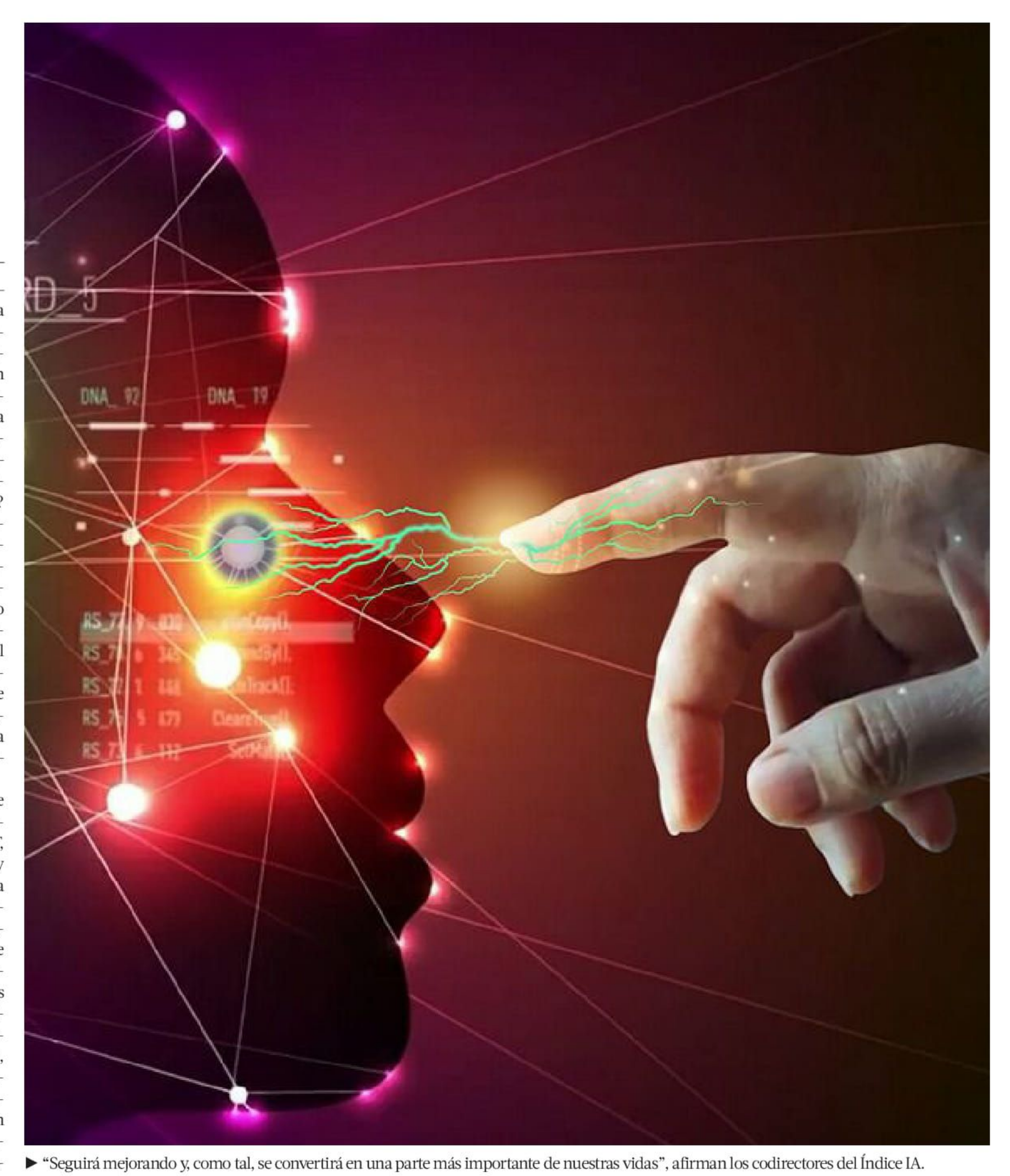


► "Seguirá mejorando y, como tal, se convertirá en una parte más importante de nuestras vidas", afirman los codirectores del Índice IA.

Stanford advierte que inteligencia artificial es "propensa a las alucinaciones y a ser engañada"

La Universidad de Stanford publicó recientemente el último Índice IA, el cual destaca los principales avances y desafíos de la inteligencia artificial. La poca incursión de la academia, el daño medioambiental y los sesgos raciales y de género resaltan como las mayores preocupaciones recogidas en este sondeo.

Francisco Corvalán

¿Quién le teme más? ¿Quién invierte más? La tecnología y las aplicaciones que utilizan algoritmos de inteligencia artificial (IA) ha tenido un explosivo avance. Ya está inmerso en gran parte de la sociedad y sus distintas herramientas han generado gran expectativa entre las personas, al mismo tiempo que incertidumbre. ¿Cómo ha cambiado el desarrollo y nuestra percepción sobre esta sorprendente incursión tecnológica en el último año?

La Universidad de Stanford publicó recientemente el Índice IA. Esta es una iniciativa independiente del Instituto Stanford para la Inteligencia Artificial centrada en el ser humano (HAI), dirigida por un grupo interdisciplinario de expertos de todo el mundo académico y la industria. El informe anual rastrea, coteja, destila y visualiza datos relacionados con la inteligencia artificial, lo que permite a los responsables de la toma de decisiones estudiar medidas significativas para promover la IA de manera responsable y ética.

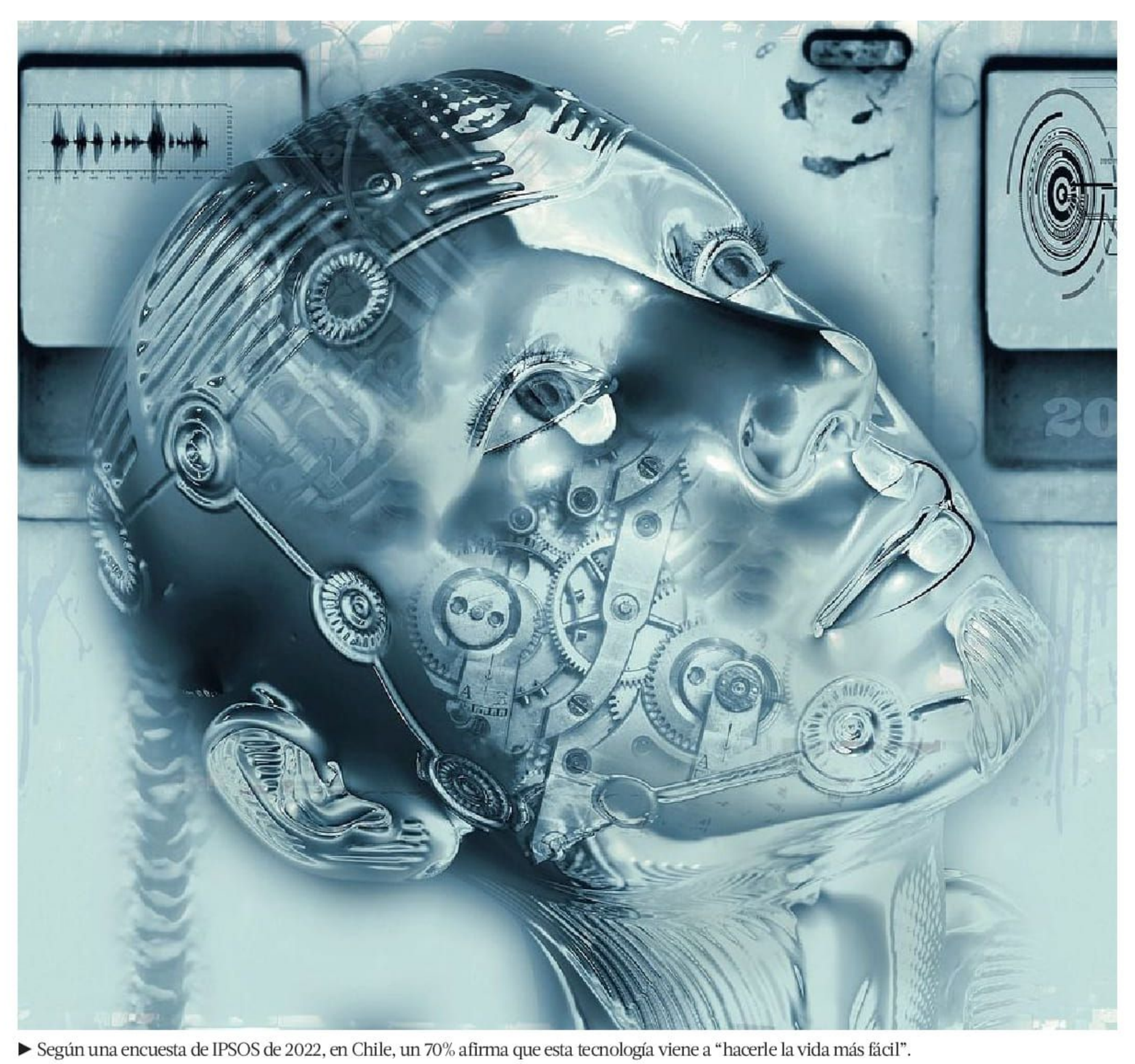
A lo largo de 2022 y principios de 2023, se han lanzado nuevos modelos de IA a gran escala cada mes. Estos modelos, como ChatGPT, Stable Diffusion, Whisper, Midjourney y DALL-E 2, son capaces de realizar una gama cada vez más amplia de tareas, desde la manipulación y el análisis de texto hasta la generación de imágenes y un reconocimiento de voz sin precedentes. Estos sistemas demuestran capacidades en la respuesta a preguntas y la generación de texto, imagen y código inimaginables hace una década, y superan el estado del arte en muchos puntos de referencia, antiguos y nuevos. "Sin embargo, son propensos a las alucinaciones, habitualmente sesgados y pueden ser engañados para que sirvan a fines nefastos, lo que destaca los complicados desafíos éticos asociados con su despliegue", manifiestan los codirectores de este Índice IA, Jack Clark y Ray Perrault.

Según detallan en un comunicado, del estudio pudieron cotejar que 2022 fue el primer año en una década en que disminuyó la inversión privada en IA. Sin embargo, esta tecnología sigue siendo un tema de gran interés para los formuladores de políticas, los líderes de la industria, los investigadores y el público. "Seguirá mejorando y, como tal, se convertirá en una parte más importante de nuestras vidas. Dada la mayor presencia de esta tecnología y su potencial de disrupción masiva, todos deberíamos comenzar a pensar de manera más crítica sobre cómo queremos exactamente que se desarrolle e implemente la IA", expresaron.

A esto también se suma la invitación a cuestionarse más sobre quién está actualmente implementando algoritmos y tecnologías de IA. Ante esto, Clark y Perrault explican que "como muestra nuestro análisis, la IA se define cada vez más por las acciones de un pequeño conjunto de actores del sector privado, en lugar de una gama más amplia de actores sociales".

¿Cuáles fueron las principales conclusiones

SIGUE ►►



► Según una encuesta de IPSOS de 2022, en Chile, un 70% afirma que esta tecnología viene a "hacerle la vida más fácil".

SIGUE ►►

de este estudio? La primera, y quizás la más destacada, apunta a que la industria corre por delante de la academia. Hasta 2014, los modelos de aprendizaje automático más importantes fueron publicados por universidades e instituciones académicas sin fines de lucro. Desde entonces, la industria se ha hecho cargo. En 2022, hubo 32 modelos importantes de aprendizaje automático producidos por la industria en comparación con sólo tres producidos por la academia.

Ante esto, el director de Vinculación con el Medio del Centro Nacional de Inteligencia Artificial (Cenia), Rodrigo Durán, destaca que "para ningún Estado ha sido posible hacer financieramente algo como el ChatGPT, algo que saque la pelota del estadio. La capacidad de cómputo que tiene OpenAI, por ejemplo es sideralmente mayor a la que podría generar la academia". Asimismo, el especialista agrega que existe una preocupación de la comunicación de los investigadores en el rol que tienen los privados de todo esto. "Finalmente los privados son los que están empujando el carro, y hoy por hoy no tienen mucho control", dice.

Estamos en un momento previo a la Revolución Industrial. Chat GPT podría ser como la máquina a vapor. Las implicancias que puede tener son inimaginables, y en general muy positivas. Sin embargo, Durán antepone que estas repercusiones "tenemos que verlas con cuidado, porque hemos aprendido de las revoluciones industriales anteriores, pero en general tienen cambios que son tremendamente positivos".

Otra conclusión que realizó este informe apunta a lo paradójico del desarrollo de esta tecnología y su relación con el ecosistema. La IA está ayudando y dañando el medioambiente al mismo tiempo. Una nueva investigación sugiere que los sistemas de IA pueden tener graves impactos ambientales. Según un estudio realizado a fines de 2022, la carrera de entrenamiento de BLOOM emitió 25 veces más carbono que un viaje de ida desde Nueva York a San Francisco, equivalentes a más de 4.000 kilómetros. Aun así, los nuevos modelos de aprendizaje por refuerzo como BCOOLER muestran que los sistemas de IA se pueden usar para optimizar el uso de energía.

Confianza en la IA y sesgo de género

El Índice IA también sacó a la luz un fenómeno singular: los países donde principalmente se desarrollan estas tecnologías no tienen un gran nivel de confianza, a diferencia de otras latitudes. En una encuesta de IPSOS de 2022, mostró que China tiene un mayor grado de confianza ante esta tecnología. Un 78 % de los encuestados estuvo de acuerdo con la afirmación de que los productos y servicios que utilizan IA tienen más ventajas que inconvenientes. Después de ellos, los de Arabia Saudita (76 %) e India (71 %) se sintieron más positivos acerca de los productos de IA. En el caso de Chile, un 67% de los encuestados manifestaron que la IA "va a solucionar sus vidas de aquí a tres años más", y un 70% afirma que esta tecnología viene a "hacerle la vida más fácil".

Por otro lado, solo el 35% de los estadounidenses incluidos en la muestra estuvo de acuerdo en que los productos y servicios que utilizan IA tenían más beneficios que inconvenientes. "Desde mi punto de vista eso se explica porque hay mayor conciencia a las consecuencias negativas de la IA en los países donde se desarrolla y aplica esta tecnología", comenta Durán.

Según comenta, han existido casos en Estados Unidos y Países Bajos, en los cuales se ha usado la IA para evaluar personas en distintas cosas, tales como para un seguro de vida, una calificación crediticia o para el acceso a beneficios sociales. "En el fondo, esos países tienen mayor desconfianza porque han sufrido la curva de aprendizaje en el para qué se deben usar los algoritmos y para qué no. No se deberían usar algoritmos para entregar derechos sociales", comenta.

Por otro lado, un 55% de los hombres cree que los productos y servicios que utilizan inteligencia artificial poseen más beneficios que desventajas, mientras que esto mismo solo lo declara el 49% de las mujeres. En prácticamente todos los ámbitos evaluados por el Índice, los hombres declaran sentirse más seguros y optimistas frente a estas tecnologías en comparación a las mujeres. ¿Por qué podría ocurrir esto?

Los modelos de lenguaje son más capaces que nunca de comprender y producir texto. Sin embargo, éstas pueden reflejar también cada vez más sesgos. Según declara la subsecretaria de Ciencia, Carolina Gainza, abordar estos sesgos de género en los sistemas de IA "requiere analizar cómo estos se construyen, ya que los sesgos pueden surgir en todas las etapas del proceso: desde el factor humano en el diseño de los algoritmos, a las bases de datos con lo que estos sistemas son entrenados y los posteriores mecanismos de retroalimentación".

La autoridad de gobierno remarca que existen principios que deben ser parte fundamental del diseño de estos sistemas. "La equidad, la ética, la transparencia, la rendición de cuentas, la perspectiva de derechos y la explicabilidad deben formar parte del proceso de diseño e implementación de las inteligencias artificiales", explica Gainza.

Desde 2021, Chile tiene una política de Inteligencia Artificial que ahora retomó el Ministerio de Ciencia para profundizarla y abordar las principales preocupaciones y desafíos que aparecen con el desarrollo de esta tecnología. "Nuestro objetivo actual es que la investigación financiada en esta materia informe las políticas públicas sobre IA, habilite su desarrollo y permita realizar acciones para empoderar a las personas en la comprensión, uso y desarrollo de herramientas de inteligencia artificial. Con esto nos sumamos a las voces desde el sur global que tienen que necesariamente participar de las discusiones respecto al desarrollo presente y futuro de la IA", declara la subsecretaria de Ciencia.

En adición, la entidad de Gobierno realizó una alianza con CAF, el Banco de Desarrollo de América Latina y la UNESCO para impulsar el uso inclusivo y ético de la inteligencia artificial por parte de los gobiernos nacionales, subnacionales y locales de América Latina y el Caribe. Se ha acordado realizar la primera cumbre regional sobre Ética de la IA en Santiago de Chile durante octubre, donde se conformará el Consejo Regional de Ética de Inteligencia Artificial, principal espacio político y técnico para la gobernanza de la IA en la región. ●