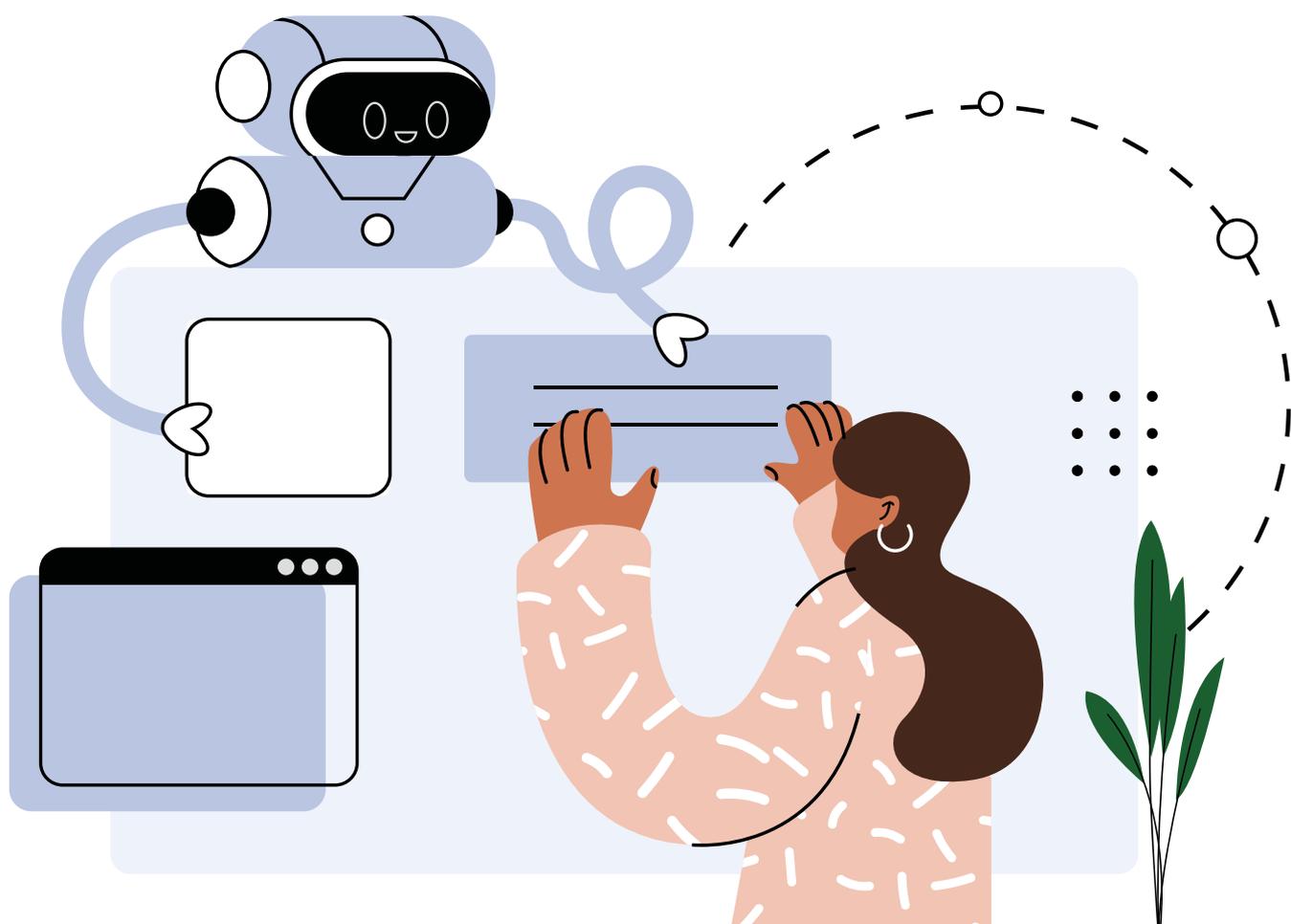


Guía para el uso de IA generativa en educación e investigación



La UNESCO: líder mundial en educación

La educación es la máxima prioridad de la UNESCO porque es un derecho humano esencial y la base para consolidar la paz y el desarrollo sostenible. La UNESCO es la agencia de las Naciones Unidas especializada en educación. Proporciona un liderazgo a nivel mundial y regional para reforzar el desarrollo, la resiliencia y la capacidad de los sistemas educativos nacionales al servicio de todos los estudiantes. La UNESCO lidera los esfuerzos para responder a los desafíos mundiales actuales mediante un aprendizaje transformador, con un enfoque especial en la igualdad de género y África a través de todas sus acciones.



La Agenda Mundial de Educación 2030

En calidad de organización de las Naciones Unidas especializada en educación, la UNESCO ha recibido el encargo de dirigir y coordinar la Agenda de Educación 2030. Este programa forma parte de un movimiento mundial encaminado a erradicar la pobreza mediante la consecución, de aquí a 2030, de 17 Objetivos de Desarrollo Sostenible. La educación, fundamental para alcanzar todos estos objetivos, cuenta con su propio objetivo específico, el ODS 4, que se ha propuesto *“garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos”*. El Marco de Acción de Educación 2030 ofrece orientación para la aplicación de este ambicioso objetivo y sus compromisos.



Publicado en 2024 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
7, place de Fontenoy, 75352 París 07 SP, Francia

© UNESCO 2024

ISBN: 978-92-3-300221-0



Esta publicación está disponible en acceso abierto bajo la licencia Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Al utilizar el contenido de la presente publicación, los usuarios aceptan las condiciones de utilización del Repositorio UNESCO de acceso abierto (<https://www.unesco.org/es/open-access/cc-sa>).

Las imágenes acompañadas de un asterisco (*) no están cubiertas por la licencia [CC-BY-SA](https://creativecommons.org/licenses/by-sa/3.0/igo/) y no pueden usarse ni reproducirse sin previa autorización por escrito de sus propietarios.

Título original: *Guidance for generative AI in education and research*

Publicado en 2023 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Los términos empleados en esta publicación y la presentación de los datos que en ella aparecen no implican toma alguna de posición de parte de la UNESCO en cuanto al estatuto jurídico de los países, territorios, ciudades o regiones ni respecto de sus autoridades, fronteras o límites.

Las ideas y opiniones expresadas en esta obra son las de los autores y no reflejan necesariamente el punto de vista de la UNESCO ni comprometen a la Organización.

La traducción de la Guía al español incluyó una revisión técnica por parte del Centro Regional de Estudios para el Desarrollo de la Sociedad de la Información bajo los auspicios de la UNESCO (Cetic.br), del Núcleo de Información y Coordinación del Ponto BR (NIC.br). La traducción fue realizada por Prioridade Consultoria Ltda. La revisión técnica de la traducción al español estuvo a cargo de Ana Laura Martínez Tessore (Cetic.br).

Créditos de cubierta: Olexandra Simkina/Shutterstock.com*

Diseñado e impreso por la UNESCO

Impreso en Francia

R E S U M E N

Hacia un enfoque centrado en el ser humano en el uso de la IA generativa

Las herramientas de IA generativa de acceso público (IAGen) están surgiendo rápidamente, y el lanzamiento de versiones iterativas supera la velocidad de adaptación de los marcos regulatorios nacionales. La falta de regulaciones nacionales sobre IAGen en la mayoría de los países deja desprotegida la privacidad de los datos de los usuarios y a las instituciones educativas, en gran medida, sin preparación para validar las herramientas.

La primera orientación global de la UNESCO sobre IAGen en educación apunta a apoyar a los países en la implementación de acciones inmediatas, la planificación de políticas de largo plazo y el desarrollo de capacidades humanas para garantizar una visión de estas nuevas tecnologías centrada en el ser humano.

Esta guía presenta una evaluación de los posibles riesgos que la IAGen podría plantear para los valores humanísticos fundamentales que promueven la intervención humana, la inclusión, la equidad, la igualdad de género, la diversidad lingüística y cultural, así como las opiniones y expresiones plurales.

Propone pasos clave para que las agencias gubernamentales regulen el uso de las herramientas de IAGen, incluyendo la obligación de proteger la privacidad de los datos y considerar un límite de edad para su uso. Establece requisitos para los proveedores de IAGen para permitir su uso ético y efectivo en la educación.

La guía destaca la necesidad de que las instituciones educativas validen la idoneidad ética y pedagógica de los sistemas de IAGen para la educación. Hace un llamado a la comunidad internacional para reflexionar sobre sus implicancias a largo plazo en el conocimiento, la enseñanza, el aprendizaje y la evaluación.

La publicación ofrece recomendaciones concretas para los formuladores de políticas y las instituciones educativas sobre cómo diseñar el uso de herramientas de IAGen, con el fin de proteger la intervención humana y beneficiar genuinamente a estudiantes, aprendices e investigadores.

Mientras que
ChatGPT alcanzó los
100 millones
de usuarios activos
mensuales en enero de 2023,
solo **un país** publicó
normativas sobre **IAGen**
en julio 2023



unesco

“Puesto que las guerras nacen en la mente de los hombres y de las mujeres, es en la mente de los hombres y de las mujeres donde deben erigirse los baluartes de la paz”



unesco

Guía para el uso de IA generativa en educación e investigación

Prefacio



© UNESCO

La inteligencia artificial generativa (IAGen) irrumpió en la escena pública a finales de 2022 con el lanzamiento de ChatGPT, que se convirtió en la aplicación de más rápido crecimiento en la historia. Con el poder de emular las capacidades humanas para generar resultados en formatos como texto, imágenes, videos, música y códigos de software, estas aplicaciones de IAGen han causado sensación. Millones de personas utilizan ahora la IAGen en su vida diaria, y el potencial de adaptar los modelos a aplicaciones de IA de dominios específicos parece ilimitado.

Este amplio rango de capacidades de procesamiento de información y de producción de conocimiento tiene, potencialmente, grandes implicancias para la educación, dado que replica el pensamiento de orden superior que constituye la base del aprendizaje humano. Dado que las herramientas de IAGen son cada vez más capaces de automatizar algunos niveles básicos de escritura y creación artística, están obligando a los formuladores de políticas educativas y a las instituciones a reconsiderar el por qué, el qué y el cómo aprendemos. Estas son consideraciones críticas para la educación en esta nueva fase de la era digital.

Esta publicación apunta a respaldar la planificación de regulaciones, políticas y el desarrollo de capacidades humanas apropiadas, para garantizar que la IAGen se convierta en una herramienta que realmente beneficie y empodere a docentes, estudiantes e investigadores.

Propone pasos clave para que los organismos gubernamentales regulen el uso de la IA generativa. Asimismo, presenta marcos y ejemplos concretos para la formulación de políticas y el diseño instruccional que permitan un uso ético y efectivo de esta tecnología en la educación. Finalmente, hace un llamado a la comunidad internacional para que considere las profundas implicancias, a largo plazo, de la IA generativa en la forma en que entendemos el conocimiento y definimos el contenido, los métodos y los resultados de aprendizaje, así como en la manera en que lo evaluamos y validamos.

Con base en la *Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO* de 2021, esta guía se fundamenta en un enfoque humanístico de la educación que promueve la intervención humana, la inclusión, la equidad, la igualdad de género, la diversidad cultural y lingüística, así como las opiniones y expresiones plurales. Además, responde al llamado del informe de 2021 de la Comisión Internacional sobre los Futuros de la Educación, *Reimaginar juntos nuestros futuros: un nuevo contrato social para la educación*, para redefinir nuestra relación con la tecnología, como parte integral de nuestros esfuerzos con el fin de renovar el contrato social de la educación.

La IA no debe usurpar los ámbitos propios de la inteligencia humana. Más bien nos invita a reconsiderar nuestras concepciones establecidas del conocimiento y el aprendizaje humano. Es mi esperanza que esta guía nos ayude a redefinir nuevos horizontes para la educación y a fundamentar nuestro pensamiento colectivo y acciones colaborativas que puedan conducir a un futuro de aprendizaje digital centrado en el ser humano para todos.

Stefania Giannini,
Subdirectora General de Educación de la UNESCO

Agradecimientos

Bajo la dirección de Stefania Giannini, Subdirectora de Educación, y la orientación de Sobhi Tawil, Director de la División del Futuro del Aprendizaje y la Innovación de la UNESCO, la redacción de la publicación estuvo a cargo de Fengchun Miao, Jefe de la Unidad de Tecnología e IA en la Educación.

Un especial agradecimiento a Wayne Holmes, Profesor Asociado del University College London, que ha colaborado en la redacción de algunas secciones de la publicación.

Esta publicación es fruto del esfuerzo colectivo de líderes educativos y de expertos en el campo de la IA y la educación.

Se ha beneficiado de las opiniones y contribuciones de muchos expertos, entre ellos: Mutlu Cukurova, profesora del University College London; Colin de la Higuera, Catedrático UNESCO de Tecnologías para la Formación de Profesores con Recursos Educativos Abiertos de la Universidad de Nantes; Shařika Isaacs, investigadora Asociada de la Universidad de Johannesburgo; Natalie Lao, Directora Ejecutiva de la App Inventor Foundation; Qin Ni, Profesor Asociado de la Universidad Normal de Shanghai; Catalina Nicolin, experta en TIC en educación del Centro Europeo de Educación Digital de Rumanía; John Shaw-Taylor, Catedrático UNESCO de IA y Profesor de Estadística Computacional y Aprendizaje Automático del University College London; Kelly Shirohira, Directora Ejecutiva de Jet Education Services; Ki-Sang Song, Profesor de la Universidad Nacional de Educación de Corea; e Ilkka Tuomi, Científico Jefe de Meaning Processing Ltd de Finlandia.

Muchos colegas de toda la UNESCO también contribuyeron de diversas maneras, entre ellos: Dafna Feinholz, Jefa de la Sección de Bioética y Ética de Ciencia y Tecnología; Francesc Pedró, Director del Instituto Internacional para la Educación Superior en América Latina y el Caribe; Prateek Sibal, Especialista de Programa, Sección de Políticas Digitales y Transformación Digital; Saurabh Roy, Oficial Superior de Proyectos en la Sección de Desarrollo Docente, División de Políticas y Sistemas de Aprendizaje a lo Largo de la Vida; Benjamin Vergel De Dios, Especialista de Programa en TIC en Educación, Sección de Innovación Educativa y Desarrollo de Habilidades de la Oficina de Bangkok; los colegas de la Entidad de Diversidad de Expresiones Culturales del Sector de Cultura; y Mark West, Especialista de Programa, División de Futuro del Aprendizaje y la Innovación.

Merecen también nuestro agradecimiento Glen Hertelendy, Luisa Ferrara y Xianglei Zheng, de la Unidad de Tecnología e Inteligencia Artificial en la Educación, Futuro del Aprendizaje y la Innovación, por coordinar la producción de la publicación.

También damos gracias a Jenny Webster por la corrección de estilo y de pruebas del texto, y a Ngoc-Thuy Tran por el diseño del formato.

Índice

Prefacio	2
Agradecimientos	3
Acrónimos y abreviaturas	6
Introducción	7
1. ¿Qué es la IA generativa y cómo funciona?	8
1.1 ¿Qué es la IA generativa?	8
1.2 ¿Cómo funciona la IA generativa?	8
1.2.1 Cómo funcionan los modelos de IAGen de texto	9
1.2.2 Cómo funcionan los modelos de IAGen de imágenes	11
1.3 Ingeniería de <i>prompts</i> para generar los resultados deseados	12
1.4 Surgimiento de EdGPT y sus implicancias	13
2. Controversias en torno a la IA generativa y sus implicancias para la educación	14
2.1 Empeoramiento de la pobreza digital	14
2.2 Superar las adaptaciones regulatorias nacionales	14
2.3 Uso de contenidos sin consentimiento	15
2.4 Modelos inexplicables utilizados para generar resultados	15
2.5 Contenido generado por IA que contamina Internet	16
2.6 Falta de comprensión del mundo real	16
2.7 Reducir la diversidad de opiniones y marginar aún más a las voces ya marginadas	17
2.8 Generación de <i>deepfakes</i> más profundos	17
3. Regulación del uso de la IA generativa en la educación	18
3.1 Un enfoque centrado en el ser humano para la IA	18
3.2 Pasos para regular la IAGen en la educación	18
3.3 Regulaciones sobre la IAGen: Elementos clave	20
3.3.1 Organismos reguladores gubernamentales	20
3.3.2 Proveedores de herramientas de IAGen	22
3.3.3 Usuarios institucionales	23
3.3.4 Usuarios individuales	23
4. Hacia un marco de políticas para el uso de la IA generativa en la educación y la investigación	24
4.1 Promover la inclusión, la equidad y la diversidad lingüística y cultural	24
4.2 Proteger la acción humana	25
4.3 Monitorear y validar los sistemas de IAGen para la educación	25
4.4 Desarrollar competencias en IA, incluyendo habilidades de IAGen para los estudiantes	26

4.5	Desarrollar las capacidades de docentes e investigadores para hacer un uso adecuado de IAGen.....	26
4.6	Promover opiniones diversas y la expresión plural de ideas	27
4.7	Testear localmente modelos de aplicaciones relevantes y crear una base acumulativa de evidencias	27
4.8	Considerar las implicancias a largo plazo, de forma intersectorial e interdisciplinaria	28
5.	Facilitar el uso creativo de la IAGen en la educación y la investigación	29
5.1	Estrategias institucionales para facilitar el uso responsable y creativo de la IAGen	29
5.2	Un enfoque de “interacción centrada en el ser humano y pedagógicamente apropiada”	30
5.3	Codiseño del uso de la IAGen en la educación y la investigación.....	30
5.3.1	IA generativa para la investigación.....	30
5.3.2	IA generativa para facilitar la enseñanza	31
5.3.3	IA generativa como entrenador 1:1 para la adquisición autodidacta de habilidades básicas	32
5.3.4	IA generativa para facilitar la investigación o el aprendizaje basado en proyectos.....	34
5.3.5	IA generativa como apoyo a estudiantes con necesidades especiales	35
6.	IAGen y el futuro de la educación y la investigación.....	37
6.1	Cuestiones éticas inexploradas.....	37
6.2	Derechos de autor y propiedad intelectual.....	37
6.3	Fuentes de contenido y aprendizaje.....	37
6.4	Respuestas homogeneizadas frente a resultados diversos y creativos	38
6.5	Replanteamiento de la evaluación y los resultados de aprendizaje	38
6.6	Procesos de reflexión.....	39
	Conclusiones finales.....	39
	Referencias.....	40

Lista de tablas

Tabla 1.	Técnicas usadas en IA generativa	8
Tabla 2.	GPTs de OpenAI	9
Tabla 3.	Codiseño de usos de la IAGen para la investigación.....	31
Tabla 4.	Codiseño de usos de la IAGen como apoyo a docentes y a la enseñanza	32
Tabla 5.	Codiseño de usos de la IAGen como entrenador 1:1 para la adquisición autodidacta de habilidades básicas en idiomas y en las artes	33
Tabla 6.	Codiseño de usos de la IAGen para facilitar la investigación o el aprendizaje basado en proyectos	35
Tabla 7.	Codiseño de usos de la IAGen para apoyar a estudiantes con necesidades especiales.....	35

Acrónimos y abreviaturas

Conceptos y tecnologías

AA	Aprendizaje automático
GB	Gigabytes
GPT	Transformador generativo preentrenado, por sus siglas en inglés
IA	Inteligencia artificial
IAD	Inteligencia artificial distribuida
IAG	Inteligencia artificial general
IAGen	Inteligencia artificial generativa
IPA	Interfaz de programación de aplicaciones
LaMDA	Modelo de lenguaje para aplicaciones de diálogo, por sus siglas en inglés
LLM	Modelo de lenguaje de gran tamaño, por sus siglas en inglés
RNA	Red neuronal artificial
RGA	Red generativa antagónica
RGPD	Reglamento General de Protección de Datos
TIC	Tecnología de la información y la comunicación
VAE	Autocodificadores variacionales, por sus siglas en inglés

Organizaciones

ACC	Administración del Ciberespacio de China
AGCC	Clúster Gubernamental de IA en la Nube (Singapur)
OCDE	Organización para la Cooperación y el Desarrollo Económicos
UE	Unión Europea
UNCTAD	Conferencia de las Naciones Unidas sobre Comercio y Desarrollo
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Introducción

El lanzamiento de ChatGPT a finales de 2022, la primera herramienta de inteligencia artificial generativa (IAGen) de uso sencillo y amplia disponibilidad para el público,¹ seguida de versiones iterativamente más sofisticadas, provocó una conmoción en todo el mundo y está impulsando una carrera entre las grandes empresas tecnológicas para posicionarse en el campo del desarrollo de modelos de IAGen.²

En todo el mundo, la preocupación inicial en el ámbito educativo era que ChatGPT y herramientas similares de IAGen fueran utilizadas por los estudiantes para hacer trampa en sus tareas, socavando el valor de la evaluación del aprendizaje, la certificación y las calificaciones (Anders, 2023). Mientras algunas instituciones educativas prohibieron su uso, otras recibieron cautelosamente la llegada de la IAGen (Tlili, 2023). Muchas escuelas y universidades, por ejemplo, adoptaron un acercamiento progresivo con la creencia de que, “más que intentar prohibir su uso, los estudiantes y el personal necesitan ser apoyados para utilizar herramientas de IAGen de forma efectiva, ética y transparente” (Russell Group, 2023). Este enfoque reconoce la vasta disponibilidad de la IAGen, que probablemente se volverá cada vez más sofisticada y que cuenta con un potencial negativo específico y un potencial positivo único para la educación.

De hecho, la IAGen tiene infinidad de usos posibles. Puede automatizar el tratamiento de la información y la presentación de resultados en todas las representaciones simbólicas clave del pensamiento humano. Permite la entrega de resultados, ofreciendo productos de conocimiento semiacabados. Al liberar a los humanos de algunas categorías de habilidades de pensamiento de orden inferior, esta nueva generación de herramientas de IA podría tener profundas implicancias en la forma en que entendemos la inteligencia y el aprendizaje humanos.

No obstante, la IAGen también plantea múltiples problemas inmediatos relacionados con cuestiones como la seguridad, la privacidad de los datos, los derechos de autor y la manipulación. Algunos de ellos son riesgos más amplios relacionados con la inteligencia artificial que se han visto exacerbados por la IAGen, mientras que otros han surgido recientemente con esta última generación de herramientas. Ahora urge comprender y abordar plenamente cada una de estas cuestiones y preocupaciones.

Esta guía está diseñada para responder a dicha necesidad urgente. Sin embargo, un conjunto temático de orientaciones sobre IAGen para la educación no debe entenderse como una afirmación de que sea la solución a los desafíos fundamentales de la educación. A pesar de la hipérbole mediática, es poco probable que la IAGen por sí sola resuelva cualquiera de los problemas que enfrentan los sistemas educativos de todo el mundo. Para responder a los problemas educativos de larga data, es clave mantener la idea de que la capacidad humana y la acción colectiva, y no la tecnología, son los factores determinantes para encontrar soluciones eficaces a los desafíos fundamentales enfrentados por las sociedades.

Por ello, esta guía pretende apoyar el planeamiento de normativas, políticas y programas de desarrollo de capacidades humanas adecuados, para garantizar que la IAGen se convierta en una herramienta que realmente beneficie y empodere docentes, estudiantes e investigadores. Basándose en la *Recomendación sobre la Ética de la Inteligencia Artificial* de la UNESCO, la guía está fundamentada en un enfoque centrado en el ser humano que promueve la actuación humana, la inclusión, la equidad, la igualdad de género y la diversidad cultural y lingüística, así como las opiniones y expresiones plurales.

La guía analiza, en primer lugar, qué es la IAGen y cómo funciona. Presenta las diversas tecnologías y modelos disponibles (Sección 1), antes de identificar una serie de cuestiones éticas y políticas controversiales en torno tanto a la IA en general como a la IAGen en particular (Sección 2). A continuación, se analizan los pasos y los elementos clave que deben examinarse para regular la IAGen desde un enfoque centrado en el ser humano, que garantice un uso ético, seguro, equitativo y significativo (Sección 3). La Sección 4 propone medidas que pueden tomarse para desarrollar marcos políticos coherentes e integrales para regular el uso de la IAGen en la educación y la investigación, mientras que la Sección 5 examina las posibilidades de utilizar creativamente la IAGen en el diseño curricular, la enseñanza, el aprendizaje y las actividades de investigación. La Sección 6 concluye la guía con consideraciones en torno a las implicancias a largo plazo de la IAGen para la educación y la investigación.

1. ¿Qué es la IA generativa y cómo funciona?

1.1 ¿Qué es la IA generativa?

La IA Generativa (IAGen) es una tecnología de inteligencia artificial (IA) que genera contenidos de forma automática en respuesta a instrucciones escritas en interfaces conversacionales de lenguaje natural (*prompts*). En vez de limitarse a conservar las páginas web existentes, IAGen produce nuevos contenidos. El contenido puede presentarse en formatos que abarcan todas las representaciones simbólicas del pensamiento humano: textos escritos en lenguaje natural, imágenes (incluyendo fotografías, pinturas digitales y caricaturas), videos, música y código de software. La IAGen se entrena utilizando datos obtenidos de páginas web, conversaciones en redes sociales y otros medios en línea. Genera su contenido analizando estadísticamente la distribución de palabras, píxeles u otros elementos en los datos que ha asimilado e identificando y repitiendo patrones comunes (por ejemplo, qué palabras siguen habitualmente a cuáles otras).

Aunque la IAGen puede producir nuevos contenidos, no puede crear nuevas ideas o soluciones a los retos del mundo real, ya que no entiende objetos reales ni las relaciones sociales que sustentan el lenguaje. Además, a pesar de su fluidez y su impresionante capacidad de producción, no se puede confiar en la precisión de la IAGen. De hecho, hasta el proveedor de ChatGPT

reconoce que “aunque herramientas como ChatGPT pueden generar frecuentemente respuestas que suenan razonables, no se puede confiar en que sean certeras” (OpenAI, 2023). La mayoría de las veces, los errores pasarán desapercibidos a menos que el usuario posea sólidos conocimientos sobre el tema en cuestión.

1.2 ¿Cómo funciona la IA generativa?

Las tecnologías específicas en las que se basa la IAGen forman parte de la familia de tecnologías de IA denominada aprendizaje automático (AA), que utiliza algoritmos que le permiten mejorar de forma continua y automática su rendimiento a partir de los datos. El tipo de AA que ha dado lugar a muchos de los avances en IA que hemos visto en los últimos años, como la IA de reconocimiento facial, se conoce como redes neuronales artificiales (RNAs), que se inspiran en el funcionamiento del cerebro humano y sus conexiones sinápticas entre neuronas. Hay muchos tipos de RNAs.

Tanto las tecnologías de IA generativa de texto como las de imagen, se basan en un conjunto de tecnologías de IA que llevan varios años a disposición de los investigadores.¹ ChatGPT, por ejemplo, utiliza un transformador generativo preentrenado (GPT), mientras que la IAGen de imagen suele utilizar lo que se conoce como redes generativas antagónicas (RGAs) (ver **Tabla 1**).³

Tabla 1. Técnicas usadas en IA generativa

Aprendizaje automático (AA)		Un tipo de IA que utiliza datos para mejorar automáticamente su desempeño.
Red neuronal artificial (RNA)		Un tipo de AA inspirado en la estructura y funcionamiento del cerebro humano (como las conexiones sinápticas entre neuronas).
IA generativa de texto	Transformador de propósito general	Un tipo de RNA capaz de hacer foco en diferentes partes de los datos para determinar cómo se relacionan entre sí.
	Modelo de lenguaje de gran tamaño (LLM)	Un tipo de transformador de propósito general que es entrenado con vastas cantidades de datos textuales.
	Transformador generativo preentrenado (GPT)⁴	Un tipo de LLM que es preentrenado con cantidades aún mayores de datos, lo cual permite que el modelo capture los matices del lenguaje y genere textos coherentes en función del contexto.
IA generativa de imágenes	Red generativa antagónica (RGA)	Tipos de redes neuronales utilizadas para la generación de imágenes.
	Autocodificador variacional (VAE)	

1.2.1 Cómo funcionan los modelos de IAGen de texto

La IA generativa de textos utiliza un tipo de RNA conocido como transformador de propósito general, y un tipo de transformador de propósito general llamado modelo de lenguaje de gran tamaño. Por eso, los sistemas de IAGen de texto suelen denominarse modelos de lenguaje de gran tamaño, o LLM. El tipo de LLM utilizado por la IAGen se conoce como transformador generativo preentrenado o GPT, por sus siglas en inglés (de ahí el “GPT” en “ChatGPT”).

ChatGPT está basado en GPT-3, desarrollado por OpenAI. Esta fue la tercera evolución de su GPT; la primera fue lanzada en 2018 y la más reciente, la GPT-4, en marzo de 2023 (ver **Tabla 2**). Cada GPT de OpenAI mejoró iterativamente la versión anterior a través de avances en arquitecturas de IA, métodos de entrenamiento y técnicas de optimización. Una faceta

bien conocida de su progreso continuo es el uso de volúmenes cada vez mayores de datos para entrenar su cantidad exponencialmente creciente de “parámetros”. Los parámetros pueden considerarse metafóricamente como perillas que se ajustan para ajustar el rendimiento del GPT. Incluyen los “pesos” del modelo, parámetros numéricos que determinan la forma en que el modelo procesa sus entradas y produce sus resultados.

Además de los avances en la optimización de las arquitecturas de IA y de los métodos de entrenamiento, esta rápida evolución también ha sido posible gracias a las ingentes cantidades de datos⁵ y a las mejoras en las capacidades informáticas disponibles para las grandes empresas. Desde 2012, las capacidades informáticas utilizadas para entrenar modelos de IAGen se han duplicado cada 3-4 meses. Para comparar, la Ley de Moore mencionaba un periodo de duplicación de dos años (OpenAI, 2018; Universidad de Stanford, 2019).

Tabla 2. GPTs de OpenAI

Modelo	Lanzamiento	Cantidad de datos para entrenamiento	Cantidad de parámetros	Características
GPT-1	2018	40 GB	117 millones	Capaz de realizar tareas de procesamiento del lenguaje natural, como completar textos y responder preguntas.
GPT-2	2019	40 GB	1,500 millones	Capaz de realizar tareas más complejas de procesamiento del lenguaje natural, como traducción automática y elaboración de resúmenes.
GPT-3	2020	17,000 GB	175,000 millones	Capaz de realizar tareas avanzadas de procesamiento del lenguaje natural, como escribir párrafos coherentes y generar artículos completos. También es capaz de adaptarse a nuevas tareas a partir de unos pocos ejemplos.
GPT-4 ⁶	2023	1,000,000 GB (informado, sin confirmación)	170,000,000 millones (informado, sin confirmación)	Confiabilidad aumentada, capacidad para procesar instrucciones más complejas.

Una vez que el GPT ha sido entrenado, la generación de una respuesta de texto a un prompt involucra los siguientes pasos:

1. El *prompt* es dividido en unidades más pequeñas (llamadas *tokens*) que son ingresadas en el GPT.
2. El GPT utiliza patrones estadísticos para predecir las posibles palabras o frases que podrían conformar una respuesta coherente al *prompt*.
 - El GPT identifica patrones de palabras y frases que suelen coexistir en su modelo
 - prediseñado de datos de gran tamaño (que incluye texto extraído de Internet y de otros sitios).
 - Utilizando estos patrones, el GPT estima la probabilidad de que aparezcan palabras o frases específicas en un contexto determinado.
 - Partiendo de una predicción aleatoria, el GPT utiliza estas probabilidades estimadas para predecir la siguiente palabra o frase probable en su respuesta.
3. Las palabras o frases predichas son convertidas en texto legible.

4. El texto legible es filtrado a través de lo que se conoce como “barandillas” (*guardrails*) para eliminar cualquier contenido ofensivo.
5. Los pasos 2 a 4 se repiten hasta que finaliza una respuesta. La respuesta se considera finalizada cuando alcanza un límite máximo de *tokens* o cumple con criterios de finalización predefinidos.
6. La respuesta se postprocesa para mejorar su legibilidad aplicando formato, puntuación y otras adaptaciones (como empezar la respuesta con palabras que podría utilizar un humano, como “Claro”, “Por supuesto” o “Lo siento”).

Aunque los GPTs y su capacidad para generar texto automáticamente han estado disponibles para los investigadores desde 2018, el lanzamiento de ChatGPT fue novedoso por su acceso gratuito a través de una interfaz de uso fácil, lo que significa que cualquier persona con acceso a Internet podría explorar la herramienta. El lanzamiento de ChatGPT causó conmoción en todo el mundo y rápidamente impulsó a otras empresas tecnológicas transnacionales a readecuarse al nuevo paradigma, sucediendo lo mismo con numerosas start-ups. Esto debería llevarse a cabo tanto lanzando sistemas similares de desarrollo propio como creando herramientas basadas en el aprovechamiento del recurso pionero.

Hacia julio de 2023, algunas de las alternativas a ChatGPT incluían a las siguientes:

- **Alpaca:**⁷ Una versión depurada del Llama de Meta, de la Universidad de Stanford, que apunta a abordar la información falsa, los estereotipos sociales y el lenguaje tóxico de los LLM.
- **Bard:**⁸ Un LLM de Google, basado en sus sistemas LaMDA y PaLM 2, con acceso a Internet en tiempo real, lo cual le permite ofrecer información actualizada.
- **Chatsonic:**⁹ Desarrollado por Writesonic, está basado en ChatGPT y rastrea los datos directamente desde Google. En consecuencia, tiene menos posibilidades de generar respuestas incorrectas.
- **Ernie** (conocido también como **Wenxin Yiyan** 文心一言):¹⁰ Un LLM bilingüe de Baidu, aún en desarrollo, que integra extensos conocimientos con conjuntos de datos masivos para generar texto e imágenes.

- **Hugging Chat:**¹¹ Creado por Hugging Face, que puso énfasis en la ética y la transparencia durante todo su desarrollo, entrenamiento e implementación. Todos los datos utilizados para entrenar sus modelos son de código abierto.
- **Jasper:**¹² Un paquete de herramientas y IPA que, por ejemplo, puede ser entrenado para escribir en el estilo particular preferido de un usuario. También puede generar imágenes.
- **Llama:**¹³ Un LLM de código abierto de Meta que requiere menos potencia computacional y menos recursos para probar nuevos enfoques, validar el trabajo de otros y explorar nuevos casos de uso.
- **Open Assistant:**¹⁴ Un sistema de código abierto diseñado para que cualquier persona con conocimientos suficientes pueda desarrollar su propio LLM. Ha sido creado a partir de datos de entrenamiento recogidos por voluntarios.
- **Tongyi Qianwen** (通义千问):¹⁵ Un LLM de Alibaba que puede responder a indicaciones en inglés o chino. Está siendo integrado al paquete de herramientas de negocios de Alibaba.
- **YouChat:**¹⁶ Un LLM que incorpora funciones de búsqueda en tiempo real para brindar contexto y perspectivas adicionales a fin de generar resultados más precisos y confiables.

La mayoría son de uso gratuito (dentro de ciertos límites), mientras que algunos son de código abierto. Se están lanzando muchos otros productos basados en alguno de estos LLM. Ejemplos de ello son los siguientes:

- **ChatPDF:**¹⁷ Resume y responde preguntas sobre documentos PDF ingresados.
- **Elicit: The AI Research Assistant:**¹⁸ Apunta a automatizar partes de los flujos de trabajo de los investigadores, identificando artículos relevantes y resumiendo la información clave.
- **Perplexity:**¹⁹ Ofrece un “centro de conocimiento” para las personas que buscan respuestas rápidas, precisas y personalizadas en función de sus necesidades.

De manera similar, las herramientas basadas en LLM están siendo integradas en otros productos, como los

navegadores web. Por ejemplo, las extensiones para Chrome basadas en ChatGPT incluyen las siguientes:

- **WebChatGPT:**²⁰ Permite que ChatGPT acceda a Internet para mantener conversaciones más precisas y actualizadas.
- **Compose AI:**²¹ Autocompleta oraciones en correos electrónicos y en otros lugares.
- **TeamSmart AI:**²² Ofrece un “equipo de asistentes virtuales”.
- **Wiseone:**²³ Simplifica la información en línea.

Por otra parte, ChatGPT ha sido integrado a algunos motores de búsqueda,²⁴ y está siendo implementado en grandes conjuntos de herramientas de productividad (como Microsoft Word y Excel), lo que lo hace aún más accesible en oficinas e instituciones educativas de todo el mundo (Murphy Kelly, 2023).

Finalmente, como una interesante transición hacia la IAGen de imágenes, el GPT más reciente de OpenAI, GPT-4, puede aceptar imágenes, además de texto, en sus *prompts*, por lo que adquiere funciones multimodales. En consecuencia, hay quienes sostienen que la denominación “modelo de lenguaje de gran tamaño” (LLM) es cada vez menos apropiada, por lo que investigadores de la Universidad de Stanford han propuesto el término “modelo fundacional” (Bommasani et al., 2021). La alternativa no ha sido plenamente adoptada aún.

1.2.2 Cómo funcionan los modelos de IAGen de imágenes

Los modelos de IAGen de imágenes y de música suelen utilizar otro tipo de RNA conocidas como redes generativas antagónicas (RGAs), que también pueden combinarse con autocodificadores variacionales. Las RGAs constan de dos partes (dos “adversarios”), el “generador” y el “discriminador”. En el caso de las RGAs de imágenes, el generador crea una imagen aleatoria en respuesta a un *prompt*, y el discriminador intenta distinguir entre la imagen generada y las reales. Posteriormente, el generador utiliza el resultado del discriminador para ajustar sus parámetros y crear otra imagen. El proceso se repite, quizás miles de veces, y el generador crea imágenes cada vez más realistas que el discriminador distingue progresivamente menos de las reales. Por ejemplo, una RGA entrenado con éxito sobre un conjunto de datos de miles de fotografías de paisajes podría generar imágenes de paisajes nuevas, aunque irreales, casi indistinguibles de las reales. Por otra parte, un GAN entrenado en un conjunto de datos

de música popular (o incluso de música de un mismo artista) puede generar nuevas piezas musicales que respeten la estructura y la complejidad de la música original.

Desde julio de 2023, los modelos de **IAGen de imágenes** disponibles incluyen a los siguientes, todos los cuales generan imágenes a partir de *prompts* de texto. La mayoría son de uso libre, dentro de ciertos límites:

- **Craiyon:**²⁵ Antes conocido como DALL·E mini.
- **DALL·E 2:**²⁶ Herramienta de IAGen de imágenes de OpenAI.
- **DreamStudio:**²⁷ Herramienta de IAGen de imágenes de Stable Diffusion.
- **Fotor:**²⁸ Integra la IAGen a un conjunto de herramientas de edición de imágenes.
- **Midjourney:**²⁹ Una herramienta de IAGen de imágenes independiente.
- **NightCafe:**³⁰ Interfaz entre Stable Diffusion y DALL·E 2.
- **Photosonic:**³¹ Generador de arte de IA de WriteSonic.

Los siguientes son ejemplos de **IAGen de video** de fácil acceso:

- **Elai:**³² Puede convertir presentaciones, sitios web y texto en videos.
- **GliaCloud:**³³ Puede generar videos a partir de noticias, publicaciones de redes sociales, eventos deportivos en vivo y datos estadísticos.
- **Pictory:**³⁴ Puede crear automáticamente videos cortos a partir de contenidos de gran tamaño.
- **Runway:**³⁵ Ofrece un conjunto de herramientas para generación y edición de videos (e imágenes).

Por último, estos son algunos ejemplos de **IAGen de música** de fácil acceso:

- **Aiva:**³⁶ Puede crear automáticamente pistas de música personalizadas.
- **Boomy,**³⁷ **Soundraw,**³⁸ y **Voicemod:**³⁹ Pueden generar canciones a partir de textos, sin requerir conocimientos de composición musical.

1.3 Ingeniería de *prompts* para generar los resultados deseados

Aunque utilizar la IAGen puede ser tan sencillo como digitar una pregunta u otro tipo de *prompt*, la realidad es que aún no es fácil que el usuario obtenga exactamente el resultado deseado. Por ejemplo, la innovadora imagen de IA *Théâtre D'opéra Spatial*, ganadora de un premio en la Feria del Estado de Colorado, Estados Unidos, requirió semanas de redacción de *prompts* y retoques de cientos de imágenes para generar el resultado final (Roose, 2022). El desafío similar que supone escribir *prompts* efectivos de IAGen de textos ha hecho que cada vez aparezcan más puestos de trabajo de ingeniería de *prompts* en los sitios web de reclutamiento (Popli, 2023). La “ingeniería de *prompts*” hace referencia a los procesos y técnicas de composición de entradas para producir un resultado de IAGen que se aproxime más a la intención original del usuario.

La ingeniería de *prompts* es más exitosa cuando el *prompt* articula una cadena coherente de razonamiento centrada en un problema particular, o una cadena de pensamiento en un orden lógico. Entre las recomendaciones específicas cabe citar:

- Utilizar un lenguaje **simple**, claro y directo que pueda entenderse fácilmente, evitando una redacción compleja o ambigua.
- Incluir **ejemplos** que ilustren la respuesta deseada o el formato de las respuestas generadas.



Implicancias para la educación y la investigación

Aunque la IAGen puede ayudar a que docentes e investigadores generen textos útiles y otros resultados que respalden su trabajo, no se trata necesariamente de un proceso sencillo. Pueden ser necesarias varias iteraciones de un *prompt* antes de obtener el resultado deseado. Es preocupante que los jóvenes estudiantes, al ser por definición menos expertos que los docentes, puedan aceptar, sin saberlo y sin mayor involucramiento crítico, resultados de IAGen superficiales, inexactos o incluso dañinos.

- Incluir **contexto**, crucial para generar respuestas relevantes y significativas.
- **Refinar** e iterar según sea necesario, experimentando con diferentes variaciones.
- Ser **ético**, evitando *prompts* que generen contenidos inapropiados, sesgados o dañinos.

También es importante reconocer desde ya que no se puede confiar en los resultados de la IAGen sin realizar una evaluación crítica. Como escribe OpenAI acerca de su GPT más sofisticado:⁴⁰

“Más allá de sus capacidades, GPT-4 tiene limitaciones similares a las de los GPTs precedentes. Sigue sin ser totalmente confiable (‘alucina’ hechos y comete errores de razonamiento). Se debe tener mucho cuidado al utilizar los resultados de los modelos lingüísticos, sobre todo en contextos de alto nivel, debiéndose seguir un protocolo exacto (como la revisión humana, la fundamentación con contexto adicional o evitando los usos de alto nivel) que se corresponda con las necesidades de un caso de uso específico.”

A la luz de la calidad de los resultados de la IAGen, deberían llevarse a cabo rigurosas pruebas de usuario y evaluaciones de rendimiento antes de validar las herramientas para su adopción a gran escala o de alto nivel. Tales ejercicios deberían diseñarse con la métrica de rendimiento más relevante en función del tipo de tarea para la que los usuarios piden a la IAGen que ofrezca resultados. Por ejemplo, para resolver problemas matemáticos, la “precisión” podría utilizarse como métrica principal para cuantificar la frecuencia con la que una herramienta de IAGen produce la respuesta correcta; para responder a preguntas delicadas, la métrica principal para medir el rendimiento podría ser la “tasa de respuesta” (la frecuencia con la que la IAGen responde directamente a una pregunta); para la generación de código, la métrica podría ser “la fracción de código generado que es directamente ejecutable” (si el código generado pudiera ejecutarse directamente en un entorno de programación y pasar las pruebas de unidad); y para el razonamiento visual, la métrica podría ser la “coincidencia exacta” (si los objetos visuales generados coinciden exactamente con la realidad) (Chen et al., 2023).

En resumen, a un nivel superficial, la IAGen es fácil de usar; sin embargo, los resultados más sofisticados requieren la intervención de personas calificadas y deben ser evaluados críticamente antes de ser utilizados.

1.4 Surgimiento de EdGPT y sus implicancias

Considerando que los modelos de la IAGen pueden servir de base o punto de partida para desarrollar modelos más especializados o específicos de un dominio, algunos investigadores han sugerido que los GPTs deberían renombrarse como “modelos fundacionales” (Bommasani et al., 2021). En la educación, los desarrolladores e investigadores han empezado a pulir un modelo fundacional para desarrollar “EdGPT”.⁴¹ Los modelos EdGPT se entrenan con datos específicos para fines educativos. En otras palabras, EdGPT pretende refinar el modelo derivado de cantidades masivas de datos de entrenamiento generales a cantidades menores de datos educativos de dominio específico y de alta calidad.

Esto le da potencialmente a EdGPT un mayor alcance para sustentar el logro de las transformaciones listadas en la Sección 4.3. Por ejemplo, los modelos EdGPT orientados al codiseño curricular podrían permitir a docentes y estudiantes generar materiales educativos apropiados, como planes de clases, cuestionarios y actividades interactivas estrechamente alineados con un enfoque pedagógico eficaz y con objetivos curriculares y niveles de desafío específicos para determinados estudiantes. Del mismo modo, en el contexto de un entrenador de habilidades lingüísticas 1:1, un modelo fundacional perfeccionado con textos apropiados para un idioma concreto podría utilizarse para generar oraciones, párrafos o conversaciones ejemplares para la práctica. Cuando los estudiantes interactúan con el modelo, éste puede responder con textos relevantes y gramaticalmente precisos en el nivel adecuado para ellos. En teoría, los resultados de los modelos EdGPT también podrían contener menos sesgos generales o contenidos cuestionables que el GPT estándar, pero, aun así, podrían generar errores. Es fundamental señalar que, a menos que los modelos y el enfoque subyacentes de la IAGen cambien

significativamente, EdGPT podría seguir generando errores y mostrar otras limitaciones. Por ello, sigue siendo importante que los principales usuarios de EdGPT, especialmente docentes y estudiantes, adopten una perspectiva crítica frente a cualquier resultado.

En la actualidad, el refinamiento de modelos fundacionales para un uso más específico del GPT en la educación se encuentra en una fase inicial. Entre los ejemplos existentes se encuentra EduChat, un modelo fundacional desarrollado por la Universidad Normal de China Oriental para prestar servicios de enseñanza y aprendizaje, y cuyos códigos, datos y parámetros son de código abierto.⁴² Otro ejemplo es MathGPT, desarrollado por el TAL Education Group, un LLM que se enfoca en la resolución de problemas relacionados con las matemáticas y en dar conferencias para usuarios de todo el mundo.⁴³

Sin embargo, antes de que sea posible realizar avances significativos, es esencial realizar esfuerzos para perfeccionar los modelos fundacionales, no solo añadiendo conocimientos de las asignaturas y eliminando sesgos, sino también agregando conocimientos sobre métodos de aprendizaje relevantes, y sobre cómo esto puede reflejarse en el diseño de algoritmos y modelos. El desafío reside en determinar hasta dónde los modelos EdGPT pueden ir más allá del conocimiento de las asignaturas para apuntar también a la pedagogía centrada en el estudiante y a las interacciones positivas docente-estudiante. Otro desafío será determinar hasta dónde pueden recogerse y utilizarse éticamente los datos de estudiantes y docentes para alimentar un EdGPT. Finalmente, también es necesaria una investigación robusta que garantice que EdGPT no socave los derechos humanos de los estudiantes ni desempodere a los profesores.

2. Controversias en torno a la IA generativa y sus implicancias para la educación

Luego de haber analizado qué es y cómo funciona la IAGen, esta sección examina las controversias y los riesgos éticos planteados por todos los sistemas de IAGen y considera algunas de sus implicancias para la educación.

2.1 Empeoramiento de la pobreza digital

Como ha sido señalado, la IAGen depende de inmensas cantidades de datos y de una enorme potencia computacional, más allá de sus innovaciones iterativas en arquitecturas de IA y métodos de entrenamiento que, en su mayoría, solo están al alcance de las más grandes empresas tecnológicas internacionales y de unas pocas economías (como Estados Unidos, la República Popular China y, en menor medida, Europa). Esto significa que la posibilidad de crear y controlar la IAGen está fuera del alcance de la mayoría de las compañías y países, especialmente los del Sur Global.

Debido a que el acceso a los datos resulta cada vez más esencial para el desarrollo económico de los países y para las oportunidades digitales de los individuos, los países y personas que no tienen acceso o no pueden permitirse suficientes datos quedan en una situación de “pobreza de datos” (Marwala, 2023). La situación es similar en cuanto al acceso a la potencia computacional. La rápida penetración de la IAGen en países y regiones tecnológicamente avanzados aceleró exponencialmente

la generación y el procesamiento de datos y, al mismo tiempo, intensificó la concentración de la riqueza de la IA en el Norte Global. Como consecuencia inmediata, las regiones pobres en datos han quedado aún más excluidas y expuestas a largo plazo al riesgo de ser colonizadas por los estándares integrados en los modelos GPT. Los actuales modelos de ChatGPT están entrenados con datos de usuarios en línea que reflejan los valores y normas del Norte Global, lo cual los hace inapropiados para algoritmos de IA localmente relevantes en comunidades pobres en datos de muchas partes del Sur Global o en comunidades menos favorecidas del Norte Global.

2.2 Superar las adaptaciones regulatorias nacionales

Los principales proveedores de IAGen también han sido criticados por no permitir que sus sistemas sean sometidos a revisiones académicas rigurosas e independientes (Dwivedi et al., 2023).⁴⁴ Las tecnologías fundacionales de la IAGen de una empresa tienden a protegerse como propiedad intelectual corporativa. Mientras tanto, muchas de las empresas que están empezando a utilizar la IAGen encuentran cada vez más dificultades para mantener la seguridad de sus sistemas (Lin, 2023). Además, a pesar de los llamamientos a la regulación por parte de la propia industria de la IA,⁴⁵ la creación de legislación sobre elaboración y uso de todo tipo de IA, incluida la IAGen, suele ir detrás de su rápido ritmo de desarrollo. Esto explica parcialmente los desafíos experimentados por los organismos nacionales o locales para comprender y regular las cuestiones jurídicas y éticas.⁴⁶

Si bien la IAGen puede aumentar las capacidades humanas para realizar determinadas tareas, el control democrático de las empresas promotoras de la IAGen es limitado. Esto plantea la cuestión de la regulación, en particular con respecto al acceso y uso de datos domésticos, incluidos los de instituciones e individuos locales, así como los datos generados en los territorios nacionales. Se necesita una legislación adecuada para que los organismos gubernamentales locales puedan obtener cierto control sobre las crecientes oleadas de IAGen a efectos de garantizar su gobernanza como bien público.



Implicancias para la educación y la investigación

Los investigadores, los docentes y los estudiantes deberían adoptar una visión crítica de las orientaciones de valor, las normas culturales y las costumbres sociales integradas a los modelos de entrenamiento de la IAGen. Los formuladores de políticas deberían tomar conciencia del agravamiento de las desigualdades causadas por la creciente brecha en el entrenamiento y control de los modelos de IAGen, y adoptar medidas al respecto.



Implicancias para la educación y la investigación

Los investigadores, docentes y estudiantes deben ser conscientes de la falta de regulaciones adecuadas para proteger la propiedad de las instituciones nacionales e individuos y los derechos de los usuarios domésticos de la IAGen, así como para responder a las cuestiones legales que esta genera.

2.3 Uso de contenidos sin consentimiento

Como se ha señalado, los modelos de IAGen se construyen a partir de grandes cantidades de datos (tales como texto, sonidos, código e imágenes) frecuentemente obtenidos de Internet y, habitualmente, sin permiso de ningún propietario. Muchos sistemas de IAGen de imágenes y algunos de código han sido, en consecuencia, acusados de violar derechos de propiedad intelectual. A la fecha de redacción de estas líneas, existen varias demandas internacionales en curso en relación con la cuestión.

Además, se ha señalado que los GPTs pueden contravenir leyes como el Reglamento General de Protección de Datos de la Unión Europea (2016) o GDPR, especialmente el derecho de las personas a ser olvidadas, dado que actualmente es imposible eliminar los datos de alguien (o los resultados de esos datos) de un modelo GPT una vez que ha sido entrenado.



Implicancias para la educación y la investigación

- Los investigadores, docentes y estudiantes deben conocer los derechos de los propietarios de los datos y comprobar si las herramientas de IAGen que utilizan contravienen la normativa vigente.
- Los investigadores, docentes y estudiantes también deben ser conscientes de que las imágenes o códigos creados con IAGen pueden violar derechos de propiedad intelectual ajenos, y que las imágenes, sonidos o códigos que crean y comparten en Internet pueden ser explotados por otras IAGen.

2.4 Modelos inexplicables utilizados para generar resultados

Hace tiempo se sabe que las redes neuronales artificiales (RNAs) suelen ser “cajas negras”, es decir, que su funcionamiento interno no está abierto a la inspección. En consecuencia, las RNAs no son “transparentes” ni “explicables”, y no es posible establecer cómo obtienen sus resultados.

Mientras que el enfoque general, incluidos los algoritmos utilizados, es en su conjunto explicable, los modelos particulares y sus parámetros, incluidos los pesos del modelo, no son inspeccionables, razón por la cual no se puede explicar un resultado específico generado. En un modelo como el GPT-4 hay miles de millones de parámetros/pesos (ver **Tabla 2**) y son los pesos en su conjunto los que almacenan los patrones aprendidos que el modelo utiliza para generar sus resultados. Como los parámetros/pesos no son transparentes en las RNAs (**Tabla 1**), no es posible explicar la forma precisa en que estos modelos generan un resultado en particular.

La falta de transparencia y explicabilidad de la IAGen es cada vez más problemática a medida que esta se vuelve más compleja (ver **Tabla 2**). Esto produce frecuentemente resultados inesperados o no deseados. Por otra parte, los modelos de IAGen heredan y perpetúan los sesgos existentes en sus datos de entrenamiento que, dada la falta de transparencia de los modelos, son difíciles de detectar y abordar. Finalmente, esta opacidad es también una razón clave de los problemas de confianza en torno a la IAGen (Nazaretsky et al., 2022a). Si los usuarios no entienden de qué modo



Implicancias para la educación y la investigación

Los investigadores, docentes y estudiantes deben ser conscientes de que los sistemas de IAGen funcionan como cajas negras y que, en consecuencia, es difícil, si no imposible, saber por qué han sido creados determinados contenidos. La falta de explicación de cómo se generan los resultados tiende a encerrar a los usuarios en la lógica definida por los parámetros diseñados en los sistemas de IAGen. Estos parámetros pueden reflejar valores y normas culturales o comerciales específicos que sesgan implícitamente los contenidos producidos.

un sistema de IAGen ha llegado a un resultado en concreto, es menos probable que estén dispuestos a adoptarlo o utilizarlo (Nazaretsky et al., 2022b).

2.5 Contenido generado por IA que contamina Internet

Dado que los datos de entrenamiento de los GPTs suelen extraerse de Internet, que con mucha frecuencia incluye lenguaje discriminatorio u otros tipos de lenguaje inaceptable, los desarrolladores han debido implementar lo que ellos denominan “barandillas” para evitar que los resultados de los GPTs sean ofensivos y/o antiéticos. Sin embargo, debido a la ausencia de una normativa estricta y de mecanismos de control eficaces, los materiales sesgados generados por la IAGen se extienden cada vez más por Internet, contaminando una de las principales fuentes de contenidos o conocimientos para la mayoría de los estudiantes de todo el mundo. Esto es especialmente importante porque el material generado por la IAGen puede parecer bastante preciso y convincente, cuando a menudo contiene errores e ideas sesgadas. Esto supone un alto riesgo para los jóvenes estudiantes que carecen de conocimientos previos sólidos sobre el tema en cuestión. También supone un riesgo recursivo para los futuros modelos GPT que serán entrenados con textos obtenidos de Internet que los propios modelos GPT han creado y que también incluyen sus sesgos y errores.



Implicancias para la educación y la investigación

- Los investigadores, docentes y estudiantes deben ser conscientes de que los sistemas de IAGen pueden generar material ofensivo y antiético.
- También deben conocer los problemas a largo plazo que pueden surgir para la confiabilidad del conocimiento cuando los modelos GPT futuros estén basados en textos generados por modelos GPT anteriores.

2.6 Falta de comprensión del mundo real

A los GPTs de texto a veces se les denomina peyorativamente “loros estocásticos” porque, como se ha señalado antes, aunque pueden producir textos que parecen convincentes, ese texto suele contener errores y puede incluir afirmaciones dañinas (Bender et al., 2021). Esto ocurre porque los GPT solo repiten patrones lingüísticos hallados en sus datos de entrenamiento (normalmente texto obtenido de Internet), empezando por patrones aleatorios (o “estocásticos”), y sin comprender su significado, del mismo modo que un loro puede imitar sonidos sin comprender realmente lo que dice.

La desconexión entre los modelos de IAGen que “parecen” comprender el texto que utilizan y generan, y la “realidad” de que no comprenden el lenguaje ni el mundo real, puede llevar a profesores y estudiantes a depositar en el resultado un nivel de confianza injustificado. Esto plantea graves riesgos para el futuro de la educación. De hecho, la IAGen no está informada de observaciones del mundo real ni de otros aspectos clave del método científico, ni tampoco está alineada con valores humanos o sociales. Por estas razones, no puede generar contenidos realmente novedosos sobre el mundo real, los objetos y sus relaciones, las personas y las relaciones sociales, las relaciones humano-objeto o las relaciones humano-tecnología. No está claro que los contenidos aparentemente novedosos generados por los modelos de IAGen puedan ser reconocidos como conocimientos científicos.

Como ya se ha indicado, los GPT pueden producir con frecuencia textos inexactos o poco confiables. De hecho, es bien sabido que los GPTs inventan algunas cosas que no existen en la vida real. Algunos llaman a esto “alucinación”, aunque otros critican el uso de un término tan antropomórfico y, por lo tanto, engañoso. Esto lo reconocen las empresas que producen IAGen. En la parte inferior de la interfaz pública de ChatGPT, por ejemplo, se afirma: “ChatGPT puede producir información inexacta sobre personas, lugares o hechos”.²

Algunos defensores también han sugerido que la IAGen representa un paso importante en el camino hacia la inteligencia artificial general (IAG), término que sugiere una clase de IA más inteligente que los humanos. Sin embargo, esto ha sido criticado durante largo tiempo, con el argumento de que la IA nunca progresará hacia la IAG, al menos hasta que, de alguna manera, combine en simbiosis a la IA basada en el conocimiento (también conocida como IA simbólica

o basada en reglas) y a la IA basada en datos (también conocida como aprendizaje automático) (Marcus, 2022). Las afirmaciones sobre la IAG o la sintiencia también nos distraen de una consideración más cuidadosa sobre los daños que actualmente se están provocando con la IA, como la discriminación oculta contra grupos ya discriminados (Metz, 2021).



Implicancias para la educación y la investigación

- El resultado de una IAGen de texto puede parecer impresionantemente humano, como si entendiera el texto que ha generado. Sin embargo, la IAGen no entiende nada. En vez de eso, estas herramientas encadenan palabras de maneras que son habituales en Internet. El texto generado también puede ser incorrecto.
- Los investigadores, docentes y estudiantes deben ser conscientes de que un GPT no comprende el texto que genera; que puede generar, y a menudo lo hace, enunciados incorrectos; y que, por lo tanto, es preciso adoptar un enfoque crítico ante todo lo que sea generado.

2.7 Reducir la diversidad de opiniones y marginar aún más a las voces ya marginadas

ChatGPT y otras herramientas similares tienden a producir solo respuestas estándar que asumen los valores de los propietarios/creadores de los datos utilizados para entrenar los modelos. De hecho, si una secuencia de palabras aparece con frecuencia en los datos de entrenamiento -como es el caso de asuntos comunes y no controversiales y creencias dominantes- es probable que el GPT la repita en sus resultados.

Se corre así el riesgo de limitar y socavar el desarrollo de opiniones y expresiones de ideas plurales. Las poblaciones con pobreza de datos, incluidas las comunidades marginadas del Norte Global, tienen una mínima o limitada presencia digital en línea. Así pues, sus voces no llegan a hacerse oír y sus preocupaciones no están representadas en los datos utilizados para entrenar a los GPT, de modo que raramente están presentes en los resultados. Por estas razones, dada la metodología de preentrenamiento basada en datos de páginas web de Internet y conversaciones en redes sociales, los modelos de GPT pueden marginalizar aún más a personas ya de por sí en desventaja.



Implicancias para la educación y la investigación

- Mientras que los desarrolladores y proveedores de modelos IAGen tienen la responsabilidad principal de abordar continuamente los sesgos en los conjuntos de datos y los resultados de estos modelos, los investigadores, docentes y estudiantes como usuarios necesitan saber que los resultados de la IAGen de texto representan solo la visión más común o dominante del mundo al momento de producir sus datos de entrenamiento y que algunos de estos datos son problemáticos o sesgados (por ejemplo, los roles de género estereotipados).
- Los estudiantes, docentes e investigadores nunca deberían aceptar la información proporcionada por la IAGen al pie de la letra, sino evaluarla siempre de forma crítica.
- Los investigadores, docentes y estudiantes también deben ser conscientes del modo en que las voces minoritarias pueden verse excluidas, ya que estas son, por definición, menos frecuentes en los datos de entrenamiento.

2.8 Generación de deepfakes más profundos

Más allá de las controversias comunes a todas las IAGen, la tecnología RGA puede utilizarse para alterar o manipular imágenes o vídeos preexistentes a fin de generar otros falsos, difíciles de distinguir de los reales. La IAGen está facilitando cada vez más la creación de estos *deepfakes* y de las llamadas *fake news*, o noticias falsas. En otras palabras, la IAGen está facilitando que determinados actores cometan actos antiéticos, inmorales y delictivos, como difundir desinformación, promover discursos de odio e incorporar rostros de personas sin su conocimiento ni consentimiento, en materiales audiovisuales totalmente falsos y en ocasiones comprometedores.



Implicancias para la educación y la investigación

Aunque es obligación de los proveedores de IAGen proteger los derechos de autor y de imagen de los usuarios, los investigadores, docentes y estudiantes también deben ser conscientes de que cualquier imagen que compartan en Internet puede ser integrada en los datos de entrenamiento de la IAGen y podría ser manipulada y utilizada de forma antiética.

3. Regulación del uso de la IA generativa en la educación

Con el fin de abordar las controversias en torno a la IA generativa y aprovechar sus potenciales ventajas en la educación, es preciso, en primer lugar, regularla. La regulación de la IAGen con fines educativos requiere una serie de pasos y medidas de política fundamentadas en un enfoque centrado en el ser humano, para garantizar su uso ético, seguro, equitativo y significativo.

3.1 Un enfoque centrado en el ser humano para la IA

La *Recomendación sobre la Ética de la Inteligencia Artificial* (UNESCO, 2021) ofrece el marco normativo necesario para comenzar a abordar las múltiples controversias en torno a la IA generativa, incluyendo a las que atañen a la educación y la investigación. Está basada en un enfoque de la IA centrado en el ser humano que defiende que el uso de la IA esté al servicio del desarrollo de las capacidades humanas para un futuro inclusivo, justo y sostenible. Este enfoque debe estar orientado por los principios de los derechos humanos y la necesidad de proteger la dignidad de las personas y la diversidad cultural que define el patrimonio común del conocimiento. En términos de gobernanza, un enfoque centrado en el ser humano requiere una regulación apropiada que pueda asegurar la intervención humana, la transparencia y la rendición pública de cuentas.

El *Consenso de Beijing sobre la Inteligencia Artificial (IA) y la Educación* (2019) ahonda en lo que implica un enfoque centrado en el ser humano para el uso de la IA en el contexto de la educación. El Consenso afirma que el uso de las tecnologías de IA en la educación debería mejorar las capacidades humanas para el desarrollo sostenible y la colaboración eficaz entre humanos y máquinas en la vida, el aprendizaje y el trabajo. También reclama adoptar nuevas medidas que garanticen un acceso equitativo a la IA para apoyar a las personas marginadas y abordar las desigualdades, a la vez que se promueven las diversidades lingüísticas y culturales. El Consenso sugiere adoptar enfoques integrales, intersectoriales y de múltiples partes interesadas para la planificación de políticas sobre IA en la educación.

Inteligencia artificial y educación: Guía para las personas a cargo de formular políticas (UNESCO, 2022b) refina aún más las implicancias de un enfoque centrado en el ser humano al momento de examinar los beneficios y riesgos de la IA en la educación y el papel de esta última como medio para desarrollar competencias de IA. Propone recomendaciones concretas para la formulación de políticas que orienten el uso de la IA a fin de (i) permitir el acceso inclusivo a los programas de aprendizaje, especialmente para grupos vulnerables como los estudiantes con discapacidad; (ii) apoyar las opciones de aprendizaje personalizadas y abiertas; (iii) mejorar la provisión y la gestión basadas en datos para ampliar el acceso y mejorar la calidad del aprendizaje; (iv) monitorear los procesos de aprendizaje y alertar a los docentes sobre los riesgos de fracaso; y (v) desarrollar la comprensión y las habilidades para el uso ético y significativo de la IA.

3.2 Pasos para regular la IAGen en la educación

Antes del lanzamiento de ChatGPT, los gobiernos habían estado desarrollando o adaptando marcos para regular la recolección y el uso de datos, así como la adopción de sistemas de IA en todos los sectores, incluyendo la educación, lo que propició un contexto legislativo y de políticas para regular las nuevas aplicaciones de IA que surgían. Tras el lanzamiento de múltiples modelos competitivos de IAGen desde noviembre de 2022, los gobiernos han venido adoptando diferentes respuestas en sus políticas: desde la prohibición de la IAGen hasta la evaluación de necesidades de adaptación de los marcos existentes para formular urgentemente nuevas regulaciones.

En abril de 2023 se mapearon y revisaron las estrategias gubernamentales para regular y facilitar el uso creativo de la IAGen (UNESCO, 2023b).⁴⁷ La revisión sugiere una lista de siete pasos que los organismos gubernamentales pueden dar para regular la IA generativa y reafirmar el control público a efectos de aprovechar su potencial en todos los sectores, incluido el educativo.

Paso 1: Adscribirse a regulaciones generales de protección de datos internacionales o regionales, o desarrollar normativas nacionales

El entrenamiento de los modelos de IAGen ha implicado la recolección y el tratamiento de datos en línea de ciudadanos de muchos países. El uso de datos y contenidos sin consentimiento por parte de los modelos de IAGen está poniendo aún más en peligro la cuestión de la protección de datos.

La normativa general de protección de datos, de la cual el GDPR de la UE promulgado en 2018 es uno de los ejemplos precursores, aporta el marco legal necesario para regular la recolección y el tratamiento de datos personales por parte de los desarrolladores de IAGen. Según el portal *Data Protection and Privacy Legislation Worldwide* de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD), 137 de 194 países han establecido una legislación para salvaguardar la protección de datos y la privacidad.⁴⁸

No obstante, sigue sin quedar claro hasta qué punto estos marcos están siendo aplicados en dichos países. Por ello, es cada vez más importante garantizar su correcta aplicación, incluido el monitoreo regular de las operaciones de los sistemas de IAGen. También es urgente que los países que aún no cuentan con leyes generales de protección de datos las desarrollen.

Paso 2: Adoptar/revisar y financiar estrategias sobre IA para todo el sector gubernamental

La regulación de la IA generativa debe formar parte de estrategias nacionales de IA más amplias que aseguren un uso seguro y equitativo de la IA en todos los sectores del desarrollo, incluida la educación. La formulación, adhesión, financiación y aplicación de las estrategias nacionales de IA requieren un enfoque que involucre a todo el gobierno. Solo un planteamiento de estas características puede garantizar la coordinación de acciones intersectoriales necesaria para responder de forma integrada a los nuevos desafíos.

A principios de 2023, unos 67 países⁴⁹ habían desarrollado o planificado estrategias nacionales sobre IA, de las cuales 61 adoptaban la forma de estrategia de IA independiente y 7 eran capítulos acerca de IA integrados en estrategias nacionales más amplias sobre TIC o digitalización. Es comprensible que, dado su carácter novedoso, al momento de redactarse este informe, ninguna de estas estrategias nacionales hubiera abordado aún la IA generativa como cuestión específica.

Es fundamental que los países revisen las estrategias nacionales de IA existentes, o las desarrollen, asegurando disposiciones que regulen el uso ético de la IA en todos los sectores, incluida la educación.

Paso 3: Consolidar e implementar normativas específicas sobre la ética de la IA

Para abordar las dimensiones éticas que plantea el uso de la IA, se requiere una normativa específica.

La revisión de estrategias nacionales de IA realizada por la UNESCO en 2023 indica que la identificación de estas cuestiones éticas y la formulación de principios orientadores solo es común a unas 40 estrategias nacionales de IA.⁵⁰ E, incluso en estos casos, los principios éticos deberán traducirse en leyes o reglamentos de cumplimiento obligatorio. Rara vez ocurre algo así. De hecho, solo unos 20 países han definido normativas claras sobre la ética de la IA, incluyendo su relación con la educación, sea como parte de las estrategias nacionales de IA o de otro modo. Resulta interesante que, aunque la educación se destaca como un dominio político en unas 45 estrategias nacionales de IA,⁵¹ las referencias a la educación se articulan más en términos de habilidades de IA y el desarrollo de talentos necesarios para respaldar la competitividad nacional, y menos en términos de cuestiones éticas.

Los países que aún no cuentan con una normativa sobre la ética de la IA deben articularla y aplicarla urgentemente.

Paso 4: Adaptar o hacer cumplir las leyes de derechos de autor existentes para regular los contenidos generados por IA

El uso cada vez más generalizado de la IAGen ha presentado nuevos desafíos para los derechos de autor, tanto en lo referente a los contenidos protegidos por éstos o al trabajo sobre el que se entrenan los modelos, como al estatus de los resultados de conocimiento “no humano” que producen.

En la actualidad, solo China, los países de la UE y Estados Unidos han adaptado la legislación sobre derechos de autor para que contemple las implicancias de la IA generativa. La Oficina del Derecho de Autor de EEUU, por ejemplo, dictaminó que los resultados de los sistemas de IAGen, como ChatGPT, no están protegidos por la ley de derechos de autor estadounidense, argumentando que “los derechos de autor solo pueden proteger material que sea producto de la creatividad humana” (Oficina del Derecho de Autor de EEUU, 2023).

En la Unión Europea, la propuesta de Ley de IA exige a los desarrolladores de herramientas de IA que revelen los materiales protegidos por derechos de autor utilizados en la creación de sus sistemas (Comisión Europea, 2021). China, a través de su reglamento sobre IAGen publicado en julio de 2023, exige el etiquetado de los resultados de IAGen como contenido generado por IA, y solo los reconoce como resultados de síntesis digital.

Regular el uso de materiales protegidos por derechos de autor en el entrenamiento de modelos de IAGen y definir el estatus de los derechos de autor de los resultados de la IAGen emergen como nuevas atribuciones de este tipo de normas. Con urgencia, las leyes existentes deberán adaptarse para dar cuenta de esto.

Paso 5: Elaborar marcos regulatorios sobre la IA generativa

El rápido ritmo de desarrollo de las tecnologías de IA está forzando a los organismos de gobernanza nacionales/locales a acelerar la actualización de sus legislaciones. Hasta julio de 2023, solo un país, China, había promulgado regulaciones oficiales específicas sobre IAGen. El Reglamento Provisional sobre Regulación del Servicio de IA Generativa publicado el 13 de julio de 2023 (Administración del Ciberespacio de China, 2023a) exige a los proveedores de sistemas de IAGen que etiqueten los contenidos, imágenes y vídeos generados por IA de forma adecuada y legal, de conformidad con su actual Reglamento sobre Síntesis Profunda en el Marco de los Servicios de Información en Línea. Es preciso desarrollar más marcos nacionales específicos de la IAGen basados en una evaluación de las brechas existentes en las regulaciones y leyes locales.

Paso 6: Construir capacidades para el uso adecuado de la IAGen en educación e investigación

Las escuelas y demás instituciones educativas necesitan desarrollar capacidades para entender los beneficios y riesgos potenciales de la IA, incluyendo la IAGen, para la educación. Solo sobre esta base podrán validar la adopción de herramientas de IA. Asimismo, es necesario apoyar a los docentes e investigadores en el fortalecimiento de sus capacidades para el uso adecuado de la IAGen, incluyendo su formación y el entrenamiento continuo. Varios países han puesto en marcha este tipo de programas de desarrollo de capacidades, entre ellos Singapur, que ofrece una plataforma específica para el desarrollo de capacidades

de IA en las instituciones educativas a través de su clúster gubernamental de IA en la nube, que incluye un repositorio específico de modelos GPT (Ocampo, 2023).

Paso 7: Reflexionar sobre las implicancias a largo plazo de la IAGen para la educación y la investigación

El impacto de las actuales versiones de IAGen apenas comienza a manifestarse, y sus efectos en la educación aún no se han explorado ni comprendido plenamente. Entretanto, continúan desarrollándose e implementándose versiones cada más potentes de IAGen y otras clases de IA. Sin embargo, siguen planteándose cuestiones cruciales sobre las implicancias de la IAGen para la creación, transmisión y validación del conocimiento, para la enseñanza y el aprendizaje, para el diseño y la evaluación curricular y para la investigación y los derechos de autor. La mayoría de los países se encuentran en la fase inicial de adopción de la IAGen en la educación, aunque aún se desconocen sus repercusiones a largo plazo. Para garantizar un uso de la IA centrado en el ser humano, urge establecer un debate público abierto y diálogos de política acerca sus implicancias de largo aliento. Un debate inclusivo en el que participen los gobiernos, el sector privado y otros actores debería servir para recabar ideas y aportaciones para la actualización progresiva de normativas y políticas.

3.3 Regulaciones sobre la IAGen: Elementos clave

Todos los países necesitan regular adecuadamente la IAGen para garantizar que promueva el desarrollo en la educación y en otros contextos. Esta sección propone acciones en torno a elementos clave que pueden ser adoptadas por: (1) agencias reguladoras, (2) desarrolladores de herramientas de IA, (3) usuarios institucionales y (4) usuarios individuales. Aunque muchos de los elementos del marco son de carácter transnacional, todos deben considerarse también a la luz del contexto local, es decir, de los sistemas educativos y los marcos normativos generales vigentes en cada país.

3.3.1 Organismos reguladores gubernamentales

Para coordinar el diseño, el alineamiento y la implementación de regulaciones sobre IAGen se requiere un enfoque integral de política pública. Para ello, se recomiendan los siguientes siete elementos y acciones clave:

- **Coordinación intersectorial:** Establecer un organismo nacional que lidere el abordaje interministerial de la IAGen y coordine la cooperación intersectorial.
- **Ajuste de la legislación:** Alinear el marco a los contextos legislativos y reglamentarios relevantes de cada país con, por ejemplo, las leyes generales de protección de datos, las normativas sobre seguridad en Internet, las leyes sobre seguridad de datos producidos a partir de ciudadanos o utilizados para brindarles servicios, y otras legislaciones y prácticas habituales relevantes. Evaluar la adecuación de la normativa vigente y las adaptaciones necesarias en respuesta a las nuevas cuestiones planteadas por la IAGen.
- **Equilibrio entre la regulación de la IAGen y la promoción de la innovación en IA:** Promover la cooperación intersectorial entre compañías, organizaciones e instituciones educativas y de investigación, así como entre agencias públicas relevantes para desarrollar de forma conjunta modelos confiables; impulsar la construcción de ecosistemas de código abierto para promover el intercambio de recursos de supercomputación y conjuntos de datos de preentrenamiento de alta calidad; y promover la aplicación práctica de la IAGen en todos los sectores, así como la creación de contenidos de alta calidad para el bienestar público.
- **Evaluación y clasificación de los riesgos potenciales de la IA:** Establecer principios y un proceso para la evaluación y categorización de la eficacia, seguridad y protección de los servicios de IAGen, antes de su implementación y a lo largo del ciclo de vida del sistema. Considerar mecanismos de categorización basados en los niveles de riesgo que la IAGen pueda implicar para los ciudadanos. Clasificarlos en regulaciones estrictas (es decir, prohibir aplicaciones o sistemas habilitados para IA con riesgos inaceptables), regulaciones especiales para aplicaciones de alto riesgo y regulaciones generales sobre aplicaciones no catalogadas como de alto riesgo. Ver el proyecto de Ley de IA de la UE como ejemplo de este enfoque.
- **Protección de la privacidad de los datos:** Considerar que el uso de la IAGen casi siempre implica que los usuarios compartan sus datos con el proveedor de la IAGen. Ordenar la

elaboración e implementación de leyes para la protección de la información personal de los usuarios e identificar y combatir las prácticas ilegales en el almacenamiento, la elaboración de perfiles y el intercambio de datos.

- **Definición y cumplimiento del límite de edad para el uso de la IAGen:** La mayoría de las aplicaciones de IAGen están diseñadas principalmente para usuarios adultos. Estas aplicaciones con frecuencia implican riesgos sustanciales para los niños y niñas, incluyendo la exposición a contenidos inapropiados, así como el potencial de manipulación. A la luz de estos riesgos y dada la considerable incertidumbre que sigue rodeando a las aplicaciones de IAGen iterativas, se recomiendan enfáticamente las restricciones por edad para las tecnologías de IA de propósito general, con el fin de proteger los derechos y el bienestar de los niños y niñas.

Actualmente, los términos de uso de ChatGPT exigen que los usuarios tengan al menos 13 años de edad, y los menores de 18 deben contar con el permiso de sus padres o tutores para utilizar el servicio.⁵² Estas restricciones o límites de edad derivan de la Ley de Protección de la Privacidad Infantil en Internet de Estados Unidos (Comisión Federal de Comercio, 1998). Aprobada en 1998, antes de que se generalizara el uso de las redes sociales y mucho antes de la creación de aplicaciones de IAGen fáciles de usar y potentes como ChatGPT, la ley estadounidense especifica que las organizaciones o los proveedores individuales de redes sociales no pueden ofrecer servicios a menores de 13 años sin permiso parental. Muchos analistas consideran que este límite es demasiado bajo y han abogado por una legislación que eleve la edad a 16 años. El Reglamento General de Protección de Datos (GDPR) de la Unión Europea (2016) especifica que los usuarios deben tener al menos 16 años para utilizar servicios de redes sociales sin permiso parental.

El surgimiento de varios *chatbots* de IAGen exige que los países consideren cuidadosamente -y deliberen públicamente- sobre la edad mínima apropiada para mantener conversaciones independientes con las plataformas de IAGen. La edad mínima debería ser de 13 años. Los países también necesitan decidir si la autodeclaración de edad sigue siendo un medio apropiado de

verificación. Los países tendrán que determinar la responsabilidad de los proveedores de IAGen en la verificación etaria y la de los padres o tutores en el monitoreo de las conversaciones independientes de los niños y niñas menores de edad.

- **Propiedad nacional de los datos y riesgo de pobreza de datos:** Adoptar medidas legislativas para proteger la propiedad nacional de los datos y regular a los proveedores de IAGen que operan dentro de sus fronteras. Para los conjuntos de datos generados por los ciudadanos utilizados con fines comerciales, establecer normativas que promuevan una cooperación de beneficio mutuo de modo que esta categoría de datos no abandone el país para ser explotada exclusivamente por las grandes empresas tecnológicas.

3.3.2 Proveedores de herramientas de IAGen

Entre los proveedores de IAGen se incluyen organizaciones e individuos que desarrollan y ponen a disposición herramientas de IAGen y/o utilizan tecnologías de IAGen para brindar servicios, incluso a través de interfaces de programación de aplicaciones (IPA) programables. La mayoría de los proveedores influyentes de herramientas de IAGen son empresas muy bien financiadas. Debe quedar claro a los proveedores de la IAGen que son éticamente responsables desde el diseño, incluida la implementación de los principios éticos estipulados en las regulaciones. Deberían cubrirse las siguientes diez categorías de responsabilidades:

- **Responsabilidades humanas:** Los proveedores de IAGen deberían ser responsables de garantizar la adhesión a los valores fundamentales y a los fines legales, respetar la propiedad intelectual y defender las prácticas éticas, evitando a su vez la propagación de la desinformación y los discursos de odio.
- **Datos y modelos confiables:** Debería exigirse a los proveedores de IAGen que demuestren la confiabilidad y la ética de las fuentes de datos, así como de los métodos usados por sus modelos y resultados. Se los debería obligar a adoptar modelos de datos y fundamentación con fuentes legales probadas, y a respetar las leyes de propiedad intelectual correspondientes (por ejemplo, si los datos están protegidos por derechos de propiedad intelectual). Además, cuando los modelos necesiten utilizar información personal, la recolección solo debe

practicarse con el consentimiento informado y explícito de los titulares.

- **Generación de contenidos no discriminatorios:** Los proveedores de IAGen deben prohibir el diseño e implementación de sistemas de IAGen que generen contenidos sesgados o discriminatorios por motivos de raza, nacionalidad, género u otras características protegidas. Deberían garantizar la existencia de “barandillas” sólidas para evitar que la IAGen genere contenidos ofensivos, tendenciosos o falsos, asegurando a la vez la protección y la no explotación de los seres humanos implicados en la información resguardada por las barandillas.
- **Explicabilidad y transparencia de los modelos de IAGen:** Los proveedores deberían presentar a los organismos públicos de gobernanza sus explicaciones sobre las fuentes, la escala y los tipos de datos utilizados por los modelos; sus reglas para etiquetar los datos en el preentrenamiento; los métodos o algoritmos utilizados por sus modelos para generar contenidos o respuestas; y los servicios que ofrecen sus herramientas de IAGen. Siempre que fuera necesario, deberían brindar apoyo para ayudar a que los organismos gubernamentales comprendan la tecnología y los datos. La propensión de la IAGen a generar contenidos con errores y respuestas cuestionables debería ser transparente para los usuarios.
- **Etiquetado de los contenidos de IAGen:** De acuerdo con las leyes o regulaciones relevantes sobre la síntesis de información en línea asistida por IA, los proveedores deben etiquetar los documentos, informes, imágenes y videos generados por IAGen de forma apropiada y legal. Por ejemplo, los resultados de la IAGen deberían estar claramente etiquetados como producidos por una máquina.
- **Principios de seguridad y protección:** Los proveedores de IAGen deberían garantizar un servicio seguro, robusto y sostenible durante todo el ciclo de vida de un sistema de IAGen.
- **Especificaciones sobre la conveniencia de acceso y uso:** Los proveedores de la IAGen deberían ofrecer especificaciones claras sobre la audiencia apropiada, los escenarios de uso y los propósitos de sus servicios, y ayudar a los usuarios de herramientas de IAGen a tomar decisiones racionales y responsables.

- **Reconocer las limitaciones y prevenir los riesgos predecibles:** Los proveedores de la IAGen deberían anunciar claramente las limitaciones de los métodos usados por los sistemas y sus resultados. Necesitan desarrollar tecnologías para asegurar que los datos de entrada, los métodos y los resultados no causen daños predecibles a los usuarios, junto con protocolos para mitigar los daños no predecibles si estos ocurrieran. También deben ofrecer orientación para ayudar a los usuarios a comprender los contenidos generados por la IAGen basándose en principios éticos, y para evitar la dependencia excesiva y la adicción a los contenidos generados.
- **Mecanismos de reclamación y reparación:** Los proveedores de la IAGen deben establecer mecanismos y canales para la recepción de quejas de los usuarios y del público en general, y adoptar las medidas oportunas para aceptar y tramitar estas reclamaciones.
- **Supervisión y notificación del uso ilegal:** Los proveedores deberán cooperar con los organismos públicos de gobernanza para facilitar el seguimiento y la denuncia de usos ilícitos. Esto incluye cuando las personas utilizan los productos IAGen de forma ilegal o violando valores éticos o sociales, como la promoción de la desinformación o la incitación al odio, la generación de spam o el desarrollo de *malware*.

3.3.3 Usuarios institucionales

Entre los usuarios institucionales están incluidas las autoridades e instituciones educativas, como universidades y escuelas, que tienen la responsabilidad de determinar si la IAGen debe ser adoptada y cuáles son los tipos de herramientas de IAGen que deben contratarse e implementarse en la institución.

- **Auditoría institucional de los algoritmos, datos y resultados de IAGen:** Implementar mecanismos para monitorear de la mejor manera posible los algoritmos y datos utilizados por las herramientas de IAGen y los resultados que generan. Esto debería incluir auditorías y evaluaciones periódicas, la protección de datos de los usuarios y el filtrado automático de contenidos inapropiados.
- **Validar la proporcionalidad y proteger el bienestar de los usuarios:** Implementar mecanismos nacionales de clasificación

o elaborar una política institucional para categorizar y validar los sistemas y aplicaciones de IAGen. Garantizar que los sistemas de IAGen adoptados por la institución estén alineados con los marcos éticos validados localmente y no causen daños predecibles a las poblaciones objetivo de las instituciones, especialmente niños, niñas y grupos vulnerables.

- **Revisar y abordar los impactos a largo plazo:** Con el tiempo, la utilización de herramientas o contenidos de IAGen en la educación puede tener efectos profundos en el desarrollo de capacidades humanas como el pensamiento crítico y la creatividad. Estos efectos potenciales deberían evaluarse y abordarse.
- **Adecuación a la edad:** Considerar la implementación de restricciones de edad mínima para el uso independiente de IAGen en la institución.

3.3.4 Usuarios individuales

Entre los usuarios individuales se incluye potencialmente a todas las personas del mundo que tengan acceso a Internet y al menos a un tipo de herramienta de IAGen. El término “usuarios individuales”, tal y como se emplea aquí, hace referencia a docentes individuales, investigadores y estudiantes de instituciones educativas formales, o a aquellos que participan en programas de estudio no formales.

- **Conocimiento de los términos de referencia sobre el uso de la IAGen:** Al firmar o expresar su consentimiento a los acuerdos de servicio, los usuarios deberían ser conscientes de las obligaciones de acatar los términos de referencia establecidos en el acuerdo y las leyes o normativas que lo respaldan.
- **Uso ético de las aplicaciones de IAGen:** Los usuarios deben implementar la IAGen de manera responsable y evitar explotarla de forma que pueda dañar la reputación y los derechos legítimos de otras personas.
- **Supervisión y notificación de aplicaciones de IAGen ilegales:** Cuando se descubran aplicaciones de IAGen que infrinjan una o varias normativas, los usuarios deberían notificarlo a los organismos gubernamentales reguladores.

4. Hacia un marco de políticas para el uso de la IA generativa en la educación y la investigación

Regular la IAGen para aprovechar sus beneficios potenciales para la educación y la investigación requiere el desarrollo de políticas adecuadas. Los datos de la encuesta de 2023 citados anteriormente indican que solo un puñado de países ha adoptado políticas o planes específicos para el uso de la IA en la educación. La sección anterior esbozaba una visión, los pasos necesarios y los elementos y acciones claves que pueden adoptar las diferentes partes interesadas. Esta sección ofrece medidas que pueden aplicarse para desarrollar marcos políticos coherentes y abarcadores para reglamentar el uso de la IAGen en la educación y la investigación.

Un punto de partida para ello es el documento de 2022 *Inteligencia artificial y educación: guía para las personas a cargo de formular políticas* (UNESCO, 2022b). Propone un conjunto amplio de recomendaciones para orientar a los gobiernos en el desarrollo e implementación de políticas sectoriales sobre IA y educación con foco en promover la educación de calidad, la equidad social y la inclusión. La mayoría de las recomendaciones siguen siendo aplicables y pueden adaptarse para orientar la formulación de políticas específicas de IAGen en la educación. Las siguientes ocho medidas específicas para la planificación de políticas sobre IAGen en la educación y la investigación se presentan aquí como complemento de la guía existente.

4.1 Promover la inclusión, la equidad y la diversidad lingüística y cultural

La importancia crítica de la inclusión debe ser reconocida y abordada a lo largo del ciclo de vida de la IAGen. Más específicamente, las herramientas de IAGen no ayudarán a dar respuesta a los desafíos fundamentales en la educación o a la consecución de los compromisos del ODS 4 a menos que estas herramientas sean accesibles de manera inclusiva (independientemente del género, origen étnico, necesidades educativas especiales, estatus socioeconómico, ubicación geográfica, situación de desplazamiento, etc.), y si no promueven por diseño la equidad, las diversidades lingüísticas y el pluralismo cultural. Para lograrlo, se recomiendan las tres siguientes medidas de política:

- Identificar a quienes no tienen o no pueden costearse la conexión a Internet o el acceso a los datos, y tomar medidas para promover la conectividad universal y las competencias digitales a fin de reducir las barreras al acceso equitativo e inclusivo a las aplicaciones de IA. Establecer mecanismos de financiamiento sostenibles para el desarrollo y la oferta de herramientas de IA a los estudiantes con discapacidad o necesidades especiales. Promover el uso de la IAGen para apoyar el aprendizaje a lo largo de toda la vida en estudiantes de todas las edades, ubicaciones y orígenes.
- Desarrollar criterios de validación de los sistemas de IAGen para garantizar que no haya sesgos de género, discriminación contra grupos marginalizados o discursos de odio embebidos en datos o algoritmos.
- Desarrollar e implementar especificaciones inclusivas para los sistemas IAGen y establecer medidas institucionales para proteger las diversidades lingüísticas y culturales cuando se implemente la IAGen en la educación y la investigación a gran escala. Las especificaciones relevantes deben exigir a los proveedores de IAGen que incluyan datos en múltiples lenguas, particularmente en lenguas locales o aborígenes, en el entrenamiento de los modelos GPT para mejorar la capacidad de la IAGen de responder y generar textos multilingües. Las especificaciones y las medidas institucionales deberían impedir estrictamente a los proveedores de IA cualquier eliminación intencionada o no de las lenguas minoritarias o la discriminación de los hablantes de lenguas aborígenes, y exigir a los proveedores que impidan que los sistemas promuevan solo las lenguas o las normas culturales dominantes.

4.2 Proteger la acción humana

A medida que la IAGen se vuelve cada vez más sofisticada, un peligro clave es su potencial para socavar la acción humana. Conforme más personas utilicen la IAGen para escribir o para otras actividades creativas, es posible que involuntariamente acaben dependiendo de ella. Esto puede comprometer el desarrollo de habilidades intelectuales. Aunque la IAGen puede utilizarse para desafiar y expandir el pensamiento humano, no debería permitirse que lo sustituyera. La protección y la mejora de la intervención humana deberían ser siempre consideraciones esenciales al diseñar y adoptar la IAGen a partir de las siguientes siete perspectivas:

- Informar a los estudiantes sobre los tipos de datos que la IAGen puede recolectar de ellos, sobre cómo se utilizan estos datos y el impacto que pueden tener en su educación y en sus vidas en general.
- Proteger la motivación intrínseca de los estudiantes para crecer y aprender como individuos. Reforzar la autonomía humana respecto de sus propios enfoques de investigación, enseñanza y aprendizaje en el contexto del uso de sistemas de IAGen cada vez más sofisticados.
- Evitar el uso de la IAGen cuando pudiera privar a los estudiantes de oportunidades para desarrollar capacidades cognitivas y habilidades sociales a través de observaciones del mundo real, prácticas empíricas como experimentos, discusiones con otros humanos y razonamiento lógico independiente.
- Asegurar suficiente interacción social y una adecuada exposición a la producción creativa humana para evitar que los estudiantes se vuelvan adictos o dependientes de la IAGen.
- Utilizar las herramientas de IAGen para minimizar la presión de las tareas y los exámenes, en vez de para exacerbarla.
- Consultar a investigadores, docentes y estudiantes su opinión acerca de la IAGen y utilizar esa retroalimentación para decidir si determinadas herramientas de IAGen deberían implantarse a nivel institucional, y de qué manera. Estimular a estudiantes, docentes e investigadores a criticar y cuestionar las metodologías subyacentes a los sistemas de

IA, la precisión de los resultados y las normas o pedagogías que estos podrían imponer.

- Evitar ceder la responsabilidad humana a los sistemas de IAGen cuando se tomen decisiones importantes.

4.3 Monitorear y validar los sistemas de IAGen para la educación

Como se mencionó antes, el desarrollo e implementación de la IAGen debe ser ético por diseño. En consecuencia, a continuación, una vez que la IAGen está en uso, y durante todo su ciclo de vida, debe ser cuidadosamente monitoreada y validada en cuanto a sus riesgos éticos, su adecuación y rigor pedagógicos, así como su impacto en los estudiantes, los docentes y en las relaciones en el aula en la escuela. A este respecto, se recomiendan las siguientes cinco acciones:

- Construir mecanismos de validación para testear si los sistemas de IAGen utilizados en educación e investigación están libres de sesgos, especialmente sesgos de género, y si fueron entrenados con datos representativos de la diversidad (en términos de género, discapacidad, estatus social y económico, origen étnico y cultural, y localización geográfica).
- Abordar la compleja cuestión del consentimiento informado, especialmente en contextos en los que niños, niñas u otros estudiantes vulnerables no sean capaces de dar un consentimiento genuinamente informado.
- Auditar si los resultados de la IAGen incluyen imágenes *deepfake*, noticias falsas (imprecisas o falaces) o discursos de odio. Si se descubre que la IAGen genera contenidos inapropiados, las instituciones y los docentes deberían estar dispuestos y ser capaces de tomar medidas rápidas y enérgicas para paliar o eliminar el problema.
- Practicar una estricta validación ética de las aplicaciones de IAGen antes de que sean adoptadas oficialmente en instituciones educativas o de investigación (es decir, adoptar un enfoque de ética por diseño).
- Antes de tomar decisiones sobre la adopción institucional, asegurarse de que las aplicaciones de IAGen en cuestión no causen daños predecibles a los estudiantes, que sean

educativamente efectivas y válidas para las edades y capacidades de los estudiantes a los que van dirigidas, y que se estén alineadas con principios pedagógicos sólidos (es decir, que estén basadas en los dominios de conocimiento relevantes y en los resultados de aprendizaje y desarrollo de valores esperados).

4.4 Desarrollar competencias en IA, incluyendo habilidades de IAGen para los estudiantes

El desarrollo de competencias de IA entre estudiantes es clave para el uso seguro, ético y significativo de esta en la educación y más allá. Sin embargo, según datos de la UNESCO, a principios de 2022 solo unos 15 países habían desarrollado e implementado, o estaban en proceso de desarrollar, currículos de IA validados por los gobiernos en las escuelas (UNESCO, 2022c). Los últimos avances de IAGen han fortalecido aún más la necesidad imperiosa de que todos alcancen un nivel adecuado de alfabetización tanto en la dimensión humana como tecnológica de la IA, comprendiendo cómo funciona en términos generales, así como el impacto específico de la IAGen. Para ello, urgen ahora las siguientes cinco acciones:

- Comprometerse a impartir currículos de IA aprobados por los gobiernos para la enseñanza escolar, en la educación y formación técnica y profesional, así como para el aprendizaje a lo largo de toda la vida. Los currículos de IA deben abordar el impacto de la IA en nuestras vidas, incluidas las cuestiones éticas que plantea, así como una comprensión de los algoritmos y los datos apropiada para cada edad, y las habilidades para el uso adecuado y creativo de las herramientas de IA, incluidas las aplicaciones de la IAGen;
- Apoyar a las instituciones de educación superior e investigación para la mejora de los programas orientados al desarrollo de talento local en IA;
- Promover la equidad de género en el desarrollo de competencias avanzadas en IA y crear una cantera de profesionales con equilibrio de género;
- Desarrollar pronósticos intersectoriales de los cambios laborales nacionales y mundiales provocados por la reciente automatización de la IAGen, y mejorar las habilidades para el

futuro en todos los niveles educativos y en los sistemas de aprendizaje a lo largo de toda la vida sobre la base de los posibles cambios en la demanda; y

- Ofrecer programas especiales para trabajadores y ciudadanos mayores que puedan necesitar aprender nuevas habilidades y adaptarse a nuevos entornos.

4.5 Desarrollar las capacidades de docentes e investigadores para hacer un uso adecuado de IAGen

Según los datos de la encuesta de 2023 sobre el uso gubernamental de la IA para la educación (UNESCO, 2023c), apenas siete países (China, Finlandia, Georgia, Qatar, España, Tailandia y Turquía) informaron haber desarrollado o estar desarrollando marcos o programas de capacitación en IA para docentes. Solo el Ministerio de Educación de Singapur informó de la creación de un repositorio en línea centrado en el uso de ChatGPT en la enseñanza y el aprendizaje. Esto demuestra claramente que los docentes de la mayoría de los países no tienen acceso a una capacitación bien estructurada sobre el uso de la IA en la educación, menos aún sobre el uso de la IAGen.

Para preparar a los docentes para un uso responsable y eficaz de la IAGen, los países necesitan adoptar las siguientes cuatro medidas:

- Formular o ajustar orientaciones basadas en experiencias locales para ayudar a investigadores y docentes a explorar las herramientas de IAGen ampliamente difundidas, y guiar el diseño de nuevas aplicaciones de IA específicas para cada dominio.
- Proteger los derechos de docentes e investigadores y el valor de sus prácticas al utilizar IAGen. Más concretamente, analizar el rol único de los docentes para facilitar el pensamiento de orden superior, organizar la interacción entre las personas y promover los valores humanos.
- Definir la orientación hacia los valores, conocimientos y habilidades que los docentes precisan para entender y usar los sistemas de IAGen de forma eficaz y ética. Permitir que los docentes creen herramientas específicas basadas en la IAGen para facilitar el aprendizaje en el aula y en su propio desarrollo profesional.

- Revisar dinámicamente las competencias que necesitan los docentes para entender y usar la IA para la enseñanza, el aprendizaje y para su desarrollo profesional, e integrar conjuntos emergentes de valores, comprensión y habilidades de IA a los marcos de competencias y programas de capacitación de docentes en servicio, así como durante su formación inicial.

4.6 Promover opiniones diversas y la expresión plural de ideas

Como ya se ha mencionado, la IAGen no entiende ni la pregunta ni la respuesta. En lugar de ello, sus respuestas están basadas en probabilidades de patrones lingüísticos encontrados en los datos (de Internet) que se ingresaron al entrenarse el modelo. Para resolver algunos de los problemas fundamentales de sus resultados, nuevos métodos están siendo investigados actualmente, como la conexión de la IAGen con bases de conocimientos y motores de razonamiento. No obstante, debido a su funcionamiento, a sus fuentes de origen y a las perspectivas tácitas de sus desarrolladores, la IAGen, por definición, reproduce en sus resultados las visiones del mundo dominantes y socava las opiniones minoritarias y plurales. En consecuencia, si queremos que las civilizaciones humanas florezcan, es esencial que reconozcamos que la IAGen nunca podrá ser una fuente válida de conocimiento sobre cualquier tema que aborde.

En consecuencia, los usuarios deben considerar los resultados de IAGen de forma crítica. En particular:

- Comprender el papel de la IAGen como fuente de información rápida, pero con frecuencia poco confiable. Aunque algunos *plugins* y herramientas basadas en LLM mencionados anteriormente están diseñados para dar respuesta a la necesidad de acceder a información validada y actualizada, aún hay escasa la evidencia robusta de que sean efectivas.
- Alentar a los estudiantes e investigadores a criticar las respuestas ofrecidas por la IAGen. Reconocer que la IAGen normalmente solo repite opiniones establecidas o estándar, socavando así las opiniones minoritarias y la expresión de ideas plurales.
- Brindar a los estudiantes oportunidades suficientes para aprender por ensayo y error, experimentos empíricos y observaciones del mundo real.

4.7 Testear localmente modelos de aplicaciones relevantes y crear una base acumulativa de evidencias

Hasta ahora, los modelos de IAGen están dominados por la información procedente del Norte Global e subrepresentan las voces del Sur Global y de las comunidades aborígenes. Solo mediante esfuerzos decididos, tales como aprovechar los datos sintéticos (Marwala, 2023), se conseguirá que las herramientas de la IAGen sean sensibles al contexto y las necesidades de las comunidades locales, en particular las del Sur Global. A efectos de explorar enfoques relevantes para las necesidades locales y colaborar más ampliamente, se recomiendan las siguientes ocho acciones:

- Garantizar que el diseño y la adopción de la IAGen se planifiquen estratégicamente en lugar de facilitar un proceso de implementación pasivo y acrítico.
- Incentivar a los diseñadores de la IAGen a que apuesten por opciones de aprendizaje abiertas, exploratorias y diversas.
- Testear y escalar casos de uso basados en evidencias de aplicación de la IA en la educación y la investigación de acuerdo con las prioridades educativas, en lugar de la novedad, el mito o las expectativas exageradas.
- Orientar el uso de la IAGen para promover la innovación en la investigación, incluyendo el aprovechamiento de las capacidades computacionales, los datos a gran escala y los resultados de la IAGen para fundamentar e inspirar la mejora de las metodologías de investigación.
- Revisar las implicancias sociales y éticas de la incorporación de la IAGen a los procesos de investigación.
- Establecer criterios específicos y metodologías pedagógicas basados en investigaciones y construir una base de evidencias sobre la eficacia de la IAGen en términos de apoyo a la oferta de oportunidades de aprendizaje inclusivo, cumplimiento de los objetivos de aprendizaje e investigación y promoción de las diversidades lingüísticas y culturales.
- Adoptar medidas progresivas para reforzar la evidencia sobre el impacto social y ético de la IAGen.

- Analizar los costos medioambientales de aprovechar las tecnologías de IA a gran escala (por ejemplo, la energía y los recursos necesarios para entrenar a los modelos GPT) y desarrollar metas sostenibles que deban cumplir los proveedores de IA evitando contribuir al cambio climático.

4.8 Considerar las implicancias a largo plazo, de forma intersectorial e interdisciplinaria

Los enfoques intersectoriales e interdisciplinarios son esenciales para el uso efectivo y ético de la IAGen en la educación y la investigación. Solo cuando se apela a un amplio rango de conocimientos, y se reúne a las múltiples partes interesadas, se logra identificar rápidamente los desafíos clave y enfrentarlos eficazmente para minimizar las consecuencias negativas a largo plazo, al tiempo que se aprovechan los beneficios actuales y acumulativos. Por lo anterior, se recomiendan estas tres acciones:

- Colaborar con los proveedores de IA, docentes, investigadores y representantes de padres y estudiantes para planificar ajustes en todo el sistema de marcos curriculares y metodologías de evaluación, a fin de aprovechar plenamente el potencial y reducir los riesgos de la IAGen para la educación y la investigación.
- Reunir a expertos intersectoriales e interdisciplinarios, incluyendo a docentes, investigadores, científicos del aprendizaje, ingenieros de IA y representantes de otras partes interesadas para examinar las implicancias a largo plazo de la IAGen para el aprendizaje y la producción de conocimiento, la investigación y los derechos de autor, los currículos y la evaluación, la colaboración humana y las dinámicas sociales.
- Ofrecer asesoramiento oportuno con el fin de informar sobre las actualizaciones iterativas de regulaciones y políticas.

5. Facilitar el uso creativo de la IAGen en la educación y la investigación

Cuando se lanzó inicialmente ChatGPT, docentes de todo el mundo expresaron preocupación por su potencial para producir trabajos, lo que podría ayudar a los estudiantes a hacer trampas. Más recientemente, muchas personas y organizaciones, incluidas algunas de las principales universidades del mundo, han argumentado que “el genio ha salido de la botella” y que herramientas como ChatGPT llegaron para quedarse y pueden utilizarse de modo productivo en entornos educativos. Mientras tanto, Internet está plagado de sugerencias sobre usos de la IAGen en la educación y la investigación. Por ejemplo, para inspirar nuevas ideas, generar ejemplos con perspectivas múltiples, desarrollar planes de clases y presentaciones, resumir materiales existentes y estimular la creación de imágenes. Aunque casi todos los días aparecen nuevas ideas en Internet, los investigadores y docentes continúan estudiando qué significa exactamente la IAGen para la enseñanza, el aprendizaje y la investigación. Particularmente, las personas que están detrás de muchos de los usos propuestos pueden no haber considerado debidamente sus principios éticos, mientras que otros están motivados por las potencialidades tecnológicas de la IAGen más que por las necesidades de los investigadores, docentes o estudiantes. En esta sección se presentan maneras de facilitar el uso creativo de la IAGen en la educación.

5.1 Estrategias institucionales para facilitar el uso responsable y creativo de la IAGen

Como se indicó anteriormente, las instituciones educativas y de investigación deberían desarrollar, implementar y validar estrategias y marcos apropiados para orientar el uso responsable y ético de los sistemas y aplicaciones de IAGen a efectos de satisfacer las necesidades de la enseñanza, el aprendizaje y la investigación. Esto puede lograrse a través de las siguientes cuatro premisas:

- **Orientación y capacitación:** Brindar orientación y capacitación a investigadores, docentes y estudiantes sobre las herramientas de IAGen para garantizar que comprendan las cuestiones éticas, como los sesgos en el etiquetado de datos y en los algoritmos, y que cumplan con las regulaciones pertinentes sobre privacidad de datos y propiedad intelectual.
- **Construir capacidades de ingeniería de prompts de IAGen:** Además de los conocimientos específicos de cada asignatura, los investigadores y docentes precisarán también experiencia en la ingeniería y la evaluación crítica de los *prompts* generados por IAGen. Considerando que los desafíos planteados por la IAGen son complejos, los investigadores y docentes deben recibir capacitación y apoyo de alta calidad para ello.
- **Detectar plagio mediante IAGen en trabajos escritos:** La IAGen permitiría a los estudiantes hacer pasar por suyos textos que no han escrito, lo que constituye un nuevo tipo de “plagio”. Los proveedores de IAGen deben etiquetar sus productos con una marca de agua que indique “generado por IA”, mientras se desarrollan herramientas para identificar el material producido por la IA. Sin embargo, hay escasas evidencias de que estas medidas o herramientas sean eficaces. La estrategia institucional inmediata para mantener la integridad académica y reforzar la responsabilidad es la detección rigurosa por parte de seres humanos. La estrategia a largo plazo ha de ser que las instituciones y los docentes reconsideren el diseño de las tareas escritas para que no se utilicen en la evaluación de producciones que las herramientas de IAGen pueden hacer mejor que los estudiantes. En su lugar, deberían abordar aquello que los humanos pueden hacer y que la IAGen y otras herramientas de IA no pueden, incluyendo la aplicación de valores humanos como la compasión y la creatividad a desafíos complejos del mundo real.
- **Implementación institucional de principios éticos:** Garantizar que investigadores, docentes y estudiantes utilicen las herramientas de la IAGen de forma responsable y ética, y aborden de manera crítica la exactitud y validez de los resultados.

5.2 Un enfoque de “interacción centrada en el ser humano y pedagógicamente apropiada”

Los investigadores y docentes deberían priorizar la intervención humana y la interacción responsable y pedagógicamente apropiada entre los humanos y las herramientas de IA al momento de decidir si utilizar la IAGen y cómo hacerlo. Esto incluye las siguientes cinco consideraciones:

- el uso de la(s) herramienta(s) debería contribuir a satisfacer necesidades humanas y hacer que el aprendizaje o la investigación sean más efectivos que un enfoque sin tecnología u otro planteamiento alternativo;
- el uso de la(s) herramienta(s) por parte de docentes y estudiantes debería responder a su motivación intrínseca;
- el proceso de uso de la(s) herramienta(s) debería ser controlado por los educadores, los estudiantes o los investigadores humanos;
- la elección y organización de la(s) herramienta(s) y el contenido que generan deberían ser proporcionales, basándose en la franja de edad de los estudiantes, los resultados esperados y el tipo de conocimiento que se espera desarrollar (por ejemplo, factual, conceptual, procedimental o metacognitivo) o el problema objetivo (por ejemplo, bien estructurado o poco estructurado); y
- los procesos de uso deben garantizar el compromiso interactivo de los humanos con la IAGen y el pensamiento de orden superior, así como la responsabilidad humana en las decisiones relacionadas con la veracidad de los contenidos generados por IA, las estrategias de enseñanza o investigación y su impacto en los comportamientos humanos.

5.3 Codiseño del uso de la IAGen en la educación y la investigación

El uso de la IAGen en la educación y la investigación no debería imponerse de arriba hacia abajo ni dejarse llevar por la hipérbole comercial. Por el contrario, su uso seguro y eficaz debe ser codiseñado por docentes, estudiantes e investigadores. También se necesita un robusto proceso de ensayo y evaluación para

analizar la efectividad y el impacto a largo plazo de sus diferentes usos.

Para facilitar el codiseño recomendado, esta guía propone un marco compuesto por las siguientes seis perspectivas para consolidar interacciones pedagógicamente adecuadas y la priorización de la intervención humana:

- ámbitos de conocimiento o problemas apropiados;
- resultados esperados;
- herramientas de IAGen apropiadas y ventajas comparativas;
- requisitos para los usuarios;
- métodos pedagógicos humanos necesarios y ejemplos de prompts; y
- riesgos éticos.

Esta sección ofrece ejemplos del modo en que un proceso de codiseño en el uso de la IAGen puede informar las prácticas de investigación, ayudar en la enseñanza, brindar entrenamiento para la adquisición a ritmo propio de habilidades fundamentales, facilitar el pensamiento de orden superior y apoyar a los estudiantes con necesidades especiales. Estos ejemplos representan apenas la punta del iceberg del creciente número de ámbitos en los que la IAGen podría tener potencial.

5.3.1 IA generativa para la investigación

Los modelos de IAGen han demostrado su potencial para expandir los puntos de vista sobre los esquemas de la investigación y enriquecer la exploración de datos, así como las revisiones bibliográficas (ver la **Tabla 3**). Si bien es posible que surja un rango más amplio de casos de uso, se necesita investigación innovadora para definir el dominio potencial de los problemas de investigación y los resultados esperados, para demostrar eficacia y precisión, así como para asegurar que la intervención humana en la comprensión del mundo real a través de la investigación no se verá socavada por usar herramientas de IA.

Tabla 3. Codiseño de usos de la IAGen para la investigación

Usos potenciales pero no probados	Ámbitos de conocimiento apropiados o problemas	Resultados esperados	Herramientas adecuadas de IAGen y ventajas comparativas	Requisitos para los usuarios	Métodos pedagógicos humanos requeridos y ejemplos de prompts	Posibles riesgos
Asesor/a de IA para perfiles de investigación.	Puede ser útil en ámbitos bien estructurados de problemas de investigación.	Desarrollo y respuesta a preguntas de investigación, sugiriendo metodologías apropiadas. Transformación potencial: Entrenador 1:1 para la planificación de la investigación.	Partir de la lista del apartado 1.2 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar, además, las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	El investigador debe tener un conocimiento básico del (de los) tema(s). El investigador debería desarrollar la capacidad de verificar la información, y ser especialmente capaz de detectar citas de trabajos de investigación inexistentes.	Ideas básicas para definir problemas de investigación (como público objetivo, temas, contexto), así como metodologías, resultados esperados y formatos. Prompt de ejemplo: <i>Escriba 10 posibles preguntas de investigación para [tema x] y clasifíquelas según su importancia para [campo de investigación y]. Example prompt:</i>	Necesidad de estar alerta al alto riesgo de que la IAGen invente información (como publicaciones de investigación inexistentes) y a que los usuarios se vean tentados a copiar y pegar esquemas de investigación generados por IA, lo que puede reducir oportunidades para que investigadores júnior aprendan de la prueba y error.
Explorador generativo de datos y revisor bibliográfico.	Puede ser útil en ámbitos de problemas de investigación poco estructurados.	Recolección automática de información, exploración de un amplio rango de datos, propuesta de borradores de revisiones bibliográficas y automatización de partes de la interpretación de datos. Transformación potencial: Entrenadores de IA para exploración de datos y revisiones de literatura.	Partir de la lista del apartado 1.2 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Los investigadores deben tener sólidos conocimientos de metodologías y técnicas de análisis de datos.	Definiciones progresivas de los problemas, alcance de los datos y fuentes bibliográficas, metodologías usadas para la exploración de datos y revisiones de la literatura, y resultados esperados y sus formatos.	Debe desconfiarse de la información generada por IAGen, del tratamiento indebido de datos, de posibles violaciones de privacidad, de la elaboración no autorizada de esquemas y de sesgos de género. Se debe estar alerta de la propagación de normas dominantes y de su amenaza a normas alternativas y opiniones plurales.

5.3.2 IA generativa para facilitar la enseñanza

El uso de plataformas IAGen generales y de herramientas IAGen educativas específicas debería diseñarse para mejorar la comprensión de los docentes sobre su asignatura, así como sus conocimientos de metodologías de enseñanza, incluso mediante el codiseño por parte de los docentes y la IA de planes de clases, paquetes de cursos o currículos completos. Los asistentes conversacionales de docentes asistidos

por IAGen o “gemelos generativos de asistentes docentes”⁵³ que son preentrenados en base a datos de docentes experimentados y bibliotecas, han sido probados en algunas instituciones educativas y pueden tener un potencial desconocido, además de un riesgo ético inexplorado. Los procesos de aplicación práctica y las futuras iteraciones de estos modelos aún deben auditarse cuidadosamente a través del marco recomendado en esta guía y salvaguardarse mediante supervisión humana, como se ejemplifica en la **Tabla 4**.

Tabla 4. Codiseño de usos de la IAGen como apoyo a docentes y a la enseñanza

Usos potenciales pero no probados	Ámbitos de conocimiento apropiados o problemas	Resultados esperados	Herramientas adecuadas de IAGen y ventajas comparativas	Requisitos para los usuarios	Métodos pedagógicos humanos requeridos y ejemplos de prompts	Posibles riesgos
Codiseñador del currículo o del curso.	Conocimientos conceptuales sobre determinados temas de enseñanza y conocimientos procedimentales sobre metodologías de enseñanza.	Ayudar en el diseño del currículo y las clases, incluyendo perfilar o ampliar visiones sobre áreas clave del tema objetivo y definir la estructura curricular. También puede ayudar a los docentes a preparar pruebas y exámenes ofreciendo ejemplos de preguntas y criterios de evaluación. Transformación potencial: Currículos generados por IA	Partir de la lista del apartado 1.2 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Los docentes deben comprender y especificar cuidadosamente qué desean que abarquen y alcancen el currículo, los cursos, las lecciones o los exámenes, si desean abordar conocimientos procedimentales o conceptuales, y qué teoría pedagógica desean aplicar.	Preguntas a la IAGen sobre sugerencias de estructura y ejemplos de conocimiento factual sobre tema(s), sugerencia de métodos y procesos de enseñanza para temas o problemas, o creación de paquetes de cursos o planes de clases basados en temas y formatos. Los diseñadores de currículos humanos deben verificar los conocimientos fácticos y revisar la idoneidad de los paquetes de cursos propuestos.	El riesgo de que la IAGen imponga normas y métodos pedagógicos dominantes es alto. Puede perpetuar inadvertidamente prácticas excluyentes en favor de grupos ya ricos en datos y acentuar las desigualdades en el acceso a oportunidades educativas relevantes y de alta calidad, desfavoreciendo a grupos pobres en datos.
Chatbot generativo como asistente docente.	Conocimiento conceptual de múltiples ámbitos en problemas bien estructurados.	Brindar apoyo personalizado, responder preguntas e identificar recursos. Transformación potencial: Gemelos generativos de asistentes docentes.	Partir de la lista del apartado 1.2 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Apoya a los docentes, pero apunta directamente a los estudiantes, requiriendo que tengan suficientes conocimientos previos, capacidades y habilidades metacognitivas para verificar los resultados de la IAGen y detectar información errónea. Por ello, resultaría más apropiado para estudiantes de enseñanza superior.	Exige que los docentes comprendan claramente los problemas, monitoreen la conversación y ayuden a los estudiantes a verificar respuestas dudosas brindadas por la IAGen.	En base a las capacidades actuales de los modelos de IAGen, las instituciones educativas necesitan garantizar supervisión humana de las respuestas brindadas por las herramientas de IAGen, estando alertas al riesgo de desinformación. También puede limitar el acceso de los estudiantes a la guía y apoyo humanos, dificultando el desarrollo de una estrecha relación docente-estudiante, algo especialmente preocupante en el caso de niños y niñas.

5.3.3 IA generativa como entrenador 1:1 para la adquisición autodidacta de habilidades básicas

Aunque cada vez se presta más atención al pensamiento de orden superior y a la creatividad al definir los resultados del aprendizaje, no hay duda de la importancia de las habilidades fundamentales en el desarrollo psicológico de niños y niñas y en la progresión de sus competencias. Entre un amplio espectro de capacidades, estas habilidades

fundacionales incluyen escuchar, pronunciar y escribir en una lengua materna o extranjera, así como matemáticas básicas, arte y codificación. El “ejercicio y la práctica” no deberían considerarse un método pedagógico obsoleto, sino revigorizarse y actualizarse con las tecnologías de la IAGen para promover el ensayo autodidacta de las habilidades básicas de los estudiantes. Si se rigen por principios éticos y pedagógicos, las herramientas de IAGen tienen el potencial de convertirse en entrenadores 1:1 para dicha práctica a ritmo propio, como se ilustra en la **Tabla 5**.

Tabla 5. Codiseño de usos de la IAGen como entrenador 1:1 para la adquisición autodidacta de habilidades básicas en idiomas y en las artes

Usos potenciales pero no probados	Ámbitos de conocimiento apropiados o problemas	Resultados esperados	Herramientas adecuadas de IAGen y ventajas comparativas	Requisitos para los usuarios	Métodos pedagógicos humanos requeridos y ejemplos de prompts	Posibles riesgos
Entrenador lingüístico 1:1	Aprendizaje de idiomas, incluida la práctica de conversación.	Involucrar a los estudiantes en la práctica de conversación para ayudarles a mejorar las habilidades de escucha, habla y escritura ofreciéndoles devoluciones, correcciones y modelaciones de la lengua materna o extranjera. Ayudar a los estudiantes a mejorar sus habilidades de escritura. Transformación potencial: Tutoriales lingüísticos 1:1 a niveles de iniciación.	Partir de la lista del apartado 1.2 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Puede establecerse un límite de edad para conversaciones independientes, vistos los resultados culturalmente insensibles o etariamente inapropiados ofrecidos por los sistemas de IAGen. El estudiante debe tener motivación intrínseca inicial para entablar una conversación con un sistema de IA. El estudiante debería poder adoptar un enfoque crítico ante las sugerencias de la IAGen y comprobar si son precisas.	Al utilizar plataformas generales de IAGen, los docentes humanos pueden orientar a los estudiantes para que interactúen con las herramientas de IAGen y soliciten opiniones para mejorar, corregir la pronunciación o ejemplos de escritura. Por ejemplo: <i>Entabla una conversación conmigo en idioma [x], ayudándome a mejorar continuamente.</i> <i>Sugiere algunas ideas que me ayuden a escribir sobre [tema x].</i>	Se debe estar alerta ante un lenguaje culturalmente insensible o contextualmente incorrecto, y ante la perpetuación involuntaria de estereotipos o sesgos culturales. Sin estrategias pedagógicas adecuadas que simulen las motivaciones intrínsecas de los estudiantes, pueden verse limitadas la creatividad y la originalidad de los niños y niñas, dando lugar a una redacción formulista. También puede limitar oportunidades de interacciones en la vida real, opiniones y expresiones plurales y el pensamiento crítico.
Entrenador artístico 1:1	Habilidades técnicas en áreas artísticas como la música y el dibujo.	Ofrecer sugerencias para técnicas artísticas (como consejos sobre perspectiva y color), o composición musical (como melodía y progresión de acordes). Transformación potencial: Docente de arte 1:1 a nivel introductorio	Partir de la lista del apartado 1.2 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Los estudiantes deben tener metas iniciales para crear arte o música, una comprensión fundacional de los elementos clave del ámbito del arte o la música, y habilidades básicas para analizar las obras o las composiciones.	Los docentes humanos deberían pedir a los estudiantes que comparen las técnicas artísticas de las herramientas de IA con sus propias obras. Los docentes o mentores humanos deben animar a los estudiantes a desarrollar y aplicar su imaginación y creatividad, algo que la IAGen no puede sustituir. Prompt de ejemplo: <i>Sugiere algunas ideas que me inspiren para crear una imagen sobre [temas/ideas].</i>	Puede exponer a los niños a contenidos inapropiados u ofensivos, que vulnerarían su derecho a la salvaguarda y el bienestar. Las herramientas de IAGen aumentan el riesgo de impedir que los estudiantes desarrollen su imaginación y creatividad.
Entrenador 1:1 de codificación o aritmética	Conocimientos y habilidades conceptuales de programación a nivel introductorio. También podría aplicarse al aprendizaje de las matemáticas básicas.	Apoyar el aprendizaje a ritmo propio de conocimientos y habilidades básicas de codificación, encontrar errores en la codificación de los estudiantes y ofrecer retroalimentación inmediata, y ajustar las respuestas a las preguntas. Transformación potencial: Docente de codificación 1:1 a nivel introductorio	Partir de la lista del apartado 1.3 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Encontrar y definir un problema y diseñar algoritmos para resolverlo siguen siendo aspectos fundamentales del aprendizaje de la codificación y la programación. Los estudiantes deben tener motivaciones intrínsecas para utilizar la codificación, junto con algunos conocimientos y habilidades básicos en el uso del lenguaje de programación.	Los docentes y entrenadores humanos deben impartir conocimientos y competencias básicas, e inspirar a los estudiantes a utilizar el pensamiento computacional y la programación para resolver problemas, incluso mediante la codificación colaborativa. Prompt de ejemplo: <i>Sugiere algunas ideas poco usuales de programación.</i>	La precisión de retroalimentación y de sugerencias sigue siendo un tema problemático, ya que la IAGen no siempre estará en lo cierto. Existe un alto riesgo de que las herramientas de IAGen impidan a los estudiantes desarrollar capacidades y habilidades de pensamiento computacional para encontrar y definir problemas significativos de codificación.

5.3.4 IA generativa para facilitar la investigación o el aprendizaje basado en proyectos

Si no son utilizadas deliberadamente para facilitar el pensamiento de orden superior o la creatividad, las herramientas de IAGen tienden a estimular el plagio o resultados superficiales de “loro estocástico”. Sin embargo, dado que los modelos de IAGen han sido

entrenados a partir de datos a gran escala, cuentan con potencial para actuar como adversarios en diálogos socráticos o como asistentes de investigación en el aprendizaje basado en proyectos. No obstante, este potencial solo puede aprovecharse mediante procesos de diseño de enseñanza y aprendizaje cuyo objetivo sea activar el pensamiento de orden superior, como se muestra en la **Tabla 6**.

Tabla 6. Codiseño de usos de la IAGen para facilitar la investigación o el aprendizaje basado en proyectos

Usos potenciales pero no probados	Ámbitos de conocimiento apropiados o problemas	Resultados esperados	Herramientas adecuadas de IAGen y ventajas comparativas	Requisitos para los usuarios	Métodos pedagógicos humanos requeridos y ejemplos de prompts	Posibles riesgos
Adversario socrático	Problemas poco estructurados	Involucrar a los estudiantes en un diálogo evocador del cuestionamiento socrático del conocimiento previo, conducente al descubrimiento de nuevos conocimientos o a una comprensión más profunda. Transformación potencial: Oponente socrático 1:1	Partir de la lista del apartado 1.3 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	El estudiante debe haber alcanzado una edad que le permita mantener conversaciones independientes con herramientas de IAGen. Los estudiantes deben tener conocimientos previos y habilidades para corroborar si los argumentos y la información presentados son precisos.	Los docentes humanos pueden ayudar a preparar una lista de preguntas cada vez más profundas como ejemplos para que los estudiantes las adapten en sus prompts. Los estudiantes también pueden empezar con un prompt amplio como “Entabla conmigo un diálogo socrático que me ayude a adoptar una perspectiva crítica sobre [tema x]” y, después, profundizar gradualmente en el diálogo mediante prompts cada vez más refinados.	Las herramientas actuales de IAGen pueden generar respuestas similares o estándar que limitan la exposición de los estudiantes a diversos puntos de vista y perspectivas alternativas, conduciendo a un efecto de cámara de eco, y dificultando el desarrollo del pensamiento independiente.
Asesor para el aprendizaje basado en proyectos	Problemas de investigación poco estructurados en ciencias o estudios sociales.	Apoyar la creación de conocimiento ayudando a los estudiantes a realizar un aprendizaje basado en proyectos. Esto incluye que la IAGen cumpla un rol similar al del asesor de investigación descrito en la Tabla 3. Transformación potencial: Entrenador de aprendizaje basado en proyectos 1:1	Partir de la lista del apartado 1.3 para evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Los estudiantes podrían actuar como investigadores junior en la planificación e implementación del aprendizaje basado en proyectos. Los estudiantes deben tener edad suficiente para usar independientemente las plataformas de IAGen. Los estudiantes deben contar con la motivación y la capacidad de participar en actividades de aprendizaje basadas en proyectos autónomos, para no caer en la tentación de copiar y pegar pasivamente las respuestas ofrecidas por las herramientas de IAGen.	Los docentes humanos orientan a los estudiantes para que soliciten a la IAGen ideas básicas para la definición de problemas de investigación, como se sugiere en el apartado 5.3.1. Los estudiantes utilizan a nivel individual y grupal las herramientas de IAGen para realizar revisiones de la literatura, recolectar y procesar datos y crear reportes.	Los estudiantes sin sólidos conocimientos previos ni capacidad para verificar la veracidad de las respuestas pueden verse inducidos al error por la información ofrecida por las herramientas de IAGen. También esta puede limitar las discusiones e interacciones de los estudiantes con sus compañeros y reducir oportunidades de aprendizaje colaborativo, dañando potencialmente su desarrollo social.

5.3.5 IA generativa como apoyo a estudiantes con necesidades especiales

En teoría, los modelos de IAGen pueden ayudar a los estudiantes con discapacidad auditiva o visual. Las prácticas emergentes incluyen subtítulos o leyendas generadas por la IAGen para estudiantes sordos o con dificultades auditivas, y audio descripción generada por la IAGen para estudiantes con deficiencias visuales. Los modelos de IAGen también pueden convertir texto en habla y habla en texto para que las personas con deficiencias visuales, auditivas o del habla accedan a los contenidos, formulen preguntas y se comuniquen con sus pares. Sin embargo, esta función aún no se ha aprovechado a escala. Según la encuesta mencionada anteriormente, realizada por la UNESCO en 2023 sobre el uso de la IA en la educación por parte de los gobiernos, solo cuatro países (China, Jordania, Malasia y Qatar) informaron que sus organismos gubernamentales habían validado y recomendado herramientas asistidas por IA para apoyar el acceso inclusivo de estudiantes con discapacidad (UNESCO, 2023c).

Existe también una tendencia a entrenar las iteraciones de modelos de IAGen para ayudar a los estudiantes a

usar sus propias lenguas, incluidas las minoritarias e indígenas, para aprender y comunicarse. Por ejemplo, PaLM 2, el LLM de nueva generación de Google, está entrenado con datos paralelos que abarcan cientos de lenguas en forma de pares de textos de origen y destino. La inclusión de datos multilingües paralelos está diseñada para mejorar más la capacidad del modelo con el fin de comprender y generar texto multilingüe (Google, 2023b).

Al ofrecer traducciones, paráfrasis y correcciones automáticas en tiempo real, las herramientas de IAGen pueden ayudar a los estudiantes que emplean lenguas minoritarias a comunicar ideas y mejorar su colaboración con pares de diferentes orígenes lingüísticos. Sin embargo, esto no ocurrirá naturalmente a gran escala. Solo mediante un diseño adecuado se aprovechará este potencial para amplificar las voces de los grupos marginados.

Finalmente, también se ha sugerido que los sistemas de IAGen pueden realizar diagnósticos basados en la conversación, e identificar problemas psicológicos o socioemocionales, así como dificultades de aprendizaje. Sin embargo, sigue existiendo poca evidencia de que este enfoque sea efectivo o seguro, y cualquier diagnóstico requeriría la interpretación de profesionales especializados.

Tabla 7. Codiseño de usos de la IAGen para apoyar a estudiantes con necesidades especiales

Usos potenciales pero no probados	Ámbitos de conocimiento apropiados o problemas	Resultados esperados	Herramientas adecuadas de IAGen y ventajas comparativas	Requisitos para los usuarios	Métodos pedagógicos humanos requeridos y ejemplos de prompts	Posibles riesgos
Diagnóstico conversacional de las dificultades de aprendizaje	Esto puede ser útil para estudiantes con dificultades de aprendizaje causadas por problemas psicológicos, sociales o emocionales.	Utilizar el lenguaje natural para identificar las necesidades de estudiantes con problemas psicológicos, sociales o emocionales o con dificultades de aprendizaje, a fin de ofrecerles el apoyo o la enseñanza adecuados. Transformación potencial: Entrenador primario 1:1 para estudiantes con problemas sociales o emocionales o dificultades de aprendizaje.	Además de herramientas generales de IAGen, buscar <i>chatbots</i> controlados por IAGen. Evaluar si son localmente accesibles, de código abierto, rigurosamente probados o validados por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Los docentes o especialistas que trabajen con este grupo de estudiantes deberán asegurarse de que los consejos primarios sugeridos por el sistema IAGen sean acertados.	Los docentes o facilitadores deben crear ámbitos confortables para que el estudiante participe en una conversación a fin de diagnosticar problemas psicológicos, sociales o emocionales, o dificultades de aprendizaje.	Puede errarse inadvertidamente en el diagnóstico de las dificultades específicas del estudiante, dando lugar a que se le preste un apoyo inadecuado.

Usos potenciales pero no probados	Ámbitos de conocimiento apropiados o problemas	Resultados esperados	Herramientas adecuadas de IAGen y ventajas comparativas	Requisitos para los usuarios	Métodos pedagógicos humanos requeridos y ejemplos de prompts	Posibles riesgos
Herramientas de accesibilidad asistidas por IA	Permiten a los estudiantes con deficiencias auditivas o visuales acceder a un rango más amplio de contenidos, mejorando así la calidad de su aprendizaje.	Satisfacer las necesidades de acceso de los estudiantes y apoyar su adquisición de conocimientos específicos de la materia brindándoles subtítulos y/o interpretación en lengua de signos para contenidos de audio o vídeo, y descripciones de audio para textos u otros materiales visuales. Transformación potencial: Ayudas de lenguaje personalizadas 1:1 basadas en IA	Además de herramientas generales de IAGen, buscar generadores de subtítulos y audio descripciones relevantes y confiables basados en IA. Evaluar si son localmente accesibles, de código abierto, rigurosamente probados o validados por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Los docentes o facilitadores deben ayudar a los estudiantes a acceder a las herramientas de la IAGen y aprender a manejarlas. También deben asegurarse de que los resultados de las herramientas brinden apoyo a estos estudiantes y no refuercen los desafíos y sesgos a los que se enfrentan.	Necesidad de testear la accesibilidad de las plataformas o herramientas para identificar y solucionar problemas de accesibilidad antes de usarlas. Las herramientas de IAGen solo pueden brindar acceso al contenido, por lo que los docentes y facilitadores deben enfocarse en mejorar la calidad de su aprendizaje y su bienestar social. Los docentes y facilitadores deben enseñar a los estudiantes a crear prompts de voz o texto en función de sus capacidades.	Los subtítulos o audio descripciones generados por las plataformas de IAGen no diseñadas específicamente para apoyar la visión o la audición suelen ser imprecisos y pueden inducir a error a los estudiantes con necesidades especiales. Estas herramientas pueden reforzar inadvertidamente los sesgos existentes.
Amplificador generativo para estudiantes marginalizados	Puede ser útil para estudiantes de entornos lingüísticos o culturales minoritarios expresar y amplificar sus voces, participar en línea y realizar estudios sociales colaborativos.	Ofrecer traducciones en tiempo real, paráfrasis y corrección automática de la escritura para ayudar a los estudiantes de grupos marginalizados a usar sus propias lenguas para comunicarse con pares de diferentes orígenes lingüísticos. Transformación potencial: LLM inclusivos para estudiantes marginalizados	Un ejemplo concreto a considerar es PaLM 2. Evaluar si las herramientas de IAGen son localmente accesibles, de código abierto, rigurosamente probadas o validadas por las autoridades. Considerar además las ventajas y desafíos de cualquier herramienta concreta de IAGen, y asegurarse de que aborda correctamente necesidades humanas específicas.	Los estudiantes deben tener conocimientos u opiniones válidas sobre el tema de conversación o estudio colaborativo. Deben ser capaces de hacer aportaciones responsables y no discriminatorias y evitar los discursos de odio.	Los docentes o educadores deben diseñar estudios y tareas escritas para los estudiantes sobre temas sociales o culturales, u organizar seminarios en línea o colaboraciones interculturales para incentivar a los estudiantes a generar ideas y compartir opiniones.	Necesidad de identificar y corregir errores de traducción y paráfrasis en la IA que puedan causar malentendidos interculturales. Este uso puede ofrecer oportunidades para que los estudiantes marginalizados amplifiquen sus voces, pero no afectará la causa de la pobreza de datos y, por lo tanto, no podrá descolonizar las herramientas de IA.

6. IAGen y el futuro de la educación y la investigación

Las tecnologías de IAGen están aún en rápida evolución y es probable que tengan un profundo impacto en la educación y la investigación, además de que todavía no se las comprende en su totalidad. Por ello, sus posibles implicancias a largo plazo para la educación y la investigación requieren una atención inmediata y una revisión en profundidad.

6.1 Cuestiones éticas inexploradas

Las cada vez más sofisticadas herramientas de IAGen determinarán problemas éticos adicionales que deberán examinarse al detalle. Además de las secciones 2 y 3, se necesitan análisis más profundos y prospectivos para revelar y abordar cuestiones éticas inexploradas desde al menos las siguientes cinco perspectivas:

- **Acceso y equidad:** Los sistemas de IAGen en la educación pueden exacerbar las disparidades existentes en el acceso a la tecnología y los recursos educativos, profundizando aún más las inequidades.
- **Conexión humana:** Los sistemas de IAGen en la educación pueden reducir la interacción entre humanos y los aspectos socioemocionales esenciales del aprendizaje.
- **Desarrollo intelectual humano:** Los sistemas de IAGen en la educación pueden limitar la autonomía y la actuación de los estudiantes al ofrecer soluciones predeterminadas o reducir el abanico de posibles experiencias de aprendizaje. Es necesario investigar su impacto a largo plazo en el desarrollo intelectual de los jóvenes estudiantes.
- **Impacto psicológico:** Los sistemas de IAGen que imitan las interacciones humanas pueden tener efectos psicológicos desconocidos en los estudiantes, lo que genera preocupaciones sobre su desarrollo cognitivo y bienestar emocional, y con respecto a su potencial de manipulación.
- **Prejuicios y discriminación ocultos:** Conforme se desarrollan y aplican sistemas

de IAGen más sofisticados en la educación, éstos probablemente generen nuevos sesgos y formas de discriminación basados en los datos de entrenamiento y en los métodos usados por los modelos, pudiendo dar lugar a resultados desconocidos y potencialmente dañinos.

6.2 Derechos de autor y propiedad intelectual

La aparición de la IAGen está cambiando rápidamente la forma en que se crean, distribuyen y consumen las obras científicas, artísticas y literarias. La copia, distribución o uso no autorizados de obras protegidas por derechos de autor sin consentimiento del titular viola sus derechos exclusivos y puede acarrear consecuencias legales. Por ejemplo, el entrenamiento de modelos de IAGen ha sido acusado de infringir derechos de autor. En uno de los casos recientes, la canción generada por IA en la que aparecían “Drake” y “The Weeknd” (Abel Tesfaye) alcanzó millones de oyentes antes de ser retirada de Internet debido a una disputa sobre derechos de autor (Coscarelli, 2023). Aunque los marcos reguladores emergentes pretenden exigir a los proveedores de IAGen que reconozcan y protejan la propiedad intelectual de los propietarios de los contenidos utilizados por el modelo, cada vez resulta más desafiante determinar la propiedad y originalidad de la abrumadora cantidad de obras generadas. Esta falta de trazabilidad no solo genera preocupaciones respecto de la protección de derechos de los creadores y la garantía de una compensación justa por sus contribuciones intelectuales, sino que además presenta desafíos en contextos educativos acerca de cómo puede utilizarse responsablemente el producto de las herramientas de IAGen. Esto puede tener profundas implicancias para el sistema de investigación.

6.3 Fuentes de contenido y aprendizaje

Las herramientas de IAGen están cambiando la forma de generar y ofrecer contenidos de enseñanza y aprendizaje. En el futuro, los contenidos generados

a través de conversaciones entre humanos e IA pueden convertirse en una de las principales fuentes de producción de conocimiento. Es probable que esto socave aún más el compromiso directo de los estudiantes con los contenidos educativos basados en recursos, libros de texto y currículos creados y validados por humanos. La apariencia fidedigna del texto de la IAGen puede inducir a los jóvenes estudiantes a error, puesto que carecen de conocimientos previos suficientes como para reconocer imprecisiones o cuestionar el texto con certidumbre. También es discutible si el involucramiento de los estudiantes frente a contenidos no validados debe reconocerse como “aprendizaje”.

La concentración resultante en la información agregada de segunda mano también puede reducir las oportunidades de los estudiantes para construir conocimiento a través de métodos probados como la percepción directa y la experiencia del mundo real, el aprendizaje por ensayo y error, la realización de experimentos empíricos y el desarrollo del sentido común. También puede amenazar la construcción social del conocimiento y el estímulo de los valores sociales a través de prácticas colaborativas en el aula.

6.4 Respuestas homogeneizadas frente a resultados diversos y creativos

La IAGen restringe las narrativas plurales, ya que los resultados generados tienden a representar y reforzar los puntos de vista dominantes. La homogeneización resultante del conocimiento limita el pensamiento pluralista y creativo. La creciente dependencia de docentes y estudiantes de herramientas de IAGen en busca de sugerencias puede conducir a la estandarización y a la conformidad con las respuestas, debilitando el valor del pensamiento independiente y la indagación autónoma. La potencial homogeneización de la expresión en obras escritas y artísticas puede limitar la imaginación, la creatividad y las perspectivas alternativas de expresión de los estudiantes.

Los proveedores de IAGen y los docentes deben considerar en qué grado podrían desarrollarse y utilizarse las EdGPT para estimular la creatividad, la colaboración, el pensamiento crítico y otras capacidades de pensamiento de orden superior.

6.5 Replanteamiento de la evaluación y los resultados de aprendizaje

Las implicancias de la IAGen para la evaluación van más allá de las preocupaciones inmediatas sobre los estudiantes que hacen trampas en sus tareas escritas. Debemos enfrentarnos al hecho de que la IAGen puede generar trabajos y ensayos relativamente bien organizados e impresionantes obras de arte, así como aprobar algunos exámenes basados en el conocimiento en ciertas áreas temáticas. Por lo tanto, tenemos que replantearnos qué es exactamente lo que se debería aprender y con qué fines, y cómo se va a evaluar y validar el aprendizaje.

El debate crítico entre docentes, personas a cargo de formular políticas, estudiantes y demás partes interesadas debe considerar las siguientes cuatro categorías de resultados del aprendizaje:

Valores: Los valores requeridos para garantizar un diseño y uso de la tecnología centrados en el ser humano son fundamentales para replantearse los resultados del aprendizaje y su evaluación en la era digital. Al revisar el propósito de la educación, deben explicitarse los valores que informan el modo en que la tecnología se relaciona con ella. Es a través de esta lente normativa que los resultados del aprendizaje y su evaluación y validación deben actualizarse de forma iterativa para responder al uso cada vez más extendido de la tecnología, incluida la IA, en la sociedad.

Conocimientos y habilidades fundacionales: Incluso en los ámbitos de competencias en los que las herramientas de la IAGen pueden hacerlo mejor que los humanos, los estudiantes seguirán necesitando conocimientos y habilidades fundacionales robustos. La alfabetización básica, la aritmética elemental y la alfabetización científica básica seguirán siendo claves para la educación en el futuro. El alcance y la naturaleza de estas habilidades fundacionales deberán revisarse periódicamente para reflejar los entornos cada vez más ricos en IA en los que vivimos.

Habilidades de pensamiento de orden superior: Los resultados del aprendizaje tendrán que incluir las habilidades necesarias para apoyar el pensamiento de orden superior y la resolución de problemas basados en la colaboración humano-IA y el uso de los resultados generados por la IAGen. Esto puede incluir la comprensión del rol de los conocimientos factuales y conceptuales

como base del pensamiento de orden superior y la evaluación crítica de los contenidos generados por la IA.

Competencias profesionales necesarias para trabajar con la IA: En los ámbitos en los que la IA puede hacerlo mejor que los humanos y está automatizando unidades de tareas, los estudiantes humanos necesitan adquirir nuevas competencias que les permitan desarrollar, operar y trabajar con las herramientas de IAGen. El rediseño de los resultados de aprendizaje y la evaluación educativa deberán reflejar las habilidades profesionales necesarias para los nuevos puestos de trabajo creados por la IA.

6.6 Procesos de reflexión

La perspectiva más fundamental de las implicancias a largo plazo de la IAGen para la educación y la investigación sigue siendo la relación complementaria entre la actuación humana y las máquinas. Una de las cuestiones clave es si los humanos pueden ceder los niveles básicos de pensamiento y los procesos de adquisición de habilidades a la IA y concentrarse en las habilidades de pensamiento de orden superior basadas en los resultados generados por la IA.

La escritura, por ejemplo, está asociada normalmente con la estructuración del pensamiento. Con la IAGen, en vez de partir de cero para planificar las metas, el alcance y el esquema de un conjunto de ideas, los humanos pueden empezar ahora con un bosquejo bien estructurado facilitado por la IAGen. Algunos expertos han caracterizado el uso de la IAGen para generar texto de esta forma como “escribir sin pensar” (Chayka, 2023). A medida que se difundan estas nuevas prácticas asistidas por la IAGen, los métodos establecidos para adquisición y evaluación de habilidades de escritura deberán adaptarse. Una opción a futuro es que el aprendizaje de la escritura se enfoque en la construcción de habilidades de

planificación y redacción de *prompts*, evaluación crítica de los resultados de la IAGen y pensamiento de orden superior, así como en la coescritura basada en los esquemas de la IAGen.

Conclusiones finales

Desde la perspectiva de un enfoque centrado en el ser humano, las herramientas de IA deberían diseñarse para ampliar o aumentar las capacidades intelectuales y las habilidades sociales humanas, y no para socavarlas, entrar en conflicto con ellas o usurparlas. Desde hace tiempo se espera que las herramientas de IA puedan integrarse como parte de las herramientas disponibles para los humanos a fin de respaldar el análisis y la acción en pro de futuros más inclusivos y sostenibles.

Para que la IA sea un elemento confiable en la colaboración entre personas y máquinas -a nivel individual, institucional y sistémico-, el enfoque centrado en el ser humano de la *Recomendación sobre la ética de la IA* (UNESCO, 2021) debe concretarse e implementarse en función de las características específicas de tecnologías emergentes como la IAGen. Solo así podremos garantizar que la IAGen se convierta en una herramienta confiable para investigadores, docentes y estudiantes.

Aunque la IAGen debe utilizarse al servicio de la educación y la investigación, todos debemos ser conscientes de que también podría cambiar los sistemas establecidos y sus fundamentos en dichos ámbitos. La transformación de la educación y la investigación que pueda generar la IAGen, si se produce, debería ser rigurosamente revisada y orientada por un enfoque centrado en el ser humano. Solo así podremos asegurar que el potencial de la IA en particular, y de todas las demás categorías de tecnologías utilizadas en la educación en general, mejoren las capacidades humanas para construir futuros digitales inclusivos para todos.

Referencias

- Administración del Ciberespacio de China. 2023a. 国家互联网信息办公室关于《生成式人工智能服务管理办法（征求意见稿）》公开征求意见的通知 [Notificación de la Administración del Ciberespacio de China sobre los Comentarios Públicos acerca de las "Medidas Administrativas para los Servicios de Inteligencia Artificial Generativa (Proyecto para Comentarios)"]. Administración del Ciberespacio de China (ACC), Beijing. (En chino.) Disponible en: http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm (Consultado por última vez el 19 de julio de 2023)
- . 2023b. 生成式人工智能服务管理暂行办法 [Medidas provisionales para la gestión de los servicios de inteligencia artificial generativa]. Administración del Ciberespacio de China (ACC), Beijing. (En chino.) Disponible en: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (Consultado por última vez el 19 de julio de 2023)
- Anders, B. A. 2023. *Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking?* Cambridge, Cell Press. Disponible en: <https://doi.org/10.1016/j.patter.2023.100694> (Consultado por última vez el 23 de junio de 2023)
- Bass, D. y Metz, R. 2023. *OpenAI's Sam Altman Urges Congress to Regulate Powerful New Technology*. Nueva York, Bloomberg. Disponible en: <https://www.bloomberg.com/news/newsletters/2023-05-17/openai-sam-altman-urges-congress-to-regulate-powerful-new-ai-technology> (Consultado por última vez el 23 de junio de 2023)
- Bender, E. M., Gebru, T., McMillan-Major, A. y Shmitchell, S. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Nueva York, Association for Computing Machinery. Disponible en: <https://doi.org/10.1145/3442188.3445922> (Consultado por última vez el 23 de junio de 2023)
- Bommasani, R. et al. 2021. *On the Opportunities and Risks of Foundation Models*. Stanford, Stanford University. Disponible en: <https://crfm.stanford.edu/report.html> (Consultado por última vez el 23 de junio de 2023)
- Bove, T. 2023. *Big tech is making big AI promises in earnings calls as ChatGPT disrupts the industry: 'You're going to see a lot from us in the coming few months'*. Nueva York, Fortune. Disponible en: <https://fortune.com/2023/02/03/google-meta-apple-ai-promises-chatgpt-earnings> (Consultado por última vez el 3 de julio de 2023)
- Chayka, K. 2023. *My A.I. Writing Report*. Nueva York, The New Yorker. Disponible en: <https://www.newyorker.com/culture/infinite-scroll/my-ai-writing-robot> (Consultado por última vez el 1 de agosto de 2023)
- Chen, L., Zaharia, M., y Zou, J. 2023. *How Is ChatGPT's Behavior Changing over Time?* Ithaca, arXiv. <https://arxiv.org/pdf/2307.09009> (Consultado por última vez el 31 de julio de 2023)
- Comisión Europea. 2021. *La Ley de Inteligencia Artificial de la UE*. Bruselas, Comisión Europea. Disponible en: <https://artificialintelligenceact.eu/es> (Consultado por última vez el 23 de junio de 2023)
- Comisión Federal de Comercio. 1998. *Children's Online Privacy Protection Act of 1998*. Washington DC, Comisión Federal de Comercio. Disponible en: <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa> (Consultado por última vez el 4 de septiembre de 2023)
- Coscarelli, J. 2023. *An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World*. Nueva York, New York Times. Disponible en: <https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html> (Consultado por última vez el 30 de agosto de 2023)
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., Carter, L., Chowdhury, S., Crick, T., Cunningham, S. W., Davies, G. H., Davison, R. M., Dé, R., Dennehy, D., Duan, Y., Dubey, R., Dwivedi, R., Edwards, J. S., Flavián, C., Gauld, R., Grover, V., Hu, M.-C., Janssen, M., Jones, P., Junglas, I., Khorana, S., Kraus, S., Larsen, K. R., Latreille, P., Laumer, S., Malik, F. T., Mardani, A., Mariani, M., Mithas, S., Mogaji, E., Horn Nord, J., O'Connor, S., Okumus, F., Pagani, M., Pandey, N., Papagiannidis, S., Pappas, I. O., Pathak, N., Pries-Heje, J., Raman, R., Rana, N. P., Rehm, S.-V., Ribeiro-Navarrete, S., Richter, A., Rowe, F., Sarker, S., Stahl, B. C., Tiwari,

- M. K., van der Aalst, W., Venkatesh, V., Viglia, G., Wade, M., Walton, P., Wirtz, J. y Wright, R. 2023. Artículo de opinión: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, Vol. 71. Ámsterdam, Elsevier, p. 102642. Disponible en: <https://doi.org/10.1016/j.ijinfomgt.2023.102642> (Consultado por última vez el 25 de agosto de 2023)
- E2Analyst. 2023. *GPT-4: Everything you want to know about OpenAI's new AI model*. San Francisco, Medium. Disponible en: <https://medium.com/predict/gpt-4-everything-you-want-to-know-about-openais-new-ai-model-a5977b42e495> (Consultado por última vez el 1 de agosto de 2023)
- Giannini, S. 2023. *La IA generativa y el futuro de la educación*. París, UNESCO. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000385877_spa (Consultado por última vez el 29 de agosto de 2023)
- Google. 2023a. *Recommendations for Regulating AI*. Mountain View, Google. Disponible en: <https://ai.google/static/documents/recommendations-for-regulating-ai.pdf> (Consultado por última vez el 23 de junio de 2023)
- . 2023b. *PaLM 2 Technical Report*. Mountain View, Google. Disponible en: <https://doi.org/10.48550/arXiv.2305.10403> (Consultado por última vez el 20 de julio de 2023)
- Lin, B. 2023. *AI Is Generating Security Risks Faster Than Companies Can Keep Up*. Nueva York, The Wall Street Journal. Disponible en: <https://www.wsj.com/articles/ai-is-generating-security-risks-faster-than-companies-can-keep-up-a2bdedd4> (Consultado por última vez el 25 de agosto de 2023)
- Marcus, G. 2022. Hoping for the Best as AI Evolves. *Communications of the ACM*, Vol. 66, No. 4. Nueva York, Association for Computing Machinery. Disponible en: <https://doi.org/10.1145/3583078> (Consultado por última vez el 23 de junio de 2023)
- Marwala, T. 2023. *Algorithm Bias — Synthetic Data Should Be Option of Last Resort When Training AI Systems*. Tokio, United Nation University. Disponible en: <https://unu.edu/article/algorithm-bias-synthetic-data-should-be-option-last-resort-when-training-ai-systems> (Consultado por última vez el 31 de julio de 2023)
- Metz, C. 2021. *Who Is Making Sure the A.I. Machines Aren't Racist?* Nueva York, The New York Times. Disponible en: <https://www.nytimes.com/2021/03/15/technology/artificial-intelligence-google-bias.html> (Consultado por última vez el 23 de junio de 2023)
- Murphy Kelly, S. 2023. *Microsoft is bringing ChatGPT technology to Word, Excel and Outlook*. Atlanta, CNN. Disponible en: <https://edition.cnn.com/2023/03/16/tech/openai-gpt-microsoft-365/index.html> (Consultado por última vez el 25 de agosto de 2023)
- Nazaretsky, T., Cukurova, M. y Alexandron, G. 2022a. An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. *LAK22: 12th International Learning Analytics and Knowledge Conference*. Vancouver, Association for Computing Machinery, pp. 55-66.
- Nazaretsky, T., Ariely, M., Cukurova, M. y Alexandron, G. 2022b. Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, Vol. 53, No. 4. Hoboken, NJ, Wiley, pp. 914-931. Disponible en: <https://doi.org/10.1111/bjet.13232> (Consultado por última vez el 1 de agosto de 2023)
- Ocampo, Y. 2023. *Singapore Unveils AI Government Cloud Cluster*. Singapur, OpenGov Asia. Disponible en: <https://opengovasia.com/singapore-unveils-ai-government-cloud-cluster> (Consultado por última vez el 25 de agosto de 2023)
- Oficina del Derecho de Autor de EEUU. 2023. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence. *Federal Register*, Vol. 88, No. 51. Washington DC, EEUU, Oficina del Derecho de Autor, Biblioteca del Congreso, pp. 16190-16194. Disponible en: <https://www.federalregister.gov/d/2023-05321> (Consultado por última vez el 3 de julio de 2023)
- OpenAI. 2018. *AI and compute*. San Francisco, OpenAI. Disponible en: <https://openai.com/research/ai-and-compute> (Consultado por última vez el 23 de junio de 2023)
- . 2023. *Educator considerations for ChatGPT*. San Francisco, OpenAI. Disponible en: <https://platform.openai.com/docs/chatgpt-education> (Consultado por última vez el 23 de junio de 2023)
- Popli, N. 2023. *The AI Job That Pays Up to \$335K—and You Don't Need a Computer Engineering Background*. Nueva York, TIME USA. Disponible en: <https://time.com/6272103/ai-prompt-engineer-job> (Consultado por última vez el 23 de junio de 2023)

- Roose, K. 2022. *An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy*. Nueva York, The New York Times. Disponible en: <https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html> (Consultado por última vez el 23 de junio de 2023)
- Russell Group, 2023. *Russell Group principles on the use of generative AI tools in education*. Cambridge, Russell Group. Disponible en: https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf (Consultado por última vez el 25 de agosto de 2023)
- Stanford University. 2019. *Artificial Intelligence Index Report*. Stanford, Stanford University. Disponible en: <https://hai.stanford.edu/ai-index-2019> (Consultado por última vez el 23 de junio de 2023)
- . 2023. *Artificial Intelligence Index Report*. Stanford, Stanford University. Disponible en: <https://hai.stanford.edu/research/ai-index-2023> (Consultado por última vez el 23 de junio de 2023)
- The Verge. 2023a. *OpenAI co-founder on company's past approach to openly sharing research: 'We were wrong'*. Washington DC, Vox Media. Disponible en: <https://www.theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview> (Consultado por última vez el 1 de agosto de 2023)
- . 2023b. *OpenAI CEO Sam Altman on GPT-4: 'people are begging to be disappointed and they will be'*. Washington DC, Vox Media. Disponible en: <https://www.theverge.com/23560328/openai-gpt-4-rumor-release-date-sam-altman-interview> (Consultado por última vez el 1 de agosto de 2023)
- Tlili, A., Shehata, B., Agyemang Adarkwah, M., Bozkurt, A., Hickey, D. T., Huang, R. & Agyemang, B. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, Vol. 10, No. 15. Berlín, Springer. Disponible en: <https://doi.org/10.1186/s40561-023-00237-x> (Consultado por última vez el 23 de junio de 2023)
- UNESCO. 2019. *Consenso de Beijing sobre la inteligencia artificial y la educación*. París, UNESCO. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000368303> (Consultado por última vez el 3 de julio de 2023)
- . 2022a. *Recomendación sobre la ética de la inteligencia artificial*. París, UNESCO. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa (Consultado por última vez el 3 de julio de 2023)
- . 2022b. *Inteligencia artificial y educación: guía para las personas a cargo de formular políticas*. París, UNESCO. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000379376> (Consultado por última vez el 23 de julio de 2023)
- . 2022c. *Currículos de IA para la enseñanza preescolar, primaria y secundaria: un mapeo de los currículos de IA aprobados por los gobiernos*. París, UNESCO. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000380602_spa (Consultado por última vez el 20 de julio de 2023)
- . 2022d. *Directrices para la formulación de políticas y planes maestros de TIC en educación*. París, UNESCO. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000385091> (Consultado por última vez el 31 de julio de 2023)
- . 2023a. *Inteligencia Artificial: La UNESCO pide a los gobiernos que apliquen sin demora el Marco Ético Mundial*. París, UNESCO. Disponible en: <https://www.unesco.org/es/articles/inteligencia-artificial-la-unesco-pide-los-gobiernos-que-apliquen-sin-demora-el-marco-etico-mundial> (Consultado por última vez el 3 de julio de 2023)
- . 2023b. *Mapping and analysis of governmental strategies for regulating and facilitating the creative use of GenAI*. No publicado.
- . 2023c. *Survey for the governmental use of AI as a public good for education*. No publicado (Presentado a la UNESCO).
- . 2023. *Technology in Education: A tool on whose terms?* París, Global Education Monitoring Report Team. Disponible en: <https://unesdoc.unesco.org/ark:/48223/pf0000385723> (Consultado por última vez el 25 de agosto de 2023)
- . 2023. *ChatGPT e inteligencia artificial en la educación superior: guía de inicio rápido*. Caracas, Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC). Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000385146_spa (Consultado por última vez el 25 de agosto de 2023)
- Unión Europea. 2016. *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)*. Bruselas, Diario Oficial de la Unión Europea. Disponible en: <http://data.europa.eu/eli/reg/2016/679/oj> (Consultado por última vez el 23 de junio de 2023)

Notas finales

- 1 Los modelos de IAGen estuvieron a disposición de los investigadores y otras partes interesadas mucho antes que ChatGPT. Por ejemplo, en 2015 Google lanzó lo que denominó 'DeepDream' (<https://en.wikipedia.org/wiki/DeepDream>).
- 2 Ver <https://chat.openai.com>
- 3 Para una explicación de las técnicas y tecnologías de la IA y su relación, ver UNESCO, 2022b, pp. 8-10.
- 4 Nótese que, dado que la IAGen es aun relativamente nueva, las distintas empresas suelen utilizar estos términos de formas diferentes y, en ocasiones, emplean palabras distintas para referirse a lo mismo.
- 5 Existe la preocupación de que los datos utilizados para entrenar futuras iteraciones del GPT de OpenAI incluyan cantidades sustanciales de texto generado por versiones anteriores del mismo. Este bucle autorreferencial podría contaminar los datos de entrenamiento y comprometer así las capacidades de los futuros modelos de GPT.
- 6 NB OpenAI, la compañía que desarrolló los GPT de esta tabla, no ha hecho pública información detallada sobre el GPT-4 (The Verge, 2023a). De hecho, el número de parámetros ha sido desmentido por el CEO de OpenAI (The Verge, 2023b). Sin embargo, las cifras incluidas aquí han sido divulgadas por varios medios (por ejemplo, ver E2Analyst, 2023). En cualquier caso, la principal conclusión es que GPT-4 se basa en un conjunto de datos mucho mayor y utiliza un número de parámetros muy superior al de GPT-3.
- 7 Ver <https://crfm.stanford.edu/2023/03/13/alpaca.html>
- 8 Ver <https://bard.google.com>
- 9 Ver <https://writesonic.com/chat>
- 10 Ver <https://yiyian.baidu.com/welcome>
- 11 Ver <https://huggingface.co/chat>
- 12 Ver <https://www.jasper.ai>
- 13 Ver <https://ai.facebook.com/blog/large-language-model-llama-meta-ai>
- 14 Ver <https://open-assistant.io>
- 15 Ver <https://www.alizila.com/alibaba-cloud-debuts-generative-ai-model-for-corporate-users>
- 16 Ver <https://you.com>
- 17 Ver <https://www.chatpdf.com>
- 18 Ver <https://elicit.org>
- 19 Ver <https://www.perplexity.ai>
- 20 Ver <https://tools.zmo.ai/webchatgpt>
- 21 Ver <https://www.compose.ai>
- 22 Ver <https://www.teamsmart.ai>
- 23 Ver <https://wiseone.io>
- 24 Ver <https://www.microsoft.com/en-us/bing>
- 25 Ver <https://www.craiyon.com>
- 26 Ver <https://openai.com/product/dall-e-2>
- 27 Ver <https://dream.ai/create>
- 28 Ver <https://www.fotor.com/features/ai-image-generator>
- 29 Ver <https://www.midjourney.com>
- 30 Ver <https://creator.nightcafe.studio>
- 31 Ver <https://writesonic.com/photosonic-ai-art-generator>
- 32 Ver <https://elai.io>
- 33 Ver <https://www.gliacloud.com>
- 34 Ver <https://pictory.ai>
- 35 Ver <https://runwayml.com>
- 36 Ver <https://www.aiva.ai>
- 37 Ver <https://boomy.com>
- 38 Ver <https://soundraw.io>
- 39 Ver <https://www.voicemod.net/text-to-song>
- 40 Ver <https://openai.com/research/gpt-4>
- 41 Ver <https://www.educhat.top> and <https://www.mathgpt.com>

- 42 Ver <https://www.educhat.top>
- 43 Ver <https://www.mathgpt.com>
- 44 Existen algunas excepciones, como Hugging Face, un grupo dedicado al desarrollo de IA de código abierto.
- 45 Ver, por ejemplo, llamadas entre Google (2023a) y OpenAI (Bass y Metz, 2023).
- 46 Para conocer un proyecto de regulación de la IA, ver el Proyecto de Ley de IA de la Comisión Europea (2021).
- 47 La reseña se basó en datos recogidos en una encuesta de la UNESCO distribuida a sus 193 Estados Miembros sobre el uso gubernamental de la IA en la educación (UNESCO, 2023c), el Observatorio de Políticas de IA de la OCDE y el Informe sobre el Índice de IA de la Universidad de Stanford (Universidad de Stanford, 2023), así como en información de primera mano obtenida de un grupo de expertos internacionales.
- 48 Ver <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
- 49 De acuerdo con el mapeo, en abril de 2023, los siguientes países habían publicado estrategias nacionales de IA: Alemania, Argentina, Australia, Austria, Bélgica, Benín, Brasil, Canadá, Bulgaria, Chile, China, Chipre, Colombia, Dinamarca, Egipto, Estonia, Finlandia, Francia, Hungría, India, Indonesia, Irlanda, Islandia, Italia, Japón, Jordania, Letonia, Lituania, Luxemburgo, Malasia, Malta, Mauricio, México, Países Bajos, Noruega, Nueva Zelanda, Omán, Perú, Polonia, Portugal, Filipinas, Qatar, República de Corea, Rumania, Federación Rusa, Arabia Saudita, Serbia, Singapur, Eslovenia, España, Suecia, Tailandia, Turquía, Túnez, Emiratos Árabes Unidos, Reino Unido, Estados Unidos, Uruguay y Vietnam. Asimismo, algunos países han incorporado estrategias de IA en estrategias digitales o de TIC más amplias, como Argelia, Botswana, Kazajstán, Kenya, Sierra Leona, Eslovaquia, Suiza y Uganda.
- 50 Según una revisión rápida de todas las estrategias nacionales de IA (UNESCO, 2023b), más de 40 estrategias tienen secciones dedicadas al tema de la ética.
- 51 Según una revisión rápida de todas las estrategias nacionales de IA (UNESCO, 2023b), unas 45 estrategias cuentan con secciones dedicadas al tema de la educación.
- 52 Ver <https://openai.com/policies/terms-of-use>
- 53 En algunos países, un docente tiene un asistente de enseñanza (AE) cuya función es dedicar tiempo a responder a las preguntas de los estudiantes y cubrir el material del curso. La IAGen podría utilizarse para desarrollar un gemelo generativo de un AE, que sirva de apoyo a los estudiantes y a otros docentes, pero que también podría causar ciertos problemas (por ejemplo, respecto de las relaciones sociales en el aula).



unesco

Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura

Guía para el uso de IA generativa en educación e investigación

Esta guía apunta a apoyar la planificación de regulaciones, políticas y programas de desarrollo de capacidades humanas apropiados para garantizar que la IAGen se convierta en una herramienta que genuinamente beneficie y empodere a docentes, estudiantes e investigadores. Explica las técnicas de IA utilizadas por la IAGen y elabora una lista de modelos de GPT puestos a disposición del público, especialmente aquellos con licencias de código abierto. También abre un debate sobre el surgimiento de EdGPT: los modelos de IAGen entrenados con datos específicos para servir a fines educativos. Además, resume algunas de las controversias clave en torno a la IAGen, desde el empeoramiento de la pobreza digital a la homogeneización de opiniones, y desde los más profundos deepfakes a cuestiones de derechos de autor. Con base en una visión humanista, la guía sugiere pasos clave para la regulación de las herramientas de IAGen, que incluyen la obligación de proteger la privacidad de los datos y el establecimiento de un límite de edad para las conversaciones independientes con las plataformas de IAGen. Para guiar el uso adecuado de las herramientas en la educación y la investigación, esta guía propone un enfoque humano-agente y adecuado a la edad para los procesos de validación ética y diseño pedagógico.

