

1. La docencia en la educación superior en México en el contexto de los modelos grandes de lenguaje: retos y oportunidades



GABRIEL PÉREZ SALAZAR*

DOI: <https://doi.org/10.52501/cc.351.01>

Resumen

A partir de una revisión del proceso de entrenamiento y funcionamiento de los Modelos Grandes de Lenguaje, se ofrece una reflexión crítica sobre algunos de los principales desafíos y oportunidades que trae consigo la irrupción de la Inteligencia Artificial Generativa en la educación superior. Se subraya la necesidad de adaptar las prácticas pedagógicas, con especial atención a los mecanismos de seguimiento y evaluación, en un contexto como el mexicano, marcado por profundas desigualdades en el acceso y uso de las tecnologías de la información y comunicación. Además, se destaca la limitada representación de la diversidad cultural del país en los principales modelos de IA, lo que plantea retos adicionales para su integración en el ámbito educativo.

Palabras clave: *docencia, inteligencia artificial generativa, brecha digital.*

Introducción

La Inteligencia Artificial (IA) puede ser definida como un conjunto de dispositivos que muestran un comportamiento aparentemente *inteligente* a través de acciones que son el resultado de un análisis de su entorno que, con

* Doctor en Ciencias Políticas y Sociales. Profesor-investigador de la Universidad Autónoma de Coahuila, México. ORCID: <https://orcid.org/0000-0003-0083-4083>

cierto grado de autonomía, permiten alcanzar objetivos determinados (Dignum, 2019). Más allá de las múltiples maneras en que se puede entender dicha cualidad,¹ como veremos más adelante en este capítulo, se trata de una innovación tecnológica que tiene ya varias décadas de ser aplicada al campo de la educación.

Sin embargo, ha sido a partir de la disposición de modelos como ChatGPT 3.0, a finales de 2022, que el asunto se ha colocado de manera relevante en la atención de quienes nos dedicamos a la educación universitaria. Con base en la representación mental que tengamos alrededor de esta tecnología, se han manifestado muy diversos posicionamientos en este sentido, que van desde el desbordante entusiasmo hasta el más oscuro de los pesimismos. Así, las iniciativas de Sal Khan² contrastan con discusiones en línea³ en las que se debate si llegará el día en que la IA reemplace al personal docente en el aula. Y las respuestas institucionales han sido igual de variables, mientras que algunas incentivan su exploración y uso responsable, otras vanamente pretenden prohibir su uso.

Como posiblemente hemos ido descubriendo en los últimos meses, sabemos que se trata de artilugios capaces de generar en unos pocos segundos, contenidos que solemos incorporar en los métodos de evaluación y de seguimiento de nuestro estudiantado. Síntesis de lecturas, ensayos, composiciones y hasta tesis completas, pueden ser elaboradas a partir de modelos como ChatGPT, Gemini, Claude y muchos otros servicios que hoy en día están a la disposición de cualquier persona usuaria con un acceso a internet. Y eso, refiriéndonos solamente a los sistemas generadores de lenguaje natural, que se suman a otros modelos capaces de dar lugar a mapas mentales, esquemas, imágenes, música y videos.

¹ Hay diversos encuadres sobre la inteligencia que enfatizan distintos aspectos, entre los que destacan el aprendizaje, el almacenamiento de información (memoria) y otros procesos que en la especie humana han sido abordados desde las ciencias cognitivas, como el razonamiento, la adaptabilidad y la capacidad de resolver problemas.

² Fundador en 2006 de Khan Academy, que es una organización sin fines de lucro dedicada al desarrollo de recursos educativos. El 14 de marzo de 2023, presentó Khanmigo, que es un *chatbot* de paga basado en ChatGPT 4, diseñado como asistente en asignaturas como matemáticas, ciencia y humanidades.

³ Por ejemplo: <https://www.projectmanagement.com/discussion-topic/201407/will-the-day-come-when-artificial-intelligence-replaces-teachers-and-faculty-members-in-universities>

Desde esta base, el objetivo del presente trabajo es reflexionar en torno a algunos de los principales retos y oportunidades que plantea la Inteligencia Artificial Generativa (IAG) en los espacios de educación superior en México y otras regiones hispanohablantes, particularmente, desde los llamados Modelos Grandes de Lenguaje (LLM por sus siglas en inglés). Partiremos de una revisión de la manera en que estos dispositivos han sido entrenados, que nos permita una comprensión básica de sus posibilidades y limitaciones. Luego, haremos un recorrido por algunos de los antecedentes más relevantes relacionados con el uso de la IA en la educación, como punto de partida para la reflexión de su uso en espacios culturales como el mexicano, caracterizado por amplias desigualdades en el acceso y uso de las tecnologías de la información y la comunicación.

Se trata de un abordaje crítico que, desde el marco dado por el campo académico de la Comunicación,⁴ abona a un asunto de la mayor actualidad y pertinencia, tomando en cuenta la creciente incorporación de la IA en una gran cantidad de sectores productivos. Además, como argumentaremos, México se ubica en un contexto cultural que, en relación con el angloparlante, se encuentra en una posición de subrepresentación, que, derivada de la brecha digital, tiene el potencial de manifestarse como una condición de desigualdad en términos de la operación de los LLM en estos contextos educativos.

Sistemas de Inteligencia Artificial Generativa: Entrenamiento y operación de los LLM

Existe una gran diversidad de desarrollos y aplicaciones de la IA. Algunos, han demostrado ser sumamente desafiantes, como los sistemas de conducción autónoma de vehículos que, luego de muchos años de investigación y recursos invertidos, al momento de escribir este trabajo seguían sin ser una opción suficientemente segura (Koon, 2023). Otros desarrollos se han orientado al ámbito de la medicina donde, a partir de modelos de reconocimiento

⁴ Uno de los ejes reconocidos dentro de este campo está dado por el estudio de los procesos de generación y transmisión de la cultura (Tindall, 1976), que es precisamente donde se ubican los LLM.

de patrones, han probado su utilidad en el diagnóstico temprano de algunos tipos de cáncer (Hunter *et al.*, 2022), o en la química, en el desarrollo de nuevas moléculas (Tang *et al.*, 2024).

Existen piezas de IA que, quizá sin que seamos conscientes de ello, hemos utilizado desde hace ya varios años. Por ejemplo, los sistemas de recomendación de contenidos en plataformas de *streaming* como Netflix y Spotify, pertenecen a esta categoría. Con base en nuestros consumos previos y reacciones, constantemente nos sugieren series, películas y canciones (Pajkovic, 2022; Seaver, 2022).

Desde una línea similar, uno de los sistemas de IA que mayor impacto han tenido en las democracias actuales, son los algoritmos que, presentes en todas las plataformas sociodigitales, generan el flujo visible de publicaciones que comparten nuestros contactos.⁵ Como ha sido estudiado por autores como Benassini (2022), Javed y Javed (2023) y Aïmeur *et al.* (2023), la búsqueda por retener nuestra atención el mayor tiempo posible, ha dado como resultado la creación de *burbujas informacionales* y *cámaras de eco* que han contribuido sensiblemente a la polarización de la opinión pública, así como a una mayor visibilización de posturas extremas que se manifiestan en todo tipo de prejuicios, discursos de odio y actos de discriminación (como racismo, xenofobia, misoginia, LGBTQfobia, clasismo, etc.).

Todas estas aplicaciones de la IA que hemos mencionado han tenido diversos efectos en los consumos y patrones de exposición a contenidos portadores de toda clase de ideologías, posturas y formas de entender la realidad. Se trata de verdaderas mediaciones que, en el sentido que propone Martín-Barbero (1987), inciden en la percepción del mundo y su encuadre.

Más recientemente,⁶ han sido desarrolladas nuevas formas de IA, cuyo impacto estamos apenas debatiendo. Nos referimos a los modelos de Inteligencia Artificial Generativa capaces de procesar lenguaje natural, conocidos como Large Language Models. A esta categoría pertenecen ChatGPT, Gemini, Claude y otros sistemas similares. A partir de Kaplan (2024), explica-

⁵ Esto es llamado *newsfeed* en algunas de estas plataformas.

⁶ Las fuentes especializadas coinciden en señalar el trabajo de Vaswani *et al.* (2017), titulado *Attention is all you need*, como una de las principales bases para la evolución de las redes neuronales que posibilitan los actuales modelos de IA generativa, junto con los procesadores de última generación de Nvidia.

remos el proceso de entrenamiento y su operación básica, con el fin de identificar los factores a partir de los cuales se manifiestan sus principales posibilidades y limitaciones.

Como explica este autor, este tipo de modelos parten de redes neuronales de aprendizaje profundo (*deep learning*) que son enfrentadas a un enorme conjunto de datos que se conoce como *corpus* de entrenamiento. Luego de múltiples iteraciones más o menos autónomas, que son el resultado de diversas funciones de cálculo y de álgebra lineal, en esta red neuronal se establecen patrones de relación n-dimensionales que, desde la experiencia humana, se traducen en asociaciones de sentido,⁷ a partir de unidades mínimas conocidas como *tokens*.⁸

Este proceso de entrenamiento se ve complementado por la intervención de seres humanos, en lo que se conoce como *Aprendizaje por Reforzamiento con Retroalimentación Humana* (RLHF,⁹ por sus siglas en inglés). En esta etapa, se hacen una serie de ajustes que modulan las salidas del modelo, de manera que no brinden respuestas que de alguna manera sean peligrosas, ilegales o inapropiadas. El resultado final de todo ello se conoce como *modelo fundacional*. Cada una de las distintas versiones de ChatGPT corresponden a diferentes modelos fundacionales.¹⁰ En el caso de Google, Bard era un modelo fundacional distinto a Gemini o Gemma.

Con base en estos modelos, las personas usuarias pueden elaborar preguntas o indicaciones, que se conocen como *prompts*. A partir de estas peticiones, el modelo generará una respuesta, palabra a palabra. Para lograrlo, lo primero que hace es descomponer el *prompt* en sus respectivos *tokens*. Con base en la red neuronal establecida luego del entrenamiento, se genera un vector n-dimensional que, a partir del total de *sentidos*¹¹ generados

⁷ El sentido es un proceso cognitivo propio de los seres humanos. En la explicación que estamos presentando, recurriremos a términos asociados a este tipo de procesos y experiencias humanas, que a modo de metáforas, pueden facilitar la comprensión de los procesos llevados a cabo por la IA. En tales ocasiones, colocaremos estas palabras en *italicas* para denotarlo.

⁸ Aunque no siempre es así, en muchos LLM un *token* suele corresponder con una palabra.

⁹ Reinforcement Learning from Human Feedback, en el original.

¹⁰ En los modelos fundacionales producidos por Open AI, algunos de ellos han sido: ChatGPT 3.5, 4, 4o, o1, etcétera.

¹¹ Para el sistema, en realidad se trata del valor estadísticamente más alto, no de un auténtico

por este *prompt*, apuntará a una región¹² determinada dentro de dicha red. A partir de dicha *posición*, el modelo tomará la primera palabra que tenga el valor más alto, en correspondencia con el sentido del *prompt*. La siguiente palabra será seleccionada bajo este mismo método, y así sucesivamente hasta completar una respuesta.

En los modelos de LLM más recientes, además del *sentido* de cada *prompt*, cada interacción contribuye a determinar con mayor precisión el contexto de las peticiones, y por lo tanto, inciden en el resultado del vector a partir del cual se genera una respuesta determinada. Hay, incluso, la posibilidad de activar una función de *memoria*, que en ChatGPT 4 y superiores, permite al modelo fundacional *recordar* el *sentido* de sesiones previas para quienes tengan registro en esta plataforma.

El resultado de estos algoritmos es la generación de respuestas que suelen tener altos niveles de congruencia en relación con el *prompt* introducido, desde una lógica interna. A pesar de que los modelos fundacionales más recientes han incorporado capacidades de reconocimiento en formatos multimedia (imágenes, sonidos y videos), sus respuestas siguen estando basadas en estos esquemas.

No obstante la naturalidad y aparente precisión de las salidas generadas por los LLM, hay varios aspectos en su entrenamiento que consideramos necesario enfatizar, con el fin de tener claros los límites de esta tecnología (al menos en su estado actual). El primer aspecto tiene que ver con el *corpus* de entrenamiento. Para que los LLM establezcan una red neuronal capaz de responder de esta forma, es necesario contar con muy grandes volúmenes de datos (en el rango de los miles de millones de unidades). Los equipos dedicados a este proceso suelen tomar contenidos disponibles de forma abierta en la Web, así como colecciones de libros digitalizados y otras fuentes similares (Silveira y Mauá, 2017; Gil y Doğru, 2019).

Desde una perspectiva técnica, esto implica la imposibilidad de curar manualmente la selección de unidades de entrenamiento y supervisar la

sentido desde una perspectiva cognitiva humana. Esta es la razón por la que la palabra aparecerá en *itálicas*.

¹² Es importante precisar que esta explicación no hace referencia a un espacio físico, sino a un sistema de coordenadas y vectores n-dimensionales en los que destacan *tokens* que sean estadísticamente relevantes en función del *prompt*.

confiabilidad de los contenidos. No sólo están ahí textos que han pasado por algún tipo de revisión, como trabajos académicos de acceso abierto y los artículos de la Wikipedia, sino también discusiones en Reddit y otros sitios similares sobre teorías conspiratorias, así como todo tipo de expresiones ajenas a la verdad y la dignidad humana. Incluso, recientes trabajos hablan de modelos de IAG que están siendo entrenados a partir de una creciente cantidad de contenidos generados por otros LLM, lo que está afectando la calidad de las redes neuronales resultantes (Shumailov *et al.*, 2024).

En lo que tiene que ver con una dimensión cultural, además de que es evidente de que este *corpus* no incluye toda la diversidad y variedad de pensamientos, ideas, tradiciones, ideologías, creencias y cosmovisiones del género humano; hay importantes brechas lingüísticas que es necesario señalar. Como ha sido señalado por Kew *et al.* (2023), la mayor parte de los *corpus* de entrenamiento empleados por empresas como Open AI, Google y Anthropic (todas de origen estadounidense) están dados por contenidos en inglés. Cuando se hace una solicitud a un LLM, éste responde de acuerdo con el idioma en que se haya hecho el *prompt*, recurriendo a las relaciones de sentido que en su red neuronal fueron establecidas a partir de su entrenamiento. En otras palabras, si la petición se hace en español, el modelo fundacional recurre a sus redes semánticas establecidas en *tokens* en este idioma, sin pasar por las más robustas relaciones establecidas en inglés.

Lo anterior implica que, cualquier respuesta que sea generada, tendrá una tendencia a ser más limitada que si hubiera sido hecha en inglés, dado que, recordemos, su precisión depende de la cantidad de unidades de entrenamiento. Así, la brecha digital que se manifiesta en una menor disponibilidad de contenidos digitales en idiomas distintos al inglés, impacta en la construcción de redes neuronales más endebles, con la consecuencia que explicaremos enseguida.

Con base en el procedimiento que ya hemos explicado, cuando el sentido resultante de un *prompt* determinado apunta a un *región* en la que no hay *tokens* semánticamente significativos, los algoritmos de la mayor parte de los actuales modelos fundacionales están programados para tomar *lo que sea* que resulte en el *token* con mayor valor relativo al *prompt*. En otras palabras, aunque se trate de una red de sentidos que podríamos calificar como *irrelevantes*, como sea el LLM empezará a generar una respuesta, *token* por

token, preservando una aparente lógica interna, aunque el resultado no corresponda con la verdad. A este fenómeno se le conoce como *alucinación* (Kaplan, 2024) y es, hasta el momento, un riesgo intrínseco a la IAG.

Esta es la razón por la que todos los LLM disponibles en la actualidad, con cierta frecuencia generan respuestas falsas, pero con una estructura argumentativa que puede ser altamente convincente. Se trata de un problema que resulta tanto de vacíos en el *corpus* de entrenamiento, como de la operación actual de sus algoritmos. Regresaremos a este punto cuando discutamos lo relativo al uso de estas tecnologías en los entornos educativos.

Por otro lado, además de las alucinaciones que eventualmente son generadas por los LLM, como Huang *et al.* (2023) y Dai *et al.* (2024) han reportado, otro efecto del *corpus* de entrenamiento son respuestas que hacen representaciones sesgadas de diversos grupos. A partir de la presencia de contenidos con este tipo de desviaciones, los modelos fundacionales han generado salidas estereotipadas en función de edad, género, orientación sexual, nacionalidad, religión, etnia, ideología, condición de discapacidad y otras características identitarias.

El segundo factor incidente en las salidas generadas por los LLM está en el ya mencionado proceso de RLHF. Para prevenir respuestas inapropiadas, los modelos fundacionales son dotados de una serie de filtros y salvaguardas que, desde una perspectiva basada en Latour (2011), pueden ser caracterizados como un conjunto de dispositivos sociotécnicos. Esto implica que, cualquiera que sea el ajuste que se haga sobre los algoritmos de salida, se trata de una conjunción de aspectos que no sólo son técnicos, sino también sociales. Los valores de las empresas que generan estos modelos fundacionales y hasta las ideologías y valoraciones personales de quienes hacen tales calibraciones, pueden incidir en la forma en que esto es hecho. Si, como sugieren Kahn y Ginther (2017), tales equipos están integrados principalmente por varones, blancos, relativamente jóvenes y cis-heterosexuales, eso determina experiencias de vida y formas muy concretas de interpretar el mundo, que tienen el potencial de transmitirse a dichas salvaguardas.

Como hemos planteado en este apartado, a pesar de la notable capacidad que tienen los LLM para generar respuestas con altos valores de aparente verosimilitud y coherencia interna, existen evidentes limitaciones a sus capacidades que se manifiestan en sesgos y alucinaciones, ambos fenómenos

derivados tanto de las características del *corpus* de entrenamiento como de los posibles efectos de los ajustes hechos a partir de los procesos de calibración hechos por humanos (RHLP). Antes de abordar la discusión sobre algunos de los impactos de la IAG en las prácticas docentes en la educación superior en México, en la siguiente sección haremos una revisión de algunos de los principales antecedentes relativos al uso de la IA en los espacios educativos.

Usos de la IAG en los entornos educativos

La educación es un campo de acción, investigación y reflexión que, en algunos sectores y regiones, ha incorporado diversas aplicaciones de IA desde hace varias décadas. En esta revisión de antecedentes, identificaremos dos conjuntos: primero, aquellos que son previos al lanzamiento de ChatGPT el 30 de noviembre de 2022 y, en segundo lugar, los que surgen a partir de esta disponibilidad de LLM.

Dentro del primer grupo, con base en una búsqueda de fuentes sobre IA y educación (AIED), Goksel y Bozkurt (2019) identifican tres ámbitos relacionados con su aplicación, posibilitados por el entonces emergente avance en los sistemas de *deep learning*: (1) el aprendizaje personalizado adaptativo,¹³ (2) los sistemas expertos y la tutoría inteligente y (3) la IA como un futuro componente del proceso educativo. Estos autores destacan el desarrollo de asistentes digitales como Siri y Cortana, que contaban con aplicaciones equivalentes para los espacios educativos, llamados *agentes pedagógicos*. Básicamente, se trataba de interfases con reconocimiento de voz, capaces de realizar búsquedas en bases de datos y otros repositorios de conocimiento, como los sistemas expertos.¹⁴

Por su parte, Ouyang y Jiao (2021) plantean la existencia de tres paradigmas relacionados con los encuadres y modelos pedagógicos que ha habido

¹³ Estos sistemas buscan adaptarse a los intereses y necesidades particulares de cada persona, en lugar de proveer un modelo general al que deban ajustarse.

¹⁴ Disponibles a partir de la década de 1970, se trata de uno de los primeras estrategias relativamente exitosas de IA que estuvieron disponibles. Se basan en árboles de decisiones que, a partir de una base de datos, permitían llegar a aspectos muy precisos sobre algún campo del conocimiento.

en torno a la aplicación de la IA en la educación: (1) IA dominante, estudiante como receptor, (2) AI como apoyo, estudiante como colaborador y (3) IA como potenciador,¹⁵ estudiante como líder:

Tabla 1. Tres paradigmas en la AIED

Paradigma	Marco teórico	Aplicaciones	Técnicas de IA
1	Conductismo	Aplicaciones tempranas de Sistemas Inteligentes de Tutoría	IA basada en relaciones estadísticas
2	Cognitivismo, Constructivismo social	Sistemas de Tutoría Basados en Diálogo Entornos Exploratorios de Aprendizaje	Redes bayesianas, procesamiento de lenguaje natural, árboles de decisión de Markov
3	Conectivismo, Sistemas complejos adaptativos	Cooperación humano-computadora: Aprendizaje personalizado / adaptado	Interfases cerebro-computadora, machine learning, deep learning

Fuente: Ouyang y Jiao (2021).

Estos autores refieren un progresivo incremento en la agencia de quienes se encuentran en los procesos educativos, así como en las posibilidades de los sistemas de IA para adaptarse y ajustarse a sus necesidades. Una de las principales conclusiones de este trabajo es que la aplicación de la IA en la educación no es un asunto exclusivamente tecnológico, sino que debe haber una integración de las dimensiones pedagógicas, sociales, culturales y económicas, en todo proceso de incorporación de estas tecnologías.

Por otro lado, Chen, Chen y Lin (2020) hacen otra revisión de fuentes en el cruce de la IA con la educación, a partir de tres ejes: el administrativo, la enseñanza y el aprendizaje. Estos autores sugieren que los usos de la IA en este ámbito inician con la introducción de los primeros sistemas de cómputo en las escuelas, pasando al posterior desarrollo de sistemas inteligentes en línea, como *chatbots* y otras aplicaciones capaces de ejecutar algunas de las funciones llevadas a cabo tradicionalmente por el personal docente, en entornos controlados conocidos como *plataformas*,¹⁶ que podían ser adaptados a las necesidades particulares del estudiantado.

De esta manera, a los sistemas de reconocimiento de voz y la instrucción personalizada señalados por Goksel y Bozkurt (2019), Chen, Chen y Lin

¹⁵ AI-empowered, en el original.

¹⁶ Algunos ejemplos de estas plataformas son *Blackboard* y *Moodle*.

(2020) agregan sistemas de realidad virtual y, en lo administrativo, el análisis de datos y los sistemas de reconocimiento visual para la evaluación de trabajos. Desde una metodología similar, el trabajo de Chen, Zou, Xie, Cheng y Liu (2022) permite agregar a esta lista aplicaciones de la IA en la educación como el análisis discursivo para el aprendizaje colaborativo, las redes neuronales para la evaluación docente y los sistemas de reconocimiento de estados emocionales.

Además de lo anterior, las reflexiones de estos mismos autores sugieren que, ante cualquier uso de la IA en la educación, es necesario tener transparencia en relación con los datos del estudiantado que se están empleando para la personalización de su aprendizaje. Con el fin de lograr una mayor aceptación del personal docente —agregan—, es recomendable involucrarlos en el diseño y desarrollo de toda aplicación de IA en sus procesos de enseñanza. Esto, además, conduce a sistemas mucho más robustos y apoyados en la experiencia.

Una crítica que es posible hacer a las aproximaciones bibliométricas, presentadas por Goksel y Bozkurt (2019) y Chen, Chen y Lin (2020), es que en sus análisis incluyen algunas tecnologías y recursos que en ocasiones sólo están vagamente relacionadas con auténticos sistemas de IA, como los campus virtuales, la incorporación de dispositivos móviles y otras aplicaciones similares. Este es un sesgo que se deriva de las palabras clave presentes en el corpus que observaron, asunto más bien relacionado con las editoriales y las personas responsables de su autoría. En todo caso, los desarrollos más relevantes de la época en este sentido que sí se encontraban dentro de la IA, consistían básicamente en aplicaciones orientadas a la gestión escolar (análisis de datos), y el apoyo en los procesos de interacción entre el personal docente y el estudiantado (sistemas expertos, *chatbots*, sistemas tutores, etc.).

El segundo conjunto de trabajos que hemos revisado ya refiere el uso de los LLM. Se trata de investigaciones publicadas a partir de 2023, en las que es posible identificar aspectos específicos asociados a estas innovaciones que, disponibles previamente, se refuerzan a partir de la aparición del ChatGPT y otros sistemas similares, como la tutoría automatizada y la personalización de la atención al estudiantado.

Las referencias que hemos localizado desde esta perspectiva pueden ser agrupadas en tres conjuntos: (1) las que aluden a las percepciones sobre el

impacto de los LLM en la enseñanza y el aprendizaje, (2) consideraciones éticas y el uso responsable de la IAG (3) sus aplicaciones pedagógicas y el diseño instruccional, especialmente en el desarrollo de agentes de AIEd.

En el primer grupo, se exploran los efectos de la IA generativa a partir de herramientas como ChatGPT en el aprendizaje, las prácticas docentes, los métodos de evaluación y los paradigmas educativos. Desde esta perspectiva, destaca el trabajo de Lan y Chen (2024), quienes identifican una serie de retos en relación con la IAG, a partir de percepciones de estudiantes y profesores:

Tabla 2. Principales retos frente a la IAG: percepciones

Estudiantes	Personal docente
Uso de la IAG como si fuera un motor de búsqueda avanzado	Temores de que el estudiantado carezca de competencias críticas para el uso de la IAG
Abuso de la IAG para la realización de tareas que requieren altas capacidades cognitivas	Preocupaciones sobre la autoría de trabajos
Subutilización de la IAG para tareas complejas (exploración, análisis, pensamiento complejo)	Falta de familiaridad con los potenciales educativos de la IA
Falta de habilidades para convertir la IAG en un recurso de apoyo	Falta de equilibrio entre el uso de la IA para aplicaciones educativas y la preservación del sentido humano

Fuente: elaboración propia a partir de Lan y Chen (2024).

Frente a esta situación, los autores sugieren el desarrollo de modelos colaborativos en los que el personal docente humano y la AIEd trabajen juntos. En este esquema, los humanos aportarían experiencia pedagógica, conocimiento de la materia e inteligencia emocional, mientras que la IA ofrecería procesamiento de información, sugerencias personalizadas de aprendizaje y retroalimentación basada en datos.

En una aproximación similar, Grande, Kiesler y Francisco (2024) coinciden en señalar que el personal docente tiene la percepción de que el estudiantado requiere de mayores competencias en el uso de los LLM, de forma que estos recursos sean vistos más como herramientas de apoyo, que como una base de datos.

Las percepciones del personal docente en torno a la IAG habían sido previamente estudiadas por Kaplan-Rakowski *et al.* (2023). En sus resultados destaca que mientras exista una mayor familiaridad y experiencia en el uso de la IAG, es más probable que sea aplicada en su práctica educativa.

Algunos de los principales motivos expresados para usarla, se relacionan con la idea de que su trabajo será más eficaz, permiten ahorrar tiempo y modificarían la forma en que se llevan a cabo las evaluaciones.

Por otro lado, en este mismo trabajo se enfatiza que algunas de las destrezas relacionadas con estos dispositivos tienen que ver con la capacidad para adaptar recursos de aprendizaje digital, altas competencias de comunicación, una personalidad flexible, apertura de mente, habilidades para gestionar la ansiedad que la IA pudiera generar y la disposición para aprender el uso de tecnologías emergentes.

Igual que en el trabajo de Lan y Chen (2024), en este trabajo se destaca la preocupación de que el estudiantado deje de lado la aplicación de destrezas académicas tradicionales, como la búsqueda de información, o que puedan asumir una actitud de indolencia o pereza cuando empleen Chat-GPT para elaborar sus trabajos escolares, o bien, dificultar el desarrollo del pensamiento crítico. Preocupaciones muy similares fueron encontradas también por Lee *et al.* (2024). En lo relativo al aspecto laboral, Kaplan-Rakowski *et al.* (2023) identifican que cerca de la tercera parte de los docentes participantes en su estudio, expresan el temor de ser reemplazados por sistemas de IA.

Por su parte, Lee *et al.* (2024) encuentran que hay muy diversas interpretaciones alrededor de la IAG de parte de docentes de educación superior. La mayor parte de quienes participaron en su estudio dicen que los principales impactos tienen que ver con cambios en las formas de evaluar, derivados del uso de la IA. Sólo alrededor del 25% dicen considerar que sus universidades están adecuadamente equipadas para la IA y el 75% de ellos indican que les gustaría tener apoyo en este sentido. Uno de los principales retos encontrados por estos autores, es mantener la correspondencia entre la formación que se brinda en la educación superior y las demandas del mercado laboral sobre la IA.

Como es posible observar, estas aproximaciones revelan algunas inquietudes en torno al uso ético y responsable de la IA en la educación. Este es un aspecto que detallaremos en el siguiente grupo de trabajos, en los que podemos identificar las siguientes preocupaciones: el uso acrítico e irresponsable de los LLM, la deshonestidad académica y la privacidad del estudiantado que emplea estas tecnologías.

En torno a este primer asunto, una preocupación reiterada es la posibilidad de que quienes están en procesos formativos dependan excesivamente de los LLM, lo que puede obstaculizar el desarrollo de su pensamiento crítico, la escritura creativa y las habilidades de resolución de problemas. De forma similar, pueden tratar a los LLM como motores de búsqueda avanzados¹⁷ para obtener respuestas rápidas, descuidando procesos de aprendizaje más profundos (Kaplan-Rakowski *et al.*, 2023; Grande *et al.*, 2024). Además, el contenido generado por LLM, aunque puede ser gramaticalmente correcto, a menudo carece de originalidad y profundidad, lo que potencialmente obstaculiza su capacidad para desarrollar su propia voz creativa (Bozkurt *et al.*, 2023).

Por otro lado, la virtual imposibilidad para discernir el texto generado por IA del trabajo de las y los estudiantes plantea un desafío significativo para la integridad académica. Los LLM pueden producir resúmenes, ensayos, composiciones y otros productos académicos aparentemente convincentes, lo que dificulta que los educadores evalúen los aprendizajes logrados y aborden el posible plagio (Orenstrakh *et al.*, 2023). Esto, sostienen Lee *et al.* (2024), requiere el desarrollo de métodos de detección robustos y estrategias de evaluación que permitan garantizar una evaluación justa.

En lo que tiene que ver con los sesgos algorítmicos planteados al inicio de este trabajo, como ya fue explicado, estos modelos se entrenan en grandes conjuntos de datos que reflejan sesgos sociales, que pueden perpetuar estereotipos dañinos y prácticas discriminatorias dentro de la educación (Moore *et al.*, 2024). Según Bozkurt *et al.* (2023), garantizar un uso equitativo e inclusivo de los LLM requeriría un examen cuidadoso de los datos de entrenamiento, incorporando perspectivas diversas y mitigando los sesgos para evitar el refuerzo involuntario de las desigualdades existentes. Estas son acciones que, evidentemente, rebasan el alcance de quienes emplean estos dispositivos en los espacios educativos y que, como explicamos en el apartado previo, además son humanamente imposibles de aplicar, dado el inmenso tamaño de los *corpus* de entrenamiento.

¹⁷ Es importante notar que el pasado 1 de noviembre de 2024, OpenAI anunció la disponibilidad de un motor de búsqueda en la Web asociado a ChatGPT 4o. Características similares ya estaban incorporadas al buscador de Google, con base en Gemini. Por lo tanto, la separación entre los *chatbots* de IA, y su capacidad de hacer búsquedas en la WWW, es cada vez menor.

En relación con la privacidad de las personas usuarias de los LLM, es que una de sus características en los modelos actualmente disponibles tiene que ver con el procesamiento remoto de los *prompts*. Como se señala en sus *términos de uso*, esto no garantiza la privacidad de los datos que son empleados en dichas peticiones, ni los relativos a sus personas usuarias. Así, el uso de los LLM en la educación a menudo implica la recopilación de datos personales, lo que plantea preguntas sobre cómo se almacena, se accede y se utiliza esta información (Lee *et al.*, 2024). Bozkurt *et al.* (2023) señalan que es esencial contar con directrices claras y marcos sólidos de gobernanza de datos para salvaguardar la privacidad de quienes se están formando y garantizar un uso responsable de sus datos con fines educativos.

En lo relativo al tercer conjunto de referencias, que tienen que ver con los sistemas de tutoría y agentes de AIEd como una aplicación pedagógica de los LLM, las consideraciones de Moore *et al.* (2023) ubican a estas tecnologías como una mediación que tiene el potencial de aumentar la agencia del estudiantado en su propio proceso formativo, a través de modelos colaborativos etiquetados bajo el nombre de *learnersourcing*. Desde esta perspectiva, la IAEd se plantea como una herramienta de apoyo para la co-creación de contenidos desarrollados desde la perspectiva de los educandos, quienes reciben una retroalimentación inmediata de parte de los LLM.

En esta misma línea, un objetivo destacado en las aplicaciones pedagógicas de la IAG tiene que ver con el desarrollo de agentes capaces de cristalizar de forma mucho más rápida y eficiente, la atención personalizada que habíamos visto en la primera etapa de la aplicación de la IA en la educación. En Lan y Chen (2024) se plantean cuatro características para dichos agentes:

1. Aprendizaje guiado de teorías y principios
2. Rutas de aprendizaje adaptativas y personalizadas
3. Evaluación y retroalimentación en tiempo real
4. Reforzamiento de lineamientos y control de calidad

Desde estos autores, el proceso de aprendizaje con un agente de IA pedagógico comienza cuando el profesor humano identifica los desafíos del estudiantado y selecciona las teorías o enfoques adecuados para diseñar tanto el procedimiento de enseñanza como las tareas específicas. El docente

evalúa el agente de IA actuando como un estudiante simulado, asegurándose de su precisión antes de implementarlo. Una vez validado, el agente es compartido con los y las estudiantes, quienes siguen sus orientaciones mientras realizan las tareas de aprendizaje. Durante este proceso, el agente de IA supervisa su progreso, evalúa resultados y proporciona retroalimentación en tiempo real, permitiendo que el alumnado corrija sus errores. Al completar con éxito las tareas, los y las estudiantes comparten sus registros de aprendizaje con el profesor, quien revisa su progreso.

Otra propuesta en este mismo sentido, parte de Zhang *et al.* (2024), quienes han desarrollado un modelo al que llaman *SimClass*, que da un paso adelante a lo planteado por Lan y Chen (2024), al aplicar el Sistema de Análisis de Interacción de Flanders¹⁸ (FIAS, por sus siglas en inglés) para su evaluación. Los principales resultados de este trabajo sugieren que el uso de este recurso tiene el potencial de mejorar las dinámicas de interacción al interior de una asignatura, incrementando tanto la presencia social como cognitiva de quienes participan (estudiantes y docentes), logrando con ello discusiones más profundas.

Una tercera propuesta de agente de IAEd se describe en el trabajo de Modran *et al.* (2024). Esta investigación propone un sistema de tutoría inteligente con *chatbot*, que integra el enfoque de Generación Aumentada por Recuperación (RAG, por sus siglas en inglés) con un LLM personalizado. El sistema desarrollado tiene como objetivo superar las limitaciones de la tutoría tradicional y los LLM de propósito general, al brindar asistencia precisa, contextualmente relevante y personalizada, mejorando así la comprensión y el compromiso. El sistema, impulsado por un agente inteligente, recupera información relevante de fuentes académicas seleccionadas, incorpora funciones interactivas para la retroalimentación del usuario y utiliza algoritmos de aprendizaje automático para la mejora continua del rendimiento, lo que garantiza una experiencia de tutoría sólida y efectiva.

¹⁸ El análisis de interacción es una técnica para captar las dimensiones cuantitativas y cualitativas del comportamiento verbal del docente en el aula. Como sistema de observación, capta el comportamiento verbal de docentes y estudiantes que están directamente relacionado con el clima socioemocional del aula. Fue desarrollado por Ned Flanders, a partir de la teoría de la psicología social y fue diseñado para probar el efecto del clima socioemocional en las actitudes y el aprendizaje de los estudiantes (Amatari, 2015).

En Iberoamérica también es posible encontrar trabajos en torno a estos mismos ejes. En general, las inquietudes, posibilidades y retos de la IAG son muy similares a lo ya referido (Garrido y Ruiz, 2024). Por ejemplo, se habla de la actualización docente y la capacitación en materia de IA (Rosignoli, 2024; Becerra *et al.*, 2024; Cotino-Arbelo *et al.*, 2024), el pensamiento crítico (Del Potro y Virgilio, 2024), el desarrollo de *chatbots* y asistentes virtuales (Nacimiento-García *et al.*, 2024; Sturla *et al.*, 2024), entre otros asuntos que en plena emergencia del fenómeno han sido abordados.

Brechas y retos en el contexto de los LLM en México e Iberoamérica

A diferencia de lo revisado en la primera parte de los antecedentes, el uso de la IA en la educación en México y América Latina, previo a la actual disponibilidad de los LLM, había sido más bien limitado. Salvo algunas experiencias en torno al uso de sistemas expertos y de entornos virtuales inmersivos, dadas las limitaciones presupuestales en muchas instituciones educativas de la región, se trata de una tecnología de la cual se hablaba con cierta frecuencia en torno a sus posibilidades, pero de la cual había pocas experiencias en su aplicación empírica (Jara y Ochoa, 2020).

A partir del 30 de noviembre de 2022, ChatGPT irrumpió como una tecnología capaz de, en apariencia, resolver una gran variedad de trabajos escolares y prácticas docentes. Esto ha despertado en la comunidad docente latinoamericana las mismas preocupaciones ya descritas en torno a su uso, especialmente en lo relativo a su empleo ético y responsable (Alonso-Martínez *et al.*, 2024).

Con base en la experiencia en el uso de los LLM, a partir de 2023 fueron propuestos diversos modelos para el uso de estos dispositivos en los espacios educativos desde una perspectiva iberoamericana (Codina, 2023). En términos generales, estas propuestas encuadran a los LLM como una herramienta que puede apoyar el desarrollo de conocimiento ampliado y fundamentado, en tanto se cumplan una serie de condiciones entre las cuales destacan la verificación de la información proporcionada por la IAG y la fundamentación en fuentes académicas. Se trata de aplicar el pensamiento

crítico ante posibles respuestas sesgadas o francas alucinaciones generadas por los modelos fundacionales.

Tomando como referente la explicación de los procesos de entrenamiento y operación de los LLM, en el contexto hispanoparlante, esta tecnología implica un conjunto de retos que se suman a tales desviaciones y respuestas que no corresponden con la realidad. En este ámbito, tanto la brecha digital (Pérez-Salazar, 2018) como la del idioma inciden en la forma en que se emplea esta tecnología.

Como ya hemos adelantado, la red neuronal resultante del proceso de aprendizaje profundo se construye a partir de los *tokens*, que son las unidades de sentido de estos sistemas de lenguaje natural. Recordemos que aunque puede haber diversas posibilidades, tales unidades suelen corresponder con palabras individuales que se presentes en el *corpus* del aprendizaje.

A partir de Guo *et al.* (2024) y Zhong *et al.* (2024), se sabe que la mayor parte de las unidades de entrenamiento empleadas en los actuales LLM se encuentran en inglés. Cualquier *prompt* que sea hecho en un idioma distinto, como ya se ha señalado, dará lugar a un proceso de generación textual que se apoyará en una red neuronal mucho menos sólida. Las salidas, por lo tanto, tenderán a ser menos profundas y con una mayor propensión a generar alucinaciones.

Por lo anterior, una estrategia que parecería salvar esta situación sería escribir los *prompts* en inglés siempre que sea posible, ya sea de forma directa, cuando se cuenten con las competencias para ello, o bien con el apoyo de otros sistemas de IA, como los traductores automáticos.¹⁹ Sin embargo, esto no sería suficiente. La relativamente menor cantidad de documentos en español disponibles para ser empleados como parte del *corpus* de entrenamiento de los LLM, implica también una brecha cultural que es necesario considerar en los entornos educativos.

Como plantean Calvas *et al.* (2019), lo local tiene una gran importancia en los procesos de discusión y generación del conocimiento alrededor de algunas asignaturas, entre las que destacan las pertenecientes a las Ciencias

¹⁹ Estos traductores son la aplicación de otros modelos neuronales que suelen ser agrupados bajo el término de *Neural Machine Translation* (Stahlberg, 2020).

Sociales. Se trata de saberes, tradiciones y elementos culturales que, pese a contribuir de forma notable a la identidad colectiva de muchas comunidades, no siempre se encuentran disponibles en formatos digitales. Por lo tanto, muchos de estos contenidos no forman parte de las redes neuronales que resultan de los procesos de entrenamiento de los LLM. Cualquier *prompt* que se haga en relación con dichas unidades tiene el potencial de generar salidas que no reflejen estos acervos.

Se trata de sesgos culturales que han sido abordados por autores como Mukherjee *et al.* (2023) y Tao *et al.* (2024). En estos trabajos se destaca la forma en que los LLM se incrustan valores pertenecientes a regiones culturales caracterizadas por el dominio del inglés y el protestantismo cristiano centroeuropeo.

Como ya hemos sustentado en Latour (2010), reiteramos que los modelos fundacionales pueden ser descritos como dispositivos sociotécnicos que no son de modo alguno neutrales, sino que incorporan dentro de sí aspectos que corresponden a los contextos en que estas tecnologías son producidas. El origen estadounidense de la mayor parte de los LLM incide de forma muy relevante en el uso que se pueda hacer de ellos en México e Iberoamérica.

Consideraciones finales

La revisión de los antecedentes nos ha permitido identificar algunas preocupaciones constantes entre el personal docente en relación con el uso de los LLM. Además del temor a ser reemplazados por instructores digitales, quizá una de las más reiteradas es su empleo para la elaboración de trabajos escritos que suelen ser requeridos como medio de evaluación.

Sin embargo, como ha afirmado Frederick M. Hess, Director de Políticas Educativas del American Enterprise Institute quizás “el problema no es la tecnología, sino el tipo de tareas que pedimos a nuestros estudiantes” (Hess, 2023). Más que a partir del establecimiento de infructuosas políticas restrictivas y persecutorias, quizá el asunto pueda ser encuadrado de otra manera.

La aparición de los LLM implica una profunda reflexión de lo que hemos estado haciendo en nuestras asignaturas y direcciones de tesis. Ciertamente,

es indispensable el desarrollo de un pensamiento crítico, no sólo hacia la IAG, sino incluso desde estos mismos dispositivos, como lo sugieren Costello, Pennycook y Rand (2024). Sin duda, el modelo de Codina (2023) resulta de la mayor pertinencia, al requerir una contrastación de fuentes y validación de las salidas generadas por los modelos fundacionales. A pesar de ello, consideramos que es necesaria una reflexión más profunda en torno a la práctica docente.

En primer lugar, como ha sido planteado por Ouyang y Jiao (2021), el uso de la IA no es sólo tecnológico, implica integrar las dimensiones pedagógicas, sociales, culturales y económicas. Como toda herramienta pedagógica, su uso requiere de una estrategia que, apoyada en un análisis de sus alcances y limitaciones actuales, permita diseñar aplicaciones relevantes y adecuadas al contexto en que ocurra cada experiencia docente. Así como prohibir su uso es un esfuerzo vano, obligar a que esta tecnología sea empleada, sin una planeación adecuada y sin analizar las condiciones de cada entorno, sería inútil y muy probablemente, ineficaz.

Como lo sugieren Chen *et al.* (2022), cualquier esfuerzo que implique la introducción de la IA en los espacios educativos, requiere de la participación formada y reflexiva de todos los agentes involucrados: estudiantes, personal docente y administrativo; involucrándose de forma colaborativa en cada etapa de su desarrollo y aplicación. En términos de Lan y Chen (2024), se trata de que el personal humano y la IAG trabajen de manera conjunta, potenciando con ello la capacidad de agencia de los primeros.

Evidentemente, como sugieren Kaplan-Rakowski *et al.* (2023), esto demanda perfiles muy precisos en todos los elementos humanos participantes en este sistema, que sean consistentes con el desarrollo y fortalecimiento de la capacidad para adaptar recursos de aprendizaje digital, destacadas competencias comunicativas, una personalidad flexible, una mentalidad abierta, habilidades para manejar la ansiedad que pueda surgir ante la inteligencia artificial (o cualquier otra tecnología) y una actitud dispuesta a aprender el uso de cualquier innovación tecnológica que pueda apoyar en los procesos educativos.

Desde esta perspectiva, cualquier modelo vertical que se sustente en el poder y la autoridad y no en la colaboración y la construcción de habilidades, como frecuentemente ha ocurrido, estará muy probablemente con-

denado al fracaso. Se trata de evitar los errores que en el pasado se han cometido en la introducción de nuevos dispositivos en los espacios de aprendizaje, de superar la fetichización y el determinismo tecnológico de iniciativas como *Enciclomedia*.²⁰ Si algo hemos aprendido en las alrededor de tres décadas que han transcurrido desde el surgimiento de las promesas de la llamada sociedad de la información es que, por sí mismo, lo virtual no llevó automáticamente a mayores niveles educativos ni de desarrollo. La IA tampoco lo hará. Es necesario acompañar todo esfuerzo de lineamientos y acuerdos que, tomando en cuenta las realidades de cada situación, conduzcan a las mejores adaptaciones y aplicaciones en cada salón de clases, en cada escuela de cada nivel.

Tales condiciones únicas, por otro lado, deben considerar las limitaciones de tipo cultural que hemos planteado a lo largo de este trabajo. Los LLM no lo *saben* todo. Lejos de ello, hay enormes porciones del conocimiento humano que, en buena medida a consecuencia de la brecha digital, han permanecido al margen de sus redes neuronales. Es fundamental no perder de vista que aquellos valores que pueden ser esenciales en lo local, podrían no estar del todo presentes en estos dispositivos. En tales casos, la intervención humana debe suplir las limitaciones de la IAG, ya sea desde sus propios algoritmos (por ejemplo, apoyándose en los modelos de generación aumentada por recuperación [RAG]) o a través de cualquier otra estrategia que sea relevante y pertinente en cada situación.

En segundo lugar, se impone una reflexión sobre muchos de los actuales métodos de evaluación que pudieran estarse aplicando. En lugar de estrategias unidimensionales como la memorización y el recuerdo literal centrado en datos que difícilmente pueden ser aplicados a otras situaciones, o la solicitud de resúmenes de lectura, que pueden ser rápidamente resueltos con el uso de los LMM sin que ello demande una auténtica presencia cognitiva al estudiantado; sería necesario considerar y, en su caso, capacitarse y adoptar otras formas de evaluación.

²⁰ Iniciativa impulsada durante la administración del presidente Vicente Fox, que consistió en el equipamiento de aulas de 5º y 6º grado de educación básica con equipo de cómputo. Según Navarro (2011), este programa careció de un análisis de alternativas que considerara los aspectos pedagógicos, los de equidad, la complejidad operativa y la limitada disponibilidad de recursos públicos.

Entre una infinidad de posibilidades, algunos de estos modelos podrían estar sustentados en las propuestas de Garrison, Anderson y Archer (2000) alrededor de las comunidades de investigación,²¹ donde se habla de la participación activa de los y las estudiantes en procesos de indagación y pensamiento crítico, facilitando la comprensión profunda y la aplicación del conocimiento. Este planteamiento implica un ciclo de cuatro fases: (1) un evento desencadenante que inicia la reflexión, (2) la exploración del problema, (3) la integración de ideas y (4) la resolución, donde se construye y aplica el significado.

Como es evidente, este proceso no excluye de ninguna manera el uso de los LLM, pero sí debería ocurrir desde una perspectiva en la que, bajo una supervisión docente muy cercana (idealmente, en el mismo salón de clases o sesión virtual), quienes se encuentran en estos espacios reciban una retroalimentación y guía constante, que les oriente en la construcción del conocimiento, e incluso, les ayude a enmarcar y facilitar el empleo de la IAG de manera crítica, validando la precisión de sus respuestas y ampliándolas tanto con otras fuentes como con la reflexión colectiva.

Todo sugiere que la IAG tiene un enorme potencial transformador en una gran cantidad de situaciones y contextos, incluyendo los educativos. Existe una altísima probabilidad de que esté siendo empleada por estudiantes de todos los niveles para resolver todo tipo de trabajos y asignaturas. El reto desde lo educativo no es de ninguna manera simple, pero sí, casi con total certeza, ineludible.

Referencias

- Aïmeur, E., Amri, S., y Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1), 30. <https://doi.org/10.1007/s13278-023-01028-5>.
- Alonso-Martínez, H., Jarquín-Ramírez, M., y Díez-Gutiérrez, E. J. (2024). Cuestiones éticas y políticas de ChatGPT en educación. Una revisión sistemática. En V. Parra, A. Corica, P. Sureda, S. Schiaffino y D. Godoy (eds.), *Actas del Primer Congreso Internacional en Inteligencia Artificial y Educación*, pp. 246-254.
- Amatari, V. O. (2015). The instructional process: a review of Flanders' interaction analy-

²¹ *Communities of Inquiry*, en el original.

- sis in a classroom setting. *International Journal of Secondary Education*, 3(5), 43-49. <https://doi.org/10.11648/j.ijsedu.20150305.11>
- Becerra, G., Grinsztajn, F., Zeliz, J. G., y Estayno, M. (2024). Intersección entre la inteligencia artificial, la educación universitaria y la formación docente. En V. Parra, A. Corica, P. Sureda, S. Schiaffino y D. Godoy (eds.), *Actas del Primer Congreso Internacional en Inteligencia Artificial y Educación* (pp. 15-25).
- Benassini, C. (2022). El discurso de odio en Facebook visto desde la epistemología de la complejidad. En O. Fragoso, M. T. Olalde y G. Garduño (Coords.), *Tejiendo diálogos. Reflexiones contemporáneas sobre la expresión y el sentido* (pp. 406-429). Escuela Nacional de Antropología e Historia; Casa Editorial Analéctica.
- Bozkurt, A. et al. (2023). Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from the educational landscape. *Asian Journal of Distance Education*, 18(1), 53-130. <https://doi.org/10.5281/zenodo.7636568>
- Calvas, M. G., Espinoza, E. E., y Herrera, L. (2019). Fundamentos del estudio de la historia local en las ciencias sociales y su importancia para la educación ciudadana. *Conrado*, 15(70), 193-202. <https://conrado.ucf.edu.cu/index.php/conrado/article/view/1124>
- Chen, L., Chen, P., y Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264-75278. <https://ieeexplore.ieee.org/abstract/document/9069875>
- Chen, X., Zou, D., Xie, H., Cheng, G., y Liu, C. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, 25(1), 28-47. <https://www.jstor.org/stable/48647028>
- Codina, L. (2023). *Cómo utilizar ChatGPT en el aula con perspectiva ética y pensamiento crítico: una proposición para docentes y educadores*. <https://www.lluiscodina.com/chatgpt-educadores/>
- Costello, T. H., Pennycook, G., y Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714). <https://doi.org/10.1126/science.adq1814>
- Cotino-Arbelo, A. E., Nacimiento-García, E., Molina-Gil, J., y González-González, K. S. (2024). Shaping Tomorrow's AI Natives: Promoting AI Literacy in Early Childhood Education. En V. Parra, A. Corica, P. Sureda, S. Schiaffino y D. Godoy (eds.), *Actas del Primer Congreso Internacional en Inteligencia Artificial y Educación* (pp. 78-85).
- Dai, S., Xu, C., Xu, S., Pang, L., Dong, Z., y Xu, J. (2024). Bias and Unfairness in Information Retrieval Systems: New Challenges in the LLM Era. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 6437-6447. <https://dl.acm.org/doi/pdf/10.1145/3637528.3671458>
- Del Potro, M. A., y Virgilio, A. (2024). Trazando el futuro educativo en la formación inicial de docentes: Valorando el desarrollo del pensamiento crítico en el uso de la Inteligencia Artificial Generativa para el diseño de clases. En V. Parra, A. Corica, P. Sureda, S. Schiaffino y D. Godoy (eds.), *Actas del Primer Congreso Internacional en Inteligencia Artificial y Educación* (pp. 34-42).
- Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible way*. Springer.

- Garrido, L. O. A., y Ruiz, H. M. (2024). Perspectiva de estudiantes de nivel medio superior respecto al uso de la inteligencia artificial generativa en su aprendizaje. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 14(28). <https://doi.org/10.23913/ride.v14i28.1830>
- Garrison, D. R., Anderson, T., y Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2-3), 87-105. [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- Gil, X., y Doğru, G. (2019). *The Web as a corpus. A multilingual multipurpose corpus* [Tesis de maestría]. Universidad Autónoma de Barcelona. <https://ddd.uab.cat/record/210868>
- Goksel, N., y Bozkurt, A. (2019). Artificial Intelligence in education: current insights and future perspectives. En S. Sisman-Ugur y G. Kurubacak (eds.), *Handbook of research on learning in the age of transhumanism* (pp. 224-236). IGI Global.
- Grande, V., Kiesler, N., y Francisco, R., M. A. (2024). Student perspectives on using a Large Language Model (LLM) for an assignment on professional ethics. *Proceedings of the 2024 on Innovation and Technology in Computer Science Education*, 1, 478-484. <https://doi.org/10.1145/3649217.3653624>
- Guo, Y., Conia, S., Zhou, Z., Li, M., Potdar, S., y Xiao, H. (2024). *Do Large Language Models have an english accent? Evaluating and improving the naturalness of multilingual LLMs*. <https://doi.org/10.48550/arXiv.2410.15956>
- Hess, F. M. (2023). AI isn't the problem. It's how we use it, especially in schools. *Medium*. <https://rickhess99.medium.com/ai-isnt-the-problem-it-s-how-we-use-it-especially-in-schools-d396076cc8ff>
- Huang, D., Bu, Q., Zhang, J., Xie, X., Chen, J., y Cui, H. (2023). *Bias assessment and mitigation in LLM-based code generation*. <https://arxiv.org/pdf/2309.14345>
- Hunter, B., Hindocha, S., y Lee, R. W. (2022). The role of artificial intelligence in early cancer diagnosis. *Cancers*, 14(6), 1524. <https://doi.org/10.3390/cancers14061524>
- Jara, I., y Ochoa, J. M. (2020). *Usos y efectos de la inteligencia artificial en educación. Sector Social división educación*. BID. <http://dx.doi.org/10.18235/0002380>
- Javed, U., y Javed, U. (2023). The Influence of social media algorithms on political polarization and public opinion. *Online Media and Society*, 4(2), 44-52. <https://hnpublisher.com/ojs/index.php/OMS/article/view/35>
- Kahn, S., y Ginther, D. (2017). Women and STEM. *National Bureau of Economic Research*. <https://doi.org/10.3386/w23525>
- Kaplan, J. (2024). *Generative Artificial Intelligence: what everyone needs to know*. Oxford University Press.
- Kaplan-Rakowski, R., Grotewold, K., Hartwick, P., y Papin, K. (2023). Generative AI and teachers' perspectives on its implementation in education. *Journal of Interactive Learning Research*, 34(2), 313-338. <https://www.learntechlib.org/primary/p/222363/>
- Kew, T., Schottmann, F., y Sennrich, R. (2023). *Turning English-centric LLMs into polyglots: How Much Multilinguality Is Needed?* https://ui.adsabs.harvard.edu/link_gateway/2023arXiv231212683K/doi:10.48550/arXiv.2312.12683
- Koon, J. (2023, 27 de noviembre). *Autonomous vehicles: not ready yet*. <https://semiengineering.com/autonomous-vehicles-not-ready-yet/>

- Lan, Y. J., y Chen, N. S. (2024). Teachers' agency in the era of LLM and generative AI. *Educational Technology & Society*, 27(1), I-XVIII. <https://www.jstor.org/stable/48754837>
- Latour, B. (2011). Network theory networks, societies, spheres: Reflections of an actor-network theorist. *International Journal of Communication*, 5(15). <https://ijoc.org/index.php/ijoc/article/view/1094/558>
- Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Strelan, P., y Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Martín-Barbero, J. (1987). *De los medios a las mediaciones. Comunicación, cultura y hegemonía*. Gustavo Gili.
- Modran, H., Bogdan, I. C., Ursuțiu, D., Samoila, C., y Modran, P. L. (2024). LLM intelligent agent tutoring in higher education courses using a RAG approach. <https://doi.org/10.20944/preprints202407.0519.v1>
- Moore, S., Tong, R., Singh, A., Liu, Z., Hu, X., Lu, Y., Liang, J., Cao, C., Khosravi, H., Denny, P., Brooks, C., y Stamper, J. (2023). Empowering education with llms-the next-gen interface and content generation. *International Conference on Artificial Intelligence in Education*, 32-37. https://doi.org/10.1007/978-3-031-36336-8_4
- Mukherjee, A., Raj, C., Zhu, Z., y Anastasopoulos, A. (2023). Global voices, local biases: socio-cultural prejudices across languages. En H. Bouamor, J. Pino y K. Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp.15828-15845). Association for Computational Linguistics.
- Nacimiento-García, E., y Díaz-Kaas-Nielsen, H. S., Cotino-Arbelo, A. E. y González-González, C. S (2024). Puntal, the simplicity and power of a generative artificial intelligence based virtual assistant and its acceptance by preschool children. En V. Parra, A. Corica, P. Sureda, S. Schiaffino y D. Godoy (eds.), *Actas del Primer Congreso Internacional en Inteligencia Artificial y Educación* (pp. 96-102).
- Navarro, A. (2011). Formación de agenda en la transición del programa Enciclomedia: hacia habilidades digitales para todos. *Revista Mexicana de Investigación Educativa*, 16(50), 699-723.
- Orenstrakh, M. S., Karnalim, O., Suárez, C. A., y Liut, M. (2024, Julio). Detecting LLM-generated text in computing education: Comparative study for ChatGPT cases. *2024 IEEE 48th Annual Computers, Software, and Applications Conference*, 121-126. <https://doi.org/10.1109/COMPSAC61105.2024.00027>
- Ouyang, F., y Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 1-6. <https://doi.org/10.1016/j.caeai.2021.100020>
- Pajkovic, N. (2022). Algorithms and taste-making: Exposing the Netflix recommender system's operational logics. *Convergence*, 28(1), 214-235. <https://doi.org/10.1177/13548565211014464>.
- Pérez-Salazar, G. (2018). Capital informacional y sociedad de la información: reflexiones en torno a la brecha digital. En A. R. Alva de la Selva (coord.), *Monitorear la*

- Sociedad de la Información y el Conocimiento. Propuesta de indicadores cualitativos: El "Capital Informacional"* (pp. 59-74). UNAM.
- Rosignoli, S. (2024). Los dispositivos de formación de formadores para la Integración de la Inteligencia Artificial Generativa en la enseñanza universitaria. Avances de Tesis de Maestría en Tecnología Educativa. En V. Parra, A. Corica, P. Sureda, S. Schiaffino y D. Godoy (eds.), *Actas del Primer Congreso Internacional en Inteligencia Artificial y Educación* (pp. 6-14).
- Seaver, N. (2022). *Computing taste: Algorithms and the makers of music recommendation*. University of Chicago Press.
- Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., y Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755-759. <https://doi.org/10.1038/s41586-024-07566-y>
- Silveira, I. C., y Mauá, D. D. (2017). University entrance exam as a guiding test for artificial intelligence. *2017 Brazilian Conference on Intelligent Systems*, 426-431. <https://doi.org/10.1109/BRACIS.2017.44>
- Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69, 343-418. <https://doi.org/10.1613/jair.1.12007>
- Sturla, S., Pecori, A., Orellano, R., y Nobile, C. (2024). Proyecto ULlibot: un chatbot de Inteligencia Artificial en Moodle LMS para asistir en tiempo real a los usuarios. En V. Parra, A. Corica, P. Sureda, S. Schiaffino y D. Godoy (eds.), *Actas del Primer Congreso Internacional en Inteligencia Artificial y Educación* (pp. 165-173).
- Tang, X., Dai, H., Knight, E., Wu, F., Li, Y., Li, T., y Gerstein, M. (2024). A survey of generative AI for de novo drug design: new frontiers in molecule and protein generation. *Briefings in Bioinformatics*, 25(4). <https://doi.org/10.1093/bib/bbae338>
- Tao, Y., Viberg, O., Baker, R. S., y Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9). <https://doi.org/10.1093/pnas-nexus/pgae346>
- Tindall, B. A. (1976). Theory in the study of cultural transmission. *Annual Review of Anthropology*, 5, 195-208.
- Vaswani, A. et al. (2017, 4-9 de diciembre). *Attention is all you need* [ponencia]. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, Estados Unidos de Norteamérica.
- Zhang, Z., Zhang-Li, D., Yu, J., Gong, L., Zhou, J., Liu, Z., Hou, L., y Li, J. (2024). Simulating classroom education with LLM-empowered agents. <https://doi.org/10.48550/arXiv.2406.19226>
- Zhong, C., Cheng, F., Liu, Q., Jiang, J., Wan, Z., Chu, C., Murawaki, Y., Kurohashi, S. (2024). Beyond English-Centric LLMs: What Language Do Multilingual Language Models Think in? <https://doi.org/10.48550/arXiv.2408.10811>