

Desinformación en la IA: El riesgo oculto de la falta de cultura de uso en modelos de lenguaje grande en universitarios

Misinformation in AI: The Hidden Risk of the Lack of Culture of Use in Large Language Models Among University Students

DOI: <https://doi.org/10.61604/dl.v17i31.495>

Gibrán Aguilar Rangel¹

Universidad Autónoma de Querétaro, México

Correo: gibrán.aguilar@uag.mx,

ORCID: <https://orcid.org/0000-0002-5865-2875>



Recibido: 27 de agosto del 2025

Aceptado: 25 de octubre del 2025

Para citar este artículo: Aguilar, A. (2025). Desinformación en la IA: El riesgo oculto de la falta de cultura de uso en modelos de lenguaje grande en universitarios, *Diá-logos*, (31), 25-35. <https://doi.org/10.61604/dl.v17i31.495>

¹Licenciado en Contaduría Pública, Maestro en Gestión de la Tecnología y Doctor en Gestión Tecnológica e Innovación por la Universidad Autónoma de Querétaro. Especialista en tendencias tecnológicas y su rol en los sectores de Educación y Finanzas. Profesor de tiempo libre, Facultad de Contaduría y Administración de la Universidad Autónoma de Querétaro.



Nuestra revista publica bajo la Licencia Creative Commons: Atribución-No Comercial-Sin Derivar 4.0 Internacional

Resumen

Uno de los componentes más conocidos del hipérónimo Inteligencia Artificial (IA) son los lenguajes de modelo grande (o LLM por sus siglas en inglés), siendo uno de los más populares ChatGPT, el cual está siendo utilizado ampliamente por los estudiantes universitarios para diversas tareas, desde resumir artículos, resolver problemas e inclusive escribir trabajos completos. El problema más evidente es la imposibilidad de detectar con certeza este tipo de prácticas, sin embargo, un problema que pasa desapercibido es que las respuestas que generan este tipo de modelos no son (a la fecha) verificables y su confiabilidad es cuestionable. En el presente trabajo se presentará como primer punto una explicación breve de qué es la IA y cómo funcionan los LLM, posteriormente se analizaron los resultados de una encuesta de uso de ChatGPT, así como la información recabada de casos de uso, concluyendo con recomendaciones de cómo dar a conocer al estudiantado que es lo que hace realmente un LLM y cómo usarlo de manera correcta.

Palabras clave

Inteligencia artificial, LLM, desinformación, ChatGPT

Abstract

One of the most well-known components of the Artificial Intelligence (AI) hypernym is large model languages (or LLMs), one of the most popular being ChatGPT, which is being widely used by university students for various tasks, from summarizing articles, solving problems and even writing complete papers. The most obvious problem is the impossibility of detecting with certainty this type of practices, however, an overlooked problem is that the responses generated by this type of models are not (to date) verifiable and their reliability is questionable. In this paper, a brief explanation of what AI is and how LLMs work will be presented as a first point. Subsequently, the results of a survey on the use of ChatGPT will be analyzed, as well as the information collected from use cases, concluding with recommendations on how to make students aware of what an LLM really does and how to use it correctly.

Keywords

Artificial intelligence, LLM, misinformation, ChatGPT

Antecedentes

La idea base de la inteligencia artificial no es nueva, como concepto, una máquina o creación artificial que iguala o incluso supera el cerebro humano ha sido reflejado en diversas obras a lo largo de las décadas, en un inicio como una ciencia ficción, pero a partir del planteamiento de la máquina de Turing, llamada así por provenir del concepto de Alan Turing, se fue perfilando más como una meta a la cual se planeaba llegar. Con la introducción de las computadoras y la evolución acelerada que se produjo en su desarrollo, crece también este deseo de crear un sistema que pudiese ser llamado inteligente, prueba de esto fue la búsqueda de un programa capaz de vencer a jugadores humanos en ajedrez, específicamente campeones mundiales, considerado como uno de los deportes mentales por excelencia, lográndolo en 1996 con la supercomputadora Deep Blue, fabricada por IBM, la clave de esto fue el gran poder de procesamiento del equipo, que podía calcular una gran cantidad de movimientos posibles con el fin de elegir el movimiento ideal. Si bien este fue un logro impresionante, era un sistema que funcionaba dentro de un juego con reglas fijas y movimientos determinados, es decir, necesitaba un encuadre específico para funcionar.

En estas etapas iniciales, los sistemas se basaban en reglas y lógica simbólica, buscando emular el razonamiento humano con base en algoritmos predefinidos, el problema con este enfoque es que no servía para lidiar con problemas complejos o lenguaje ambiguo, un ejemplo sería los comandos de voz que requerían decir una frase específica para ser activados y cualquier desviación en el comando resultaría en una falla. El avance en esta técnica surge con la aplicación de modelos estadísticos a grandes volúmenes de datos con el llamado aprendizaje automático (conocido en inglés como machine learning) (Forero-Corba y Negre Bennasar, 2024) y su versión avanzada, el aprendizaje profundo (deep learning), esto sería la puerta de acceso a la visión por computadora, un reconocimiento avanzado de voz y el procesamiento de lenguaje natural.

Los modelos de lenguaje grande (Large Language Models o LLM) son una de las aplicaciones más avanzadas del modelo de aprendizaje profundo en el campo del procesamiento del lenguaje natural, y son lo que muchas personas relacionan cuando piensan en inteligencia artificial, al punto de usarlo como sinónimo. Su evolución está ligada al concepto de *transformer* (Vaswani et al., 2017), esta red arquitectónica redujo el tiempo de entrenamiento y mejoró la calidad de los resultados y fue la base para el modelo GPT (*generative pre-trained transformer*) que lanzó OpenAI. Esta arquitectura se diferencia de versiones previas, las cuales procesaban las palabras de un texto de forma secuencias, en que permite analizar todas las palabras en un texto de forma simultánea, logrando identificar relaciones y dependencias entre dichas palabras, lo cual mejora el entendimiento del contexto y permite generar mejores respuestas.

Un punto esencial para que un LLM sea exitoso tiene que ver con el entrenamiento, generalmente este se divide en dos fases, un pre-entrenamiento y un ajuste. La fase de pre-entrenamiento consiste en alimentar el modelo con grandes volúmenes de datos de texto, como libros, revistas, artículos, etc., con el objetivo de que el modelo pueda predecir la siguiente palabra en una secuencia de texto, por medio del análisis de patrones gramaticales y de contexto. En la fase de ajuste, el modelo que ya recibió un entrenamiento previo, recibe un segundo entrenamiento con set de datos más pequeños y más especializados sobre temas en específico, lo cual le permite realizar tareas especializadas (Mao et al., 2024), como los asistentes virtuales de ciertos sitios web que pueden resolver dudas de los usuarios en tiempo real (Yan et al., 2024).

Derivado del punto anterior, surgen varias desventajas, en la parte de consideraciones éticas, la mayor parte de los datos en los que son entrenados estos modelos, son tomados sin permiso, ni reconocimiento de los autores originales, esto resulta especialmente problemático cuando se busca lucrar con estos modelos, puesto que se está generando una ganancia derivada del "robo" de trabajo de miles de personas. Otro problema es el sesgo, un LLM usará como marco de referencia los datos en los que fue entrenado (Kajiwara y Kawabata, 2024), es decir que puede llegar a discriminar con base en esta falta de contexto, asimismo puesto que su conocimiento está limitado a patrones estadísticos, no tiene una comprensión del mundo real ni de los conceptos como tal. Otro punto a considerar es la parte de la desinformación, ya sea de manera intencional, con un tercero generando textos o imágenes falsas para engañar a un grupo de personas (Barman et al., 2024), o bien de manera involuntaria por los datos erróneos que debido a sus limitaciones un modelo pueda presentar como ciertos al usuario. Finalmente, un LLM consume una gran cantidad de energía y recursos computacionales (Jiang et al., 2024), dependiendo del tipo de modelo una consulta simple puede utilizar el equivalente a una botella de agua en energía, y mientras más se difunde su uso y se integra en diferentes productos, mayor será este consumo potencial.

En la parte educativa, se ha dado un gran enfoque al potencial de deshonestidad académica, las instituciones, especialmente las de educación superior, dan gran peso a evitar el plagio, existen programas especializados que ayudan a los docentes a detectar este tipo de situaciones, sin embargo frente a los LLM se encuentran limitados (Abdelaal et al., 2019), si bien ya se han dado avances en este respecto, con programas dedicados a intentar identificar texto generado por un LLM (Ihekweazu et al., 2023), estos aún no cuentan con una confiabilidad total que permita asegurar que un texto no fue creado por el estudiante. Existen asimismo proponentes que

consideran que este tipo de modelos pudiesen usarse para ayudar tanto a estudiantes como docentes en la redacción científica (Salvagno et al., 2023), mientras que los críticos argumentan que derivado de la falta de confiabilidad esto podría resultar contraproducente (Sun et al., 2024).

Objetivos

El objetivo general del trabajo es analizar los hábitos de uso de herramientas de inteligencia artificial (IA) específicamente de los LLM entre estudiantes universitarios, identificando la frecuencia de uso, las percepciones sobre su utilidad y los desafíos asociados, con el fin de comprender cómo estas tecnologías pueden potencialmente impactar en el ámbito académico y personal.

Lo que se busca es no solo describir el estado actual del uso de LLM entre estudiantes universitarios, sino también comprender las implicaciones de este fenómeno en la educación superior, al identificar hábitos, expectativas y percepciones de los estudiantes, el presente trabajo busca contribuir al debate sobre el papel de la IA en la educación, proporcionar datos para instituciones educativas que busquen adaptarse a nuevas tecnologías, así como fomentar una reflexión crítica sobre los desafíos éticos y prácticos asociados con el uso de IA en el ámbito académico.

Metodología

La presente investigación se trata de una investigación cualitativa, de corte descriptivo, puesto que se pretende generalizar con base en los resultados obtenidos, se utilizó un muestreo de avalancha (Martin-Crespo y Salamanca, 2007). Para la población objetivo de la encuesta, se tomó a los estudiantes de la Facultad de Contaduría y Administración de la Universidad Autónoma de Querétaro, esto por ser una población a la que se tenía acceso para encuestar y por tratarse de la facultad con mayor número de estudiantes de la universidad.

La encuesta fue anónima para permitir la libre expresión de los estudiantes, se aplicó a estudiantes de primero a octavo semestre cursando las licenciaturas de Administración, Contaduría, Administración Financiera, Negocios y Comercio Internacional, Turismo, Economía y Actuaria. Se realizó por medio de Google Forms y se distribuyó por medio de WhatsApp, la encuesta constó de 11 preguntas de opción múltiple y una pregunta abierta, se comenzó con grupos iniciales de encuestados y se les pidió que enviaran la encuesta a sus contactos que estuviesen dentro de la población objetivo, esto para lograr el muestreo en cadena o bola de nieve.

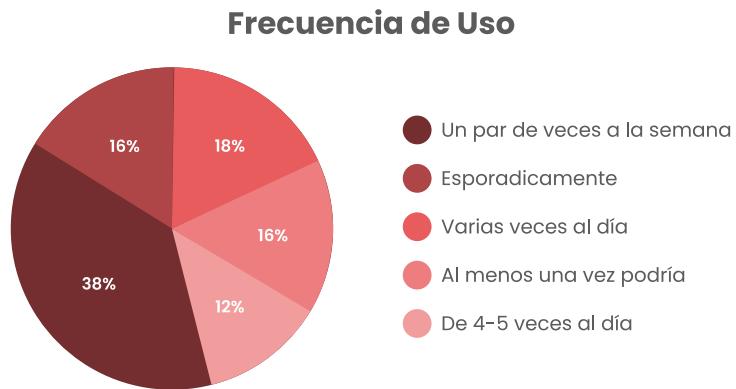
Antes de lanzar la encuesta general se realizó una prueba piloto en un grupo de 12 estudiantes, como parte de la retroalimentación se pidió que la pregunta 2 permitiera elegir varias opciones (pese a la redacción que dice "el uso principal") por lo que se modificó la opción de respuesta y por tanto los porcentajes de los resultados serán mayores a un 100%, este cambio también permite dar un mejor panorama de cómo están utilizando los LLM. En la siguiente sección se presentan los resultados con una breve reflexión sobre los mismos.

Análisis y discusión de resultados

De los estudiantes encuestados un 83.8% afirmó utilizar LLM al menos un par de veces por semana, con más de 30% (del total de encuestados) utilizándolo de manera

diaria. Esto marca una tendencia de uso alta, pudiendo compararse al uso de redes sociales, indicando una clara tendencia a la normalización del uso de modelos de texto de IA.

Figura 1
Resultados frecuencia de uso



En los usos principales resulta prudente advertir que es probable que exista un sesgo, al tratarse de una encuesta que fue distribuida con ayuda de los jefes de grupo por medio de WhatsApp, a pesar de ser anónima, puede existir desconfianza de si se pueden rastrear las respuestas o el uso que se le dará a las mismas; por lo que no es sorprendente que el uso para ayuda en tareas se encuentre en cuarta posición, en la opción de otros las respuestas fueron muy variadas para tener alguna relevancia, lo que es importante notar es la función de buscador que se le está dando a los LLM, con una abrumadora mayoría usándolo ya sea para la búsqueda de información, la resolución de dudas o ambas.

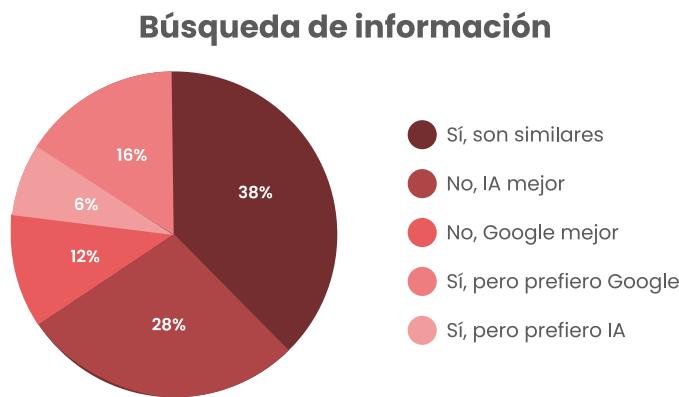
Tabla 1
Usos frecuentes de LLM en estudiantes

Usos principales que se le da a ChatGPT y otros LLM	
Resolver dudas	239
Búsqueda de información	232
Resumir lecturas o artículos	89
Escribir tareas	43
Redactar correos o mensajes	26
Otros	12

Uno de los temas más discutidos en torno al uso de LLM en el contexto del aula estudiantil tiene que ver con deshonestidad académica (Dakakni y Safa, 2023), es decir, el uso de estos LLM para la resolución de problemas, redacción de trabajos, entre otras tareas, como se mencionó en el párrafo anterior, por la forma de distribuir la encuesta se sabía que podía haber un sesgo en este aspecto, por lo que el enfoque del instrumento más que buscar reafirmar esta parte, estaba orientado a identificar cual era la percepción sobre los resultados que arrojan dichos modelos. Si bien las alucinaciones en LLM son un tema ya estudiado por la comunidad académica, específicamente los que se enfocan en el estudio de la IA, siendo estas alucinaciones una de las razones principales por las que no se recomiendan en contextos en los que la veracidad sea primordial (Lavrinovics et al., 2025), una de las preocupaciones que

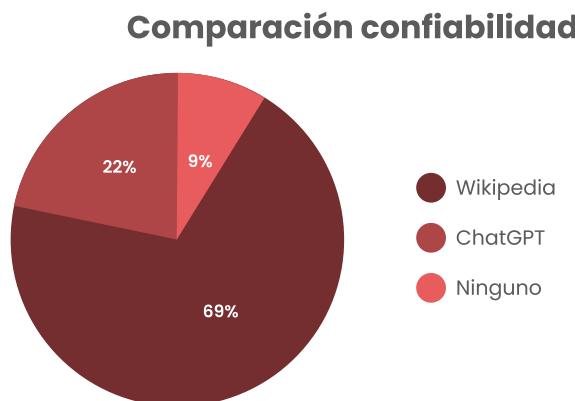
parece confirmarse con los resultados es que los estudiantes no sean conscientes de esta limitación y utilicen algún LLM en sustitución de un buscador tradicional, un 60% de los encuestados considera que son similares con un preocupante 28% que considera que un LLM es mejor a un buscador.

Figura 2
Herramienta para búsqueda de información



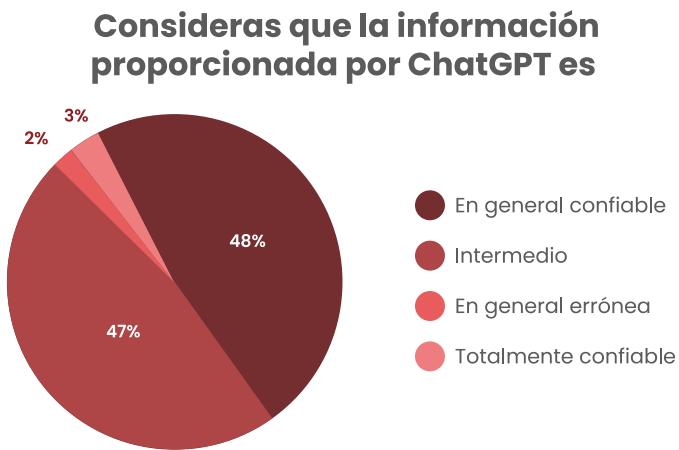
Con el surgimiento de los LLM y la popularización de ChatGPT y otros modelos, ha ido quedando fuera de la discusión uno de los previos favoritos (y controversiales) sitios para búsqueda de información, Wikipedia solía ser uno de los ejemplos de donde no obtener información, especialmente en el ámbito académico, esto debido a su forma de edición y verificación, en teoría cualquiera podía editarla y por tanto los datos ahí presentados no eran de fiar, sin embargo, en los últimos años se ha dado una mejora en la percepción de los artículos ahí presentados, con diversos análisis argumentando que la calidad en ciertos temas puede ser comparable o incluso mejor que artículos especializados (Konieczny, 2021), lo que si no es debatible es que en general sus afirmaciones deben llevar fuentes y estas fuentes son relativamente sencillas de verificar, quedando al usuario la decisión de si se puede confiar en lo ahí presentado. Resulta interesante, por tanto, que la mayoría de estudiantes considera a ChatGPT como una fuente más confiable.

Figura 3
Percepción de confiabilidad



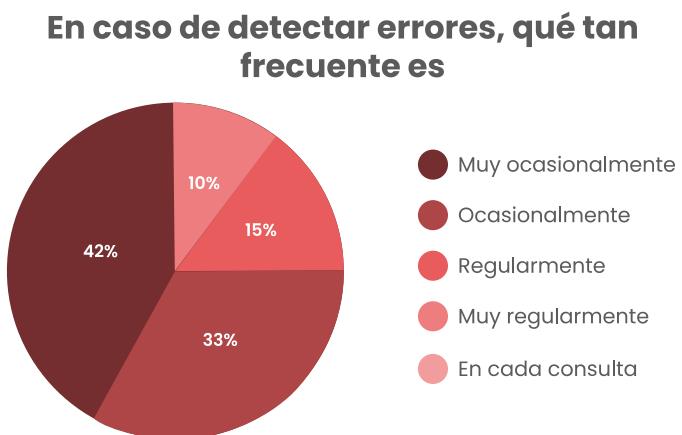
Derivado de la pregunta anterior, si ChatGPT es más confiable que Wikipedia, al menos en la percepción de los encuestados, ¿qué tanto confían en los datos emitidos? Ligeramente más del 50% tienen una muy buena percepción de la información que arroja.

Figura 4
Confiabilidad de LLM



Si se está utilizando un LLM para preguntarle de manera casi constante dudas generales o buscando datos, ¿qué tanto llegan a detectar errores? En esta pregunta un 68.8% de los encuestados respondió sí haber detectado errores, la pregunta de seguimiento tiene que ver la frecuencia de detección de errores, en esta pregunta un cierto porcentaje de los que respondió no haber detectado errores terminó respondiendo que había detectado muy ocasionalmente, siendo esta respuesta y la de ocasionalmente (ambas opciones indican que es por debajo del 50% de las consultas) la gran mayoría de respuestas, es decir, que la mayoría de consultas en estos modelos son tomadas como verdaderas por los estudiantes.

Figura 5
Frecuencia de errores detectados

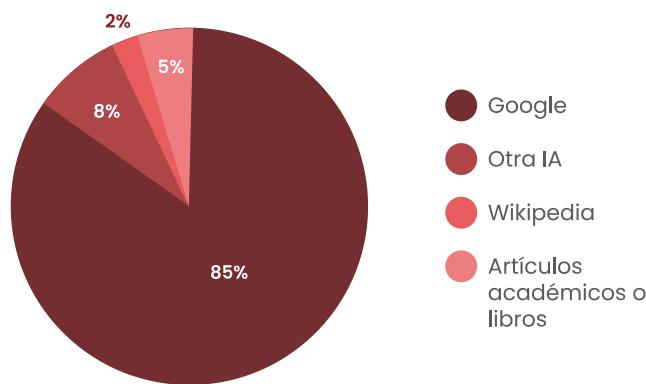


Ante la pregunta de si al dudar de una respuesta que les estuviera dando el LLM utilizaban alguna otra fuente, la mayoría de personas respondió que sí (un 92.2% de los encuestados), en la pregunta de seguimiento de cómo corroboraban el resultado, la mayoría respondió que utilizaba Google, el cual, si bien es un buscador tradicional, ya incluye en el inicio una herramienta de IA, la cual puede tener sus propios problemas de fiabilidad, aun así resulta reconfortante que un mínimo de estudiantes elijan corroborar con otro LLM (un 8%) y aún más que un 5% elija corroborar con artículos académicos y/o libros.

Figura 6

Fuente para corroborar resultados

En caso de corroborar respuestas, qué fuente alterna utilizas

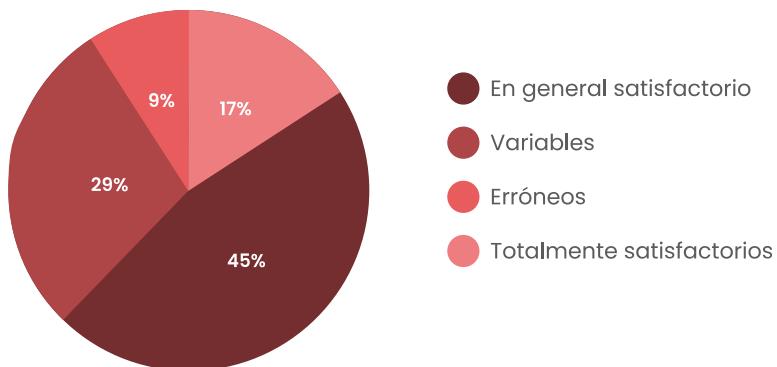


En la última sección se preguntó si habían utilizado un LLM para temas de matemáticas, en la cual un 56% de personas respondió que sí, de ese porcentaje se les preguntó cómo habían sido los resultados y en esta parte son mixtos los resultados, si bien más del 50% no han tenido mala experiencia con las preguntas de matemáticas. Esta parte de la encuesta se incluyó por que los LLM tienen dificultades en la parte de razonamiento matemático, incluso los modelos más avanzados (López Espejel et al., 2023), por lo que en teoría son más dados a equivocarse y si el estudiante tiene nociones del tema que está consultando, debería ser más fácil de identificar cuando el modelo se equivoca. El 17% que afirma siempre tener resultados satisfactorios en esta parte pudiera deberse a que hayan sido pocas las consultas y por tanto un menor porcentaje de probabilidad de fallo, o bien un desconocimiento del tema que se está consultando y por tanto se asume que los resultados son correctos, aunque en ocasiones no lo sean.

Figura 7

Confiabilidad de LLM para matemáticas

Si utilizan LLM para matemáticas, los resultados son



A manera de conclusión

Los LLM parecen haber llegado para quedarse, hasta hace unos meses parecía que específicamente ChatGPT llegaría a tener la influencia que tuvo Google en su momento, cuya marca se convirtió en sinónimo con buscar en internet. Con la llegada de otros modelos aparentemente más capaces ese futuro se vuelve incierto, pero no así el de la integración de los LLM en la vida estudiantil. Estos modelos están sustituyendo la búsqueda de información por medios tradicionales, como artículos y libros, pero también las búsquedas en portales de internet tradicionales, los cuales podían ser la puerta de acceso a información verificada.

Uno de los mayores problemas está relacionado con el desconocimiento de cómo funcionan los LLM, derivado de la encuesta, se puede percibir que hay poca crítica a los resultados que se obtienen de estos modelos, se toman como verdaderos los resultados y se llega a creer que tienen el mismo nivel de confiabilidad o incluso más que otras fuentes. No solo son los estudiantes los que desconocen el funcionamiento al ser una tecnología relativamente reciente, hay una parte importante de docentes que no están correctamente informados, prueba de esto es la creciente preocupación con la deshonestidad académica al usar estas herramientas, y una no tan marcada preocupación de que están aprendiendo cosas erróneas. Se hace énfasis en que no se deben usar porque es hacer trampa, como si fuera un tercero quien hace la tarea, pero sería importante recalcar que no se deben usar como buscador para todo porque no hay una certeza en la información que emite, para esto se debe capacitar a docentes en cómo funciona la tecnología a fondo de forma que puedan advertir de manera adecuada a sus estudiantes.

Referencias

- Abdelaal, E., Gamage, S. H. P. W., & Mills, J. E. (2019). Artificial intelligence is a tool for cheating academic integrity. *AAEE 2019 Annual Conference, (December)*, 1–7.
- Barman, D., Guo, Z., y Conlan, O. (2024). The dark side of language models: exploring the potential of LLMs in multimedia disinformation generation and dissemination. *Machine learning with applications*, 16(March).
- Dakakni, D., y Safa, N. (2023). Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. *Computers and Education: Artificial Intelligence*, 5 (August).
- Forero-Corba, W., y Negre Bennasar, F. (2024). Técnicas y aplicaciones del Machine Learning e Inteligencia Artificial en educación: una revisión sistemática. *RIED-Revista Iberoamericana de Educación a Distancia*, 27(1), 0–34.
- Ihekweazu, C., Zhou, B. (2023). The use of artificial intelligence in academic dishonesty: Ethical considerations. *Isocap.Us*, 1–10.
- Jiang, P., Sonne, C., Li, W., You, F., y You, S. (2024). Preventing the immense increase in the life-cycle energy and carbon footprints of LLM-powered intelligent chatbots. *Engineering*, 40, 202–210.
- Kajiwara, Y., y Kawabata, K. (2024). AI literacy for ethical use of chatbot: Will students accept AI ethics? *Computers and Education: Artificial Intelligence*, 6(March), 100251.
- Konieczny, P. (2021). From adversaries to allies? The uneasy relationship between experts and the Wikipedia community. *She Ji*, 7(2), 151–170.

- Lavrinovics, E., Biswas, R., Bjerva, J., y Hose, K. (2025). Knowledge graphs, large language models, and hallucinations: An NLP perspective. *Journal of Web Semantics*, 85(December 2024), 100844.
- López Espejel, J., Ettifouri, E. H., Yahaya Alassan, M. S., Chouham, E. M., & Dahhane, W. (2023). GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. *Natural Language Processing Journal*, 5(August), 100032.
- Mao, Y., Ge, Y., Fan, Y., Xu, W., Mi, Y., Hu, Z., & Gao, Y. (2024). A Survey on LoRA of large language models. 1–124.
- Martín-Crespo Blanco, M., Salamanca Castro, A., (2007). El muestreo en la investigación cualitativa. *Nure Investigación*, 27(Marzo-Abril).
- Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care*, 27(1), 1–5.
- Sun, Y., Sheng, D., Zhou, Z., y Wu, Y. (2024). AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanities and Social Sciences Communications*, 11(1), 1–14.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., y Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 2017-Decem(Nips), 5999–6009.
- Yan, B., Li, K., Xu, M., Dong, Y., Zhang, Y., Ren, Z., y Cheng, X. (2024). On protecting the data privacy of large language models (LLMs): A survey. *High-Confidence Computing*, 100300.

Anexos

Encuesta de uso de IA

1. ¿Con qué frecuencia usas Chat GPT o algún otro modelo de IA?

- Varias veces al día
- Al menos una vez por día
- De 4-5 veces por semana
- Un par de veces por semana
- Esporádicamente (una o dos veces por mes)

2. El uso principal que le das es para:

- Resolver dudas
- Búsqueda de información
- Resumir artículos o lecturas de clase
- Escribir tareas
- Redactar correos
- Otro (especifique cual)

3. Para búsqueda de información ¿Consideras qué usar ChatGPT o alguna alternativa es lo mismo que usar Google o algún otro buscador?

- Si, son similares
- No, ChatGPT (u otra IA) es mejor
- No, Google (u otro buscador) es mejor
- Sí, pero prefiero Google (u otro buscador)
- Sí, pero prefiero ChatGPT (u otra IA)

4. ¿Cuál crees que es una fuente más confiable, Wikipedia o ChatGPT (u otra IA)?

- Wikipedia
- ChatGPT (u otra IA)
- Ninguno de los dos

5. Consideras que la información que te proporciona ChatGPT (u otra IA) es:
 - Totalmente confiable
 - En general confiable
 - A veces confiable a veces errónea
 - En general errónea
6. Has detectado información errónea en los resultados que te presenta ChatGPT o su alternativa:
 - Sí
 - No
7. En caso de responder sí, con qué frecuencia:
 - En cada consulta (100% del tiempo)
 - Muy regularmente (70-90% del tiempo)
 - Regularmente (50-70% del tiempo)
 - Ocasionalmente (30-50% del tiempo)
 - Muy ocasionalmente (menos del 30%)
8. En caso de dudar de alguna respuesta, utilizas alguna otra fuente para corroborar la respuesta emitida:
 - Sí
 - No
9. En caso de responder sí, cuál utilizas normalmente:
 - Google
 - Otra IA
 - Wikipedia
 - Otra (especifique)
10. Utilizas o has utilizado ChatGPT u otra IA para temas de matemáticas:
 - Sí
 - No
11. En caso de responder sí, los resultados han sido:
 - Totalmente satisfactorios
 - En general satisfactorios
 - Variables
 - En general erróneos
12. Emite una breve opinión personal sobre ChatGPT u otras IA que hayas utilizado.