados para el contexto de la región. Su implementación exitosa requiere de un ajuste fino (fine-tuning), re-entrenado con datos socioculturales relevantes, y con una validación de comunidades que diversos niveles de alfabetización, lenguas, procedencias y realidades sociales. Sin esta adaptación, estos sistemas podrían reforzar los sesgos, y excluir a grupos completos de personas usuarias. Los desafíos específicos identificados por las personas participantes incluyen:

Disponibilidad de datos:

La disponibilidad de datos sigue siendo una limitación importante. Los países de ALC generan y gestionan muchos menos conjuntos de datos que otras regiones, y gran parte de los datos existentes está fragmentada, desactualizada o es inaccesible. Esta escasez, agravada por infraestructuras débiles, una interoperabilidad limitada y marcos de gobernanza de datos insuficientes, restringe el desarrollo de sistemas de IA adaptados a las necesidades locales. La ausencia de estándares compartidos sobre la calidad de los datos y su uso responsable debilita aún más la equidad, la transparencia y la rendición de cuentas.



Ya que la mayoría de los sistemas de IA están entrenados principalmente con bases de datos del idioma inglés, a menudo fallan a la hora de reconocer dialectos regionales, lenguas indígenas, o referencias culturales específicas. Efectivamente, el inglés sigue siendo el idioma predominante en línea, ya que aparece en el 50% de los sitios web globales, mientras que el español, ocupa el segundo lugar, con aproximadamente un 6% del contenido en línea (Petrosyan, 2025). Esto no solo limita la precisión de las salidas, sino que también reduce la legitimidad percibida y la accesibilidad de estas herramientas para millones de personas. Las personas participantes destacaron la importancia de invertir en modelos que reflejen la diversidad lingüística y cultural de la región e hicieron hincapié en que la inclusión es un prerrequisito para la seguridad y la confianza. Asimismo destacaron algunos esfuerzos actuales para satisfacer esta necesidad, como es el caso de Latam-GPT, resumido en el siguiente cuadro.

LATAM-GPT

Latam-GPT es un Modelo de Lenguaje de Gran Tamaño (LLM por sus siglas en inglés) desarrollado en América Latina y para América Latina. El proyecto tiene como objetivo crear un modelo de IA abierto, gratuito y colaborativo que refleje la diversidad lingüística y cultural de la región. Tras dos años de desarrollo, ha logrado articular la participación de ciudadanía, comunidad investigadora y gobiernos de distintos países de la región. La iniciativa está liderada por el Centro Nacional de Inteligencia Artificial de Chile (CENIA), con el apoyo del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación de Chile y el financiamiento del Banco de Desarrollo de América Latina y el Caribe (CAF).

A través de 33 alianzas estratégicas con organizaciones de toda la región, el proyecto ha recopilado un corpus que supera los 8 terabytes de texto. Este conjunto de datos ha permitido el desarrollo de un modelo de 50 mil millones de parámetros, comparable en escala a GPT-3.5, y capaz de realizar tareas complejas como razonamiento, traducción y asociación semántica. La base de datos regional de Latam-GPT abarca información de 21 países de América Latina e incluye más de 2.6 millones de documentos, con una tasa promedio de cumplimiento del 59.5 %.

Dependencia externa:
Las personas participantes también expresaron su preocupación por la creciente dependencia de modelos de IA de propósito general y de infraestructuras de computación en la nube extranjeras. La dependencia excesiva de tecnologías externas no limita únicamente la capacidad de la región para auditar o interpretar errores, sino que también incrementa la vulnerabilidad frente a procesos de toma de decisión opacos y a sesgos no internacionales integrados en sistemas importados. Las personas participantes subrayan que, en la actualidad, existen pocas alternativas confiables para enfrentar esta dependencia. La región opera bajo un entorno geopolítico complejo, marcado por dos esferas dominantes de influencia tecnológica y económica, lo que restringe el desarrollo de ecosistemas de IA autónomos y soberanos en ALC.

4.3 RIESGOS SISTÉMICOS

El Informe internacional de seguridad de la IA define a los riesgos sistémicos como aquellos que "más allá de los riesgos directamente asociados con las capacidades de modelos individuales, el despliegue generalizado de la IA de propósito general está relacionado con varios riesgos sistémicos más amplios". El Informe identifica seis áreas principales de preocupación:

1. Riesgos en el mercado laboral
2. Brecha global en I+D de IA
3. Concentración del mercado y puntos únicos de fallo
4. Riesgos medioambientales
5. Riesgos de privacidad
6. Infracciones de derechos de autor

Los riesgos sistémicos en el desarrollo y uso de la IA de propósito general en ALC están profundamente entrelazados con las desigualdades estructurales y las fragilidades institucionales de la región. Más que crear vulnerabilidades completamente nuevas, la IA tiende a amplificar desigualdades sociales, económicas y políticas preexistentes, como la desigualdad de género, la informalidad laboral y el acceso limitado a tecnologías digitales. Las personas participantes destacaron dos puntos clave.

Desplazamiento laboral:
Las personas participantes señalaron que la automatización impulsada por la IA de propósito general podría gravar el desplazamiento laboral de sectores como los call centers y los servicios de atención al cliente, donde ya se observan señales tempranas de desplazamiento laboral en algunos países de ALC. Esta preocupación contrasta con narrativas ampliamente difundidas que sostienen que las pérdidas de empleo serán limitadas o compensadas por la creación de nuevos puestos. De hecho, un informe del Banco Interamericano de Desarrollo (BID, 2023) estima que más de un tercio de los

empleos en América Latina podrían enfrentar un alto riesgo de automatización en la próxima década, en específico aquellos que dependen de tareas repetitivas, como el apoyo administrativo, operaciones en call centers, servicio al cliente. Para ilustrar esta dinámica, las personas participantes mencionaron el caso de Colombia. En este país, el sector de call centers creció un 746% en los últimos cuatro años y actualmente representa el 3.2% del PIB nacional, generando más de 750,000 empleos, siendo casi el 8% del empleo formal (Held, 2024). Sin embargo, pese a este crecimiento, muchos de estos puestos se encuentran simultáneamente en alto riesgo de automatización.

Brecha en las habilidades y la educación:

La rápida adopción de la IA de propósito general también está generando una brecha urgente en materia de habilidades y educación. Los sistemas educativos tradicionales no están logrando mantenerse al ritmo del cambio tecnológico, lo que provoca la obsolescencia de muchas de las competencias que actualmente se enseñan. Las personas participantes señalaron que este desafío es especialmente crítico en ALC, donde los sistemas educativos se encuentran entre los peores evaluados a nivel mundial. Las desigualdades estructurales implican que las zonas rurales se vean aún más afectadas en términos de calidad educativa.

Las personas participantes también destacaron la falta de infraestructura tecnológica en las escuelas, en particular en las escuelas públicas. Una realidad que limita el alcance de los programas de habilidades digitales y la capacidad de integrar de manera efectiva, así como guiar al estudiantado en el desarrollo de habilidades digitales fundamentales que el mercado laboral pide.

Falta de alineación normativa e institucional:

Las personas participantes advirtieron que, más allá de las dimensiones sociales y educativas, los riesgos sistémicos en ALC se ven agravados por la fragilidad institucional, una gobernanza fragmentada y la falta de continuidad de las políticas entre administraciones. Estas dinámicas obstaculizan el desarrollo de marcos regulatorios coherentes en materia de IA y de estrategias digitales a largo plazo. Las personas participantes también señalaron que, si bien muchos países de la región han adoptado agendas y regulaciones nacionales de IA, con frecuencia existe una brecha significativa entre el diseño de las políticas y su implementación. Los presupuestos limitados y la capacidad institucional insuficiente siguen siendo obstáculos comunes en gran parte de ALC. Las personas participantes coincidieron en que, para que la región pueda beneficiarse plenamente de la IA de propósito general, es indispensable crear los mecanismos necesarios y asignar los recursos necesarios para traducir las estrategias y regulaciones en acciones concretas.

05

FACTORES HABILITADORES DETRÁS DE LOS RIESGOS IDENTIFICADOS

05

FACTORES HABILITADORES DETRÁS DE LOS RIESGOS IDENTIFICADOS

Los riesgos identificados en toda la región no surgen de forma aislada. Más bien, se ven habilitados y amplificados por una serie de factores estructurales, institucionales y tecnológicos que definen el contexto de ALC. Además de los desafíos mencionados previamente, las personas participantes destacaron dos factores habilitadores transversales que permean en la región:

5.1 FALTA DE ALINEACIÓN INSTITUCIONAL Y COORDINACIÓN DE POLÍTICAS PÚBLICAS

Las personas participantes destacaron que muchas instituciones públicas en ALC continúan abordando la IA de manera aislada. Muchas de las agencias impulsan iniciativas digitales o de IA en paralelo, sin estándares, objetivos ni evaluaciones de riesgo compartidos. Esta fragmentación responde tanto a la ausencia de mecanismos de colaboración interinstitucional como lo son el aumento de la polarización política, lo que dificulta la construcción de acuerdos multisectoriales y la articulación de políticas digitales coherentes. Esta brecha se ve además agravada por los cambios de prioridades entre administraciones, que interrumpen la planificación a largo plazo, paralizan proyectos interadministrativos y debilitan la memoria institucional.

En muchos casos, las estrategias nacionales de IA o las agendas digitales existen en el papel, pero carecen de la coordinación operativa necesaria para traducirse a la práctica. Esto debilita la capacidad de abordar riesgos transversales como la desinformación, los ciberataques o la violencia digital, que requieren respuestas coordinadas entre los sectores de educación, justicia,

inseguridad, telecomunicaciones y los organismos reguladores. En conjunto, estos factores obstaculizan, en primer lugar, la capacidad de los países para diseñar e implementar políticas eficaces de mitigación de riesgos asociados a la IA y, en segundo lugar, a su capacidad para alinearse a nivel regional alrededor de principios y estrategias comunes.

5.2 FRAGMENTACIÓN DE LA GOBERNANZA REGIONAL

La realidad interna descrita anteriormente también se reproduce a nivel regional, donde los países de ALC suelen tomar decisiones de manera aislada, sin una alineación más amplia, lo que da lugar a una gobernanza fragmentada y a la duplicación de esfuerzos. Esta fragmentación, según las personas participantes, se origina por la polarización política, por enfoques regulatorios diferentes y en alineamientos internacionales divergentes, ya sea hacia Estados Unidos, Europa o China, que empujan a la región hacia distintas direcciones, así como lo mencionaron por la caída del multilaeralismo, lo cual crea una limitande para el desarrollo de la región. Si bien algunos países líderes participan en foros internacionales, gran parte de la región sigue estando subrepresentada. Esta falta de orientación colectiva limita la cooperación regional y debilita la influencia de ALC en los debates globales, dejando a los países de forma individual negociar desde una posición de desventaja estructural, con escasa capacidad para incidir en la definición de estándares globales. Dicho esto, algunos países, entre ellos Brasil, han logrado hacerse escuchar frente a los intentos de las grandes empresas tecnológicas por imponer sus propias reglas.

06

OPORTUNIDADES DE COLABORACIÓN EN LA REGIÓN

06

OPORTUNIDADES DE COLABORACIÓN EN LA REGIÓN

Las personas participantes estuvieron de acuerdo en que los desafíos y oportunidades presentados por la IA de propósito general requiere una respuesta regional coordinada y estratégica. Las siguientes recomendaciones fueron mencionadas por las personas participantes para impulsar el progreso hacia un desarrollo y uso más seguro de la IA de propósito general en la región:

CONSTRUIR UNA VISIÓN COMPARTIDA Y FORTALECER LA ACCIÓN COLECTIVA

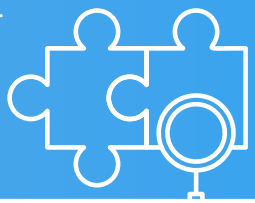
Con una población de 668 millones de personas, ALC representa un bloque importante de personas consumidoras e innovadoras. Al aprovechar su tamaño de mercado y con coordinación interna, la región puede exigir de manera colectiva estándares más altos de calidad, seguridad y equidad para las tecnologías que importa y desarrolla. Actuar de forma conjunta permitiría ampliar su voz en los debates globales y posicionar sus propios valores y perspectivas en la discusión sobre la gobernanza de la IA.

AVANZAR EN LA GOBERNANZA Y EN ESTÁNDARES REGIONALES

Los gobiernos de ALC podrían colaborar en el desarrollo de estándares locales, marcos regulatorios y códigos de conducta que promuevan la transparencia, la rendición de cuentas y la confianza en los sistemas de IA. Establecer reglas claras para el uso de la IA de propósito general en el sector público, el sistema de justicia y el sector privado contribuiría a reducir la proliferación de prácticas de "IA en la sombra" (shadow AI) (usos no regulados u opacos de la IA de propósito general) y a fortalecer la confianza pública. Cabe destacar que la cooperación regional no debe entenderse como un prerrequisito para la acción; los países pueden avanzar de manera individual al mismo tiempo que promueven la cohesión regional.

FORTALECER EL CONOCIMIENTO Y LA COLABORACIÓN EN INVESTIGACIÓN

StReforzar las plataformas de conocimiento compartido y las redes de investigación colaborativa de la región requiere no solo crear nuevas iniciativas, sino también consolidar y potenciar las ya existentes. Si bien estos mecanismos existen en ALC, a menudo carecen del mandato o de los recursos necesarios para abordar los principales desafíos sociales, económicos y ambientales que plantea la IA de propósito general. Por ejemplo, profundizar en el mapeo de riesgos regionales e identificar áreas prioritarias de acción permitiría desarrollar estrategias focalizadas y basadas en evidencia, alineadas con las realidades locales.



FORTALECER LA EDUCACIÓN Y LA ALFABETIZACIÓN DIGITAL

Para garantizar que la IA de propósito general funcione como una herramienta de empoderamiento, la región debe poner a la alfabetización en IA en el centro de su estrategia. Esto implica fomentar las competencias digitales, el pensamiento crítico y la conciencia ética entre el estudiantado, personal docente, medios de comunicación, funcionarios públicos y la ciudadanía en general. Los sistemas educativos deberían incorporar microcredenciales, trayectorias de formación flexibles y programas de aprendizaje continuo que respondan a las demandas cambiantes del mercado laboral.

AUMENTAR LA VISIBILIDAD A TRAVÉS DE LA INNOVACIÓN LOCAL

Finalmente, la región debe visibilizar los usos positivos y responsables de la IA de propósito general que ya existen dentro de sus fronteras. Documentar y difundir experiencias locales exitosas, como chatbots para la prestación de servicios públicos, herramientas predictivas en el sector salud o sistemas de apoyo en la educación, permitiría construir un repositorio de buenas prácticas, impulsar la innovación local, fortalecer la confianza pública, orientar a las personas personas formuladoras de políticas públicas y posicionar a ALC como una región capaz de liderar en innovación responsable y seguridad en IA.

07
CONCLUSIÓN

07 CONCLUSIÓN

La seguridad de la IA en América Latina y el Caribe no consiste únicamente en traer los debates de alto nivel sobre tecnologías emergentes a la región, sino también en enriquecer la conversación global con aportes basados en las propias experiencias, contextos y perspectivas de ALC. Este documento contribuye a ese objetivo al posicionar a la región como un actor activo dentro del ecosistema global de gobernanza de la IA, y no como un receptor pasivo de agendas externas. Esta iniciativa construye sobre las realidades regionales para comprender y abordar los riesgos asociados al desarrollo y uso de la IA de propósito general en ALC. A pesar de sus diferencias internas, la región comparte desafíos estructurales profundos, como lo son el desplazamiento laboral, las brechas de capacidades y técnicas, y la dependencia de tecnologías extranjeras, las cuales pueden servir como base para articular una voz colectiva más sólida. Reconocer y construir sobre estos puntos comunes permitiría fortalecer la capacidad de ALC para actuar de manera coordinada e influir en la forma en que la IA se desarrolla y utiliza tanto dentro como fuera de la región.

Lejos de ofrecer respuestas definitivas, el evento buscó sentar las bases para futuras iniciativas orientadas a profundizar el diálogo entre gobiernos, sociedad civil, academia y sector privado. Si existe un aprendizaje transversal de este primer ejercicio, es el siguiente: América Latina y el Caribe tienen tanto la responsabilidad como la oportunidad de contribuir al debate global sobre seguridad y gobernanza de la IA, no solo como países individuales, sino como una región con visión propia, prioridades compartidas y valores comunes. Alcanzar este objetivo requerirá coordinación, persistencia y voluntad política, pero también abre la posibilidad de desarrollar un modelo regional de gobernanza tecnológica, anclado en la inclusión, la equidad y la dignidad humana.

08

AGRADECIMIENTOS

AGRADECIMIENTOS

Queremos expresar nuestro más profundo agradecimiento a la Secretaría del Informe internacional sobre la seguridad de la IA de 2025 por su confianza en Eon Institute para coordinar este espacio. En participar, le extendemos nuestro agradecimiento a Freya Hempleman, antigua Responsable de las Relaciones Internacionales y Hannah Merchand, Jefa de Estrategia y Compromiso por el constante apoyo y colaboración durante el proceso. Asimismo, también estamos profundamente agradecidas con Daniel Privitera, Autor principal del Informe, y a Lucía Velasco, Integrante del equipo de redacción, por compartir generosamente sus conocimientos y opiniones con las personas participantes, y por fomentar un intercambio de ideas abierto que enriqueció enormemente el debate.

También nos gustaría extender nuestra gratitud a la Embajada Británica en la Ciudad de México por abrir sus puertas para celebrar este diálogo y por su apoyo en la organización del evento.

Finalmente, a todas las personas que fueron parte de este evento, agradecemos su tiempo, su compromiso y sus contribuciones valiosas, las cuales fueron fundamentales para la preparación de este documento. Los nombres de las personas participantes están enlistadas a continuación en orden alfabético.

Alvaro Marcelo Soto, Director del Centro Nacional de Inteligencia Artificial de Chile; Antonette Williams Barnett, Jefa de la Secretaría de Planificación Institucional y Sectorial en el Ministerio de Ciencia, Tecnología y Telecomunicaciones de Costa Rica; Carlos Carlderón, Director Fundador de la Academia de Datos, Inteligencia Artificial y Ciencias de la Computación de la Universidad de Ingeniería y Tecnología (UTEC) del Perú; Carolina Botero Cabrera, Miembro del Consejo Directivo de la Fundación Karisma y del Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (CETIC.br) y del Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE) de la Universidad de Palermo; Carolina Ines Aguerre Regusci, Profesora Adjunta en la Universidad Católica del Uruguay; Constanza Galarza Seeber, Asesora de la

AGRADECIMIENTOS

Dirección Nacional de Planeación y Asuntos Internacionales de la Secretaría de Innovación, Ciencia y Tecnología de Argentina; Daniel Álvarez, Director Nacional de la Agencia Nacional de Ciberseguridad de Chile; Diego Flores Jimenez, Titular de sector de industria electrónica y digital de la Secretaría de Economía de México; Enrique Chaparro, Secretario del Consejo Directivo de la Fundación Vía libre; Estefanía Capdeville, Directora de Proyectos Especiales y Cooperación Global de Agencia de Transformación Digital y Telecomunicaciones de México; Fabrizio Scrollini, Director Senior de Programas en Humanitarian OpenStreetMap Team; Giannina Patricia Gamio Franco de Taboada, Especialista en Sistemas Administrativos I en la Presidencia del Consejo de Ministros del Perú; Javier Barreiro, Fundador y Director de DaMa Uruguay y de IUGO por Domus Global; Jaime Alejandro Figueres Ulate, Director de Inteligencia Artificial en Zeta Intelligence; Jean Antonio García Periche, Presidente del Center for Public Intelligence (CPI); Kathia Anette De la Rosa, Directora y Fundadora de Action Intelligence-WSI; Liliana Fernández Gómez, Líder de Investigación en Spiral Centro de Tecnologías para el Desarrollo; Manuel Pliego Ramos, Director de Asuntos Gubernamentales en Microsoft México; Marushka Chocobar, Director of Generativa LATAM; Melvin Ángel Gago Rodrigo, Jefe de la Oficina General de Tecnología de la Información de la Presidencia del Consejo de Ministros del Perú; Óscar Alexander Ballen Cifuentes, Director de Asignación de Recursos del Ministerio de Tecnologías de la Información de Colombia; Rodrigo Roa, Director General de Data Observatory.