

Resumen

Información sobre el presente informe

- Esta es la publicación preliminar del primer “Informe científico internacional sobre la seguridad de la IA avanzada”. A este informe contribuyó un grupo heterogéneo de 75 expertos en inteligencia artificial (IA), entre los que se cuentan integrantes de un grupo consultivo internacional de expertos en cuyo nombramiento participaron 30 países, la Unión Europea (UE) y las Naciones Unidas.
- Los expertos independientes encargados de la redacción conjunta del presente informe, que estuvieron bajo la dirección de la persona a cargo de la coordinación del mismo, tuvieron plena discreción sobre su contenido.
- En una era de progresos sin precedentes en el desarrollo de la IA, esta primera publicación se centra en una rama que ha avanzado con especial rapidez en los últimos años: la IA de uso general, esto es, aquella capaz de realizar una gran diversidad de tareas. En un contexto de rápidos avances, la investigación sobre este tipo de IA sigue evolucionando y arrojando nuevos resultados y aún no es un campo científico consolidado.
- Los ciudadanos de todo el mundo solo podrán disfrutar con seguridad de las numerosas ventajas que puede ofrecer la IA de uso general si se gestionan debidamente sus riesgos. Este informe se centra en la detección de los mismos y en la evaluación de métodos técnicos para valorarlos y mitigarlos. El objetivo no es evaluar de forma exhaustiva todas las posibles repercusiones sociales de este tipo de IA, ni tampoco abordar los numerosos beneficios que puede generar.
- Por primera vez en la historia, este informe preliminar ha congregado a expertos designados por 30 países, la UE y las Naciones Unidas y a otros especialistas destacados a escala mundial para proporcionar unos fundamentos científicos comunes y basados en datos empíricos que contribuyan a los debates y las decisiones sobre la seguridad de la IA de uso general. Seguimos discrepando en varias cuestiones, menores y mayores, en torno a las capacidades de esta tecnología y a los riesgos conexos y su mitigación. Con todo, consideramos que este proyecto es esencial para mejorar nuestra comprensión colectiva de esta tecnología y sus posibles riesgos y para lograr el consenso y la mitigación eficaz de los riesgos que garantice que las personas puedan aprovechar los posibles beneficios de la IA de uso general en condiciones de seguridad. Es mucho lo que está en juego. Esperamos poder seguir avanzando en este sentido.

Aspectos destacados del resumen

- Si se gestiona adecuadamente, cabe utilizar este tipo de IA para promover el interés público, lo que puede redundar en mayores bienestar y prosperidad e impulsar los descubrimientos científicos. Sin embargo, un defecto de funcionamiento o un uso malintencionado también pueden resultar perjudiciales, por ejemplo, propiciando la adopción de decisiones sesgadas en situaciones transcendentales o allanando el camino a las estafas, las noticias falsas o las violaciones de la privacidad.

- Con el avance de las capacidades relativas a la IA de uso general, podrían surgir riesgos como son repercusiones a gran escala en el mercado laboral, ataques con agentes biológicos o de piratería informática asociados al uso de la IA y que la sociedad perdiese el control de esta tecnología, si bien los investigadores y expertos discuten la probabilidad de que estas hipótesis lleguen a materializarse. Las opiniones divergentes sobre estos riesgos suelen deberse a expectativas distintas sobre las medidas que adoptará la sociedad para limitarlos, la eficacia de esas medidas y la rapidez con que avanzarán las capacidades relacionadas con la IA de uso general.
- Existe una gran incertidumbre sobre el ritmo al que avanzarán las capacidades relacionadas con la IA de uso general en el futuro. Algunos expertos creen que lo más probable es que los avances se ralenticen, mientras que otros consideran que es posible o probable que vayan a una velocidad vertiginosa.
- Se dispone de varios métodos técnicos para evaluar y reducir los riesgos de la IA de uso general a los que pueden recurrir los responsables de desarrollar esta tecnología y que pueden exigir los encargados de regular su uso, pero todos tienen limitaciones. Por ejemplo, las técnicas actuales que se emplean para explicar por qué los modelos de este tipo de IA producen un determinado resultado son muy limitadas.
- El futuro de la IA de uso general es incierto y se plantean distintas posibilidades respecto a la senda que seguirá en lo sucesivo, las cuales arrojan resultados tanto muy positivos como muy negativos. Aun así, nada de lo que traiga consigo la IA resulta inevitable. Serán las decisiones de las sociedades y los Gobiernos las que determinen el futuro de esta tecnología. Este informe preliminar pretende facilitar un debate constructivo sobre tales decisiones.

Este informe sintetiza los conocimientos que ha acumulado hasta la fecha la comunidad científica sobre la IA de uso general —aquella capaz de realizar una gran diversidad de tareas— centrándose en la comprensión y gestión de sus riesgos.

La capacidad de los sistemas que utilizan IA ha evolucionado a gran velocidad. Estos avances han puesto de relieve las numerosas oportunidades que la IA brinda a las empresas, el mundo de la investigación, la administración pública y en la esfera privada, y asimismo han llevado a aumentar la concienciación sobre los daños que causa y los posibles riesgos futuros asociados a la IA avanzada.

El propósito del “Informe científico internacional sobre la seguridad de la IA avanzada” es propiciar la adopción de una concepción común a escala internacional sobre los riesgos vinculados a la IA y cómo pueden mitigarse. Esta primera publicación preliminar del citado informe se centra en un tipo de IA que ha evolucionado con especial rapidez: la IA de uso general, capaz de realizar una gran diversidad de tareas.

En un contexto de rápidos avances, la investigación sobre la IA de uso general sigue evolucionando y arrojando nuevos resultados y aún no es un campo científico consolidado. El informe presenta un panorama general de los conocimientos que ha acumulado hasta la fecha la comunidad científica sobre este tipo de IA y los riesgos que lleva aparejados. Para ello, se

señalan ámbitos en los que existe consenso científico y esferas en las que se observan divergencias o se plantean preguntas de investigación pendientes de respuesta.

Los ciudadanos de todo el mundo solo podrán disfrutar con seguridad de las numerosas ventajas que puede ofrecer la IA de uso general si se gestionan debidamente sus riesgos. Este informe se centra en la detección de los mismos y en la evaluación de métodos técnicos para valorarlos y mitigarlos, como por ejemplo el aprovechamiento de la propia IA para reducirlos. El objetivo no es evaluar de forma exhaustiva todas las posibles repercusiones sociales de esta tecnología, ni tampoco abordar los numerosos beneficios que puede traer consigo.

Según muchos indicadores, las capacidades relativas a la IA de uso general se han multiplicado con gran rapidez en los últimos años, y no hay consenso en cuanto a su evolución de cara al futuro, lo que da pie a que se plantee una gran diversidad de situaciones hipotéticas.

Muchos indicadores apuntan al rápido avance de las capacidades relacionadas con la IA de uso general. Hace cinco años, eran pocas las veces en que los principales modelos lingüísticos basados en IA llegaban a producir un párrafo de texto coherente. Hoy en día, algunos modelos pueden entablar conversaciones de varios turnos sobre una amplia gama de temas, programar a nivel básico o generar vídeos a partir de una descripción. Sin embargo, resulta difícil realizar una estimación fiable sobre las capacidades relativas a la IA de uso general y definir las con precisión.

La velocidad a la que evolucionará este tipo de IA depende tanto del ritmo de los avances tecnológicos como del marco normativo. Este informe se centra en los aspectos tecnológicos y no aborda la incidencia que las iniciativas normativas en este ámbito podrían tener en el ritmo al que se desarrolle y se vaya implantando esta tecnología.

Los responsables de desarrollar IA han logrado rápidos avances en la IA de uso general en los últimos años, sobre todo gracias a que los recursos destinados a poner a prueba nuevos modelos han seguido aumentando (tendencia conocida como “escalamiento”) y al perfeccionamiento de los algoritmos existentes. Por ejemplo, en términos anuales, los recursos computacionales (“computación”) utilizados para entrenar a los modelos de IA de última generación han llegado a cuadruplicarse, el tamaño de los conjuntos de datos para su entrenamiento se ha duplicado con creces, y la eficiencia de los algoritmos se han multiplicado por entre 1,5 y 3 veces (“rendimiento computacional”). Los investigadores debaten si el “escalamiento” ha permitido avanzar en desafíos fundamentales como el razonamiento causal.

El ritmo de los futuros avances en la IA de uso general tiene importantes repercusiones en la gestión de los riesgos que surjan, pero los expertos albergan opiniones divergentes en cuanto a lo que esta tecnología puede deparar aún en un futuro próximo. Las tesis de los expertos varían con respecto a la posibilidad de que las capacidades relacionadas con la IA de uso general avancen lento, rápido o sumamente rápido. Este desacuerdo gira en torno a una pregunta esencial: ¿Será suficiente el continuo “escalamiento” de los recursos y el

perfeccionamiento de las técnicas existentes para lograr avances rápidos y atajar problemas como la fiabilidad y la exactitud de la información, o se requieren nuevos avances en el campo de la investigación para lograr adelantos significativos en el ámbito de la IA de uso general?

Varias empresas líderes que se ocupan de desarrollar este tipo de IA están apostando por el “escalamiento” para seguir logrando mejoras en el rendimiento computacional. Si las tendencias recientes continúan, para finales de 2026 algunos modelos de IA de uso general se entrenarán utilizando entre 40 y 100 veces más computación que los modelos que en 2023 requerían el uso de recursos informáticos más intensivo, y los métodos de entrenamiento de esta tecnología que se utilicen serán entre 3 y 20 veces más eficientes. Sin embargo, existen posibles obstáculos para seguir aumentando tanto los datos como la computación, en especial la disponibilidad de datos, los chips de IA, los gastos de inversión y la capacidad energética a escala local. Las empresas que se encargan de desarrollar IA de uso general están trabajando para sortear dichos obstáculos.

El objetivo de varias líneas de investigación es comprender y evaluar la IA de uso general de manera más fiable, pero nuestra comprensión general de cómo funcionan los modelos y sistemas basados en este tipo de tecnología es limitada.

Los enfoques para gestionar los riesgos asociados a la IA de uso general a menudo se basan en el supuesto de que los responsables de desarrollar esta tecnología y los responsables de la formulación de políticas en este ámbito pueden evaluar las capacidades y posibles repercusiones de los modelos y sistemas de esta tecnología. No obstante, reconociendo que los métodos técnicos pueden contribuir a la evaluación, todos los que existen presentan limitaciones y no pueden ofrecer sólidas garantías frente a la mayoría de los daños que puede provocar la IA de uso general. En general, los conocimientos de la comunidad científica sobre el funcionamiento interno, la capacidad y la repercusión social de esta IA son muy limitados, y existe un amplio acuerdo entre expertos en que debería otorgarse prioridad a mejorar nuestra comprensión de esta tecnología. Algunos de los desafíos clave son los siguientes:

- Los responsables de desarrollar modelos de IA de uso general todavía saben poco sobre cómo funcionan. Esto se debe a que dichos modelos no están programados como solía hacerse, sino que están entrenados: los desarrolladores de IA disponen un entrenamiento que involucra una gran cantidad de datos, y el resultado de ese proceso es el modelo de IA de uso general. Estos modelos pueden constar de billones de componentes, llamados parámetros, y la mayoría de sus mecanismos de funcionamiento interno son inescrutables, incluso para los propios desarrolladores del modelo. Las técnicas de explicabilidad e interpretabilidad de modelos pueden contribuir a que los investigadores y desarrolladores entiendan el funcionamiento de estos modelos de IA, pero la investigación en este campo es incipiente.
- Este tipo de IA se evalúa principalmente probando el modelo o sistema con distintos datos de entrada. Estas verificaciones aleatorias son útiles para evaluar los puntos fuertes y débiles, en especial las vulnerabilidades y su capacidad de provocar posibles daños, pero no brindan garantías de seguridad cuantitativas. Las pruebas a menudo

pasan por alto peligros y sobrevaloran o subestiman funciones o aspectos porque estos sistemas de IA pueden comportarse de manera distinta en diferentes circunstancias y con distintos usuarios o si se modifican ulteriormente sus componentes.

- Los auditores externos pueden, en principio, auditar modelos o sistemas de IA de uso general diseñados por una empresa. Sin embargo, las empresas a menudo no les facilitan el nivel de acceso directo que necesitan a los modelos o la información sobre los datos y métodos utilizados para poder realizar una evaluación rigurosa. Algunos Gobiernos están empezando a crear capacidad para llevar a cabo evaluaciones y auditorías técnicas en esta esfera.
- Resulta difícil evaluar la repercusión social que tendrá un sistema de IA de uso general porque la investigación en materia de evaluación de riesgos no ha llegado a elaborar metodologías de evaluación rigurosas y exhaustivas. Además, este tipo de IA presenta numerosos casos de utilización, que a menudo no han sido definidos previamente y las restricciones a las que están sujetos son escasas, lo que complica aún más la evaluación de riesgos. Comprender las posibles repercusiones sociales de los modelos y sistemas basados en este tipo de IA exige un análisis matizado y multidisciplinario. Un reto técnico e institucional que está abordándose es la inclusión de una mayor diversidad de perspectivas en los procesos de desarrollo y evaluación de la IA de uso general.

La IA de uso general puede plantear graves riesgos para la seguridad y el bienestar tanto de los individuos como a nivel público.

Este informe clasifica los riesgos de esta tecnología en tres categorías: riesgos relativos al uso malintencionado, riesgos relacionados con defectos de funcionamiento y riesgos sistémicos. También analiza varios factores transversales que contribuyen a muchos riesgos.

Uso malintencionado. Como todas las tecnologías avanzadas, los sistemas de IA de uso general pueden utilizarse de forma malintencionada para provocar daños. Los posibles usos malintencionados van desde los que empiezan a ser relativamente conocidos, como las estafas a través de sistemas de este tipo de IA, hasta los que algunos expertos creen que podrían producirse en los próximos años, como el uso malintencionado de la capacidad científica que brinda la IA de uso general.

- Causar perjuicios a las personas a través de contenidos falsos generados por IA de uso general es un tipo de uso malintencionado del que se sabe bastante. Esta clase de IA puede utilizarse para aumentar la escala y la sofisticación de estafas y fraudes, por ejemplo mediante ataques por suplantación de identidad que se han optimizado con el uso de esta tecnología. También puede utilizarse para generar contenidos falsos comprometedores que muestren la imagen de personas sin su consentimiento, como la pornografía sintética no consentida.
- Otro motivo de preocupación es el uso malintencionado de este tipo de IA para desinformar y manipular a la opinión pública. La IA de uso general y otras tecnologías modernas facilitan la generación y difusión de desinformación, sobre todo para influir en los procesos políticos. Por lo general, pese a la utilidad de las contramedidas de

carácter técnico, como la marca de agua en los contenidos, los actores informáticos con un nivel de sofisticación intermedio en este ámbito consiguen eludirlas.

- También podría observarse un uso malintencionado de este tipo de IA en la ciberdelincuencia, al reforzar las destrezas cibernéticas de los usuarios y permitiendo que aquellos que utilizan esta tecnología de forma malintencionada perpetren ciberataques eficaces. Los sistemas de IA de uso general pueden utilizarse para aumentar la escala y automatizar parcialmente algunos tipos de operaciones cibernéticas, como los ataques de ingeniería social. Sin embargo, estos sistemas también podrían utilizarse en ciberdefensa. En general, todavía no se dispone de pruebas contundentes que sugieran que la IA de uso general pueda automatizar tareas sofisticadas de ciberseguridad.
- Algunos expertos también han expresado su preocupación por que esta IA pueda utilizarse para apoyar el diseño y uso malintencionado de armas, como las biológicas. No hay pruebas sólidas de que los actuales sistemas de IA de uso general planteen este riesgo. Por ejemplo, si bien las capacidades de dichos sistemas en el ámbito de la biología son mayores, los estudios de que se dispone, que son limitados, no aportan pruebas claras de que estos sistemas permitan a los usuarios maliciosos obtener patógenos biológicos con mayor facilidad que utilizando Internet. Sin embargo, las futuras amenazas a gran escala apenas se han evaluado y son difíciles de descartar.

Riesgos relacionados con defectos de funcionamiento. Incluso cuando los usuarios no tienen intención de causar daños, pueden surgir riesgos graves debido a defectos en el funcionamiento de este tipo de IA. Estos fallos pueden deberse a varias causas y tener distintas consecuencias:

- Puede que los usuarios no entiendan bien las funciones de los productos basados en modelos y sistemas de IA de uso general, por ejemplo debido a un problema de comunicación o a publicidad engañosa, pudiendo provocar daños si los utilizan de forma inadecuada o con fines inadecuados.
- Los sesgos en los sistemas de IA en general constituyen un conocido problema que sigue sin resolverse también en el caso de la IA de uso general. Los resultados de este tipo de IA pueden ser parciales con respecto a características que cuentan con protección como la raza, el género, la cultura, la edad y la discapacidad. Esta circunstancia puede crear riesgos, sobre todo en ámbitos delicados como la sanidad, la contratación de personal y los préstamos financieros. Además, muchos modelos de IA cuyo uso está muy generalizado se entrenan principalmente con datos que eminentemente proceden de las culturas occidentales, lo que puede aumentar los daños que pueden sufrir las personas sobre las que los datos usados para el entrenamiento de esos modelos no dan suficiente cuenta.
- Las hipótesis de “pérdida de control” son posibles situaciones futuras en que la sociedad ya no puede poner un freno real a los sistemas de IA de uso general, incluso cuando resulta evidente que están provocando daños. Existe amplio consenso en que la concepción de los actuales sistemas de IA de uso general no plantea este riesgo. Algunos expertos creen que las iniciativas en curso para desarrollar IA *autónoma* de uso

general (sistemas que pueden tomar decisiones, planificar y perseguir objetivos) podrían desembocar en una pérdida de control si prosperasen. Los expertos discrepan sobre la verosimilitud de las situaciones de pérdida de control, cuándo podrían producirse y lo difícil que sería mitigarlas.

Riesgos sistémicos. El desarrollo y la adopción generalizados de sistemas de IA de uso general plantea varios riesgos sistémicos, que van desde posibles repercusiones en el mercado laboral hasta riesgos para la privacidad y efectos ambientales:

- La IA de uso general, especialmente si sigue avanzando a gran velocidad, puede emplearse para automatizar muchas tareas, lo que podría tener efectos importantes en el mercado laboral. Esta automatización podría significar que muchas personas perdiesen sus actuales empleos. Sin embargo, muchos economistas esperan que la posible pérdida de empleos pueda compensarse, posiblemente en su totalidad, con la creación de nuevos empleos y con una mayor demanda en sectores no automatizados.
- La investigación y el desarrollo de IA de uso general se concentra actualmente en unos pocos países occidentales y China. Esta “brecha de la IA” obedece a diversas causas, pero en parte se debe a los diferentes niveles de acceso a la computación necesaria para desarrollar un sistema de IA de uso general. El menor acceso que tienen los países de bajos ingresos y las instituciones académicas a la computación con respecto a los países de altos ingresos y las empresas tecnológicas los pone en situación de desventaja.
- La concentración del mercado relativo al desarrollo de IA de uso general que resulta de ello hace que las sociedades sean más vulnerables a varios riesgos sistémicos. Por ejemplo, el uso generalizado de un reducido número de sistemas de IA de este tipo en sectores fundamentales como las finanzas o la salud podría causar fallas e interrupciones simultáneas a gran escala en estos sectores interdependientes, como consecuencia de, por ejemplo, errores o vulnerabilidades.
- El creciente uso de la computación en el desarrollo y la implantación de modelos de IA de uso general ha provocado un rápido aumento del consumo de energía asociado con esta tecnología. Esta tendencia no parece que vaya a moderarse, lo que podría llevar aparejado un incremento de las emisiones de CO₂ y del consumo de agua.
- Los modelos o sistemas de IA de uso general pueden plantear riesgos para la privacidad. Por ejemplo, las investigaciones han demostrado que, si se utilizan datos de entrada contradictorios, los usuarios pueden acceder a información de carácter personal al extraer datos de entrenamiento de un determinado modelo de IA. Para futuros sistemas que se han entrenado con datos personales sensibles, como información sobre la salud o de índole financiera, lo anterior puede conducir a intromisiones en la privacidad de especial gravedad.
- Las posibles violaciones de los derechos de autor en el desarrollo de IA de uso general representan un reto para la legislación tradicional en materia de propiedad intelectual, así como para los sistemas de obtención del consentimiento, compensación y control de los datos. Un régimen de derechos de autor poco claro desincentiva a los desarrolladores de este tipo de IA a declarar qué datos utilizan y no precisa qué

protección se brinda a los creadores cuyo trabajo se utiliza sin su consentimiento para entrenar modelos de IA de uso general.

Factores de riesgo transversales. Entre los riesgos asociados a los sistemas de IA de uso general se encuentran varios factores de riesgo transversales, que consisten en características de los propios sistemas de IA que aumentan la probabilidad o la gravedad no de uno, sino de varios riesgos:

- Entre los factores de riesgo transversales de carácter técnico cabe citar la dificultad de garantizar que los sistemas de IA de uso general se comporten de forma fiable según lo previsto, nuestra falta de comprensión sobre su funcionamiento interno y los “agentes inteligentes” que se están diseñando, que funcionan de forma autónoma y se someten a una menor supervisión.
- Entre los factores de riesgo transversales de carácter social destaca el posible desajuste entre el ritmo al que avanza la tecnología y la llegada de las medidas reglamentarias en este ámbito, así como los incentivos a la competencia para que los encargados del desarrollo de sistemas de IA saquen productos al mercado con rapidez, lo que posiblemente iría en detrimento de la gestión exhaustiva de los riesgos.

Hay varios enfoques técnicos que pueden contribuir a mitigar los riesgos, pero ningún método conocido en la actualidad ofrece certeza ni garantías sólidas frente a los daños asociados a los sistemas de IA de uso general.

Aunque en este informe no se abordan las intervenciones normativas para mitigar los riesgos que plantea la IA de uso general, sí se analizan los métodos técnicos de mitigación de riesgos en los que los investigadores están avanzando. A pesar de estos avances, los métodos actuales no resultan fiables a la hora de impedir tan siquiera los resultados manifiestamente dañinos de estos sistemas de IA en contextos reales. Se utilizan varios enfoques técnicos para evaluar y mitigar los riesgos:

- Se está avanzando en el entrenamiento de modelos de IA de uso general para que funcionen de forma más segura. Los desarrolladores también entrenan a los modelos para que sean más resistentes a los datos de entrada que se crean para hacerlos fallar (“entrenamiento contradictorio”). A pesar de ello, quienes están detrás de esos datos contradictorios suelen encontrar alternativas que reducen la eficacia de las salvaguardias establecidas, sin que para ello tengan que esforzarse mucho. Limitar las capacidades de un sistema de IA de este tipo a casos de uso específicos puede ayudar a reducir el riesgo de fallos imprevistos o los usos malintencionados.
- Existen varias técnicas para detectar riesgos, comprobar el funcionamiento del sistema y evaluar el rendimiento una vez que se ha implantado un sistema de IA de uso general. Estas prácticas a menudo se denominan “seguimiento”.
- La mitigación de los sesgos en los sistemas de IA de uso general se puede abordar durante todo el ciclo de los sistemas, desde el diseño y el entrenamiento hasta la implantación y la utilización. Sin embargo, evitar por completo los sesgos es todo un

reto porque requiere una recopilación sistemática de datos para el entrenamiento de los sistemas, una evaluación continua y una detección eficaz de los sesgos. También puede ser necesario llegar a una solución a medio camino entre la equidad y otros objetivos como la precisión y la privacidad, y decidir qué se considera conocimiento útil y qué es un sesgo indeseable que no debería reflejarse en los resultados.

- La protección de la privacidad es un campo de investigación y desarrollo prolífero. Con solo reducir al mínimo el uso de datos personales sensibles en el entrenamiento de los sistemas se pueden mitigar considerablemente los riesgos ligados a la privacidad. Sin embargo, cuando se utilizan datos sensibles, ya sea de forma intencionada o no, no resulta fácil adaptar los recursos técnicos que existen y se utilizan para reducir los riesgos relacionados con la privacidad a grandes modelos de IA de uso general y puede que no brinden un control efectivo a los usuarios.

Conclusión: La IA de uso general puede avanzar por muy distintos caminos y mucho dependerá de cómo actúen las sociedades y los Gobiernos.

El futuro de la IA de uso general es incierto y se plantean distintas posibilidades respecto a la senda que seguirá en lo sucesivo, las cuales arrojan resultados tanto muy positivos como muy negativos. Aun así, nada de lo que traiga consigo la IA resulta inevitable. Cómo se desarrollan los sistemas de IA de uso general y quién los diseña, qué problemas debe resolver esta tecnología, si las sociedades podrán aprovechar todas las posibilidades económicas que ofrece la IA de uso general, quién se beneficia de ella, a qué tipos de riesgos nos exponemos y cuánto invertimos en investigación para mitigarlos... estas y muchas otras preguntas dependen de las decisiones que las sociedades y los Gobiernos tomen en la actualidad y en el futuro para determinar cómo seguirá evolucionando esta tecnología.

En aras de facilitar un debate constructivo sobre estas decisiones, este informe proporciona una visión general del punto en que se encuentran la investigación científica y el debate sobre la gestión de los riesgos vinculados a estos sistemas de IA en la actualidad. Es mucho lo que está en juego. Esperamos poder seguir avanzando en este campo.