



Alfabetización crítica en IA. Recursos educativos para una pedagogía de la descajanegrización  
Alejandro Artopoulos, Alejandra Lliteras  
Trayectorias Universitarias, 10 (19), e168, 2024  
ISSN 2469-0090 | <https://doi.org/10.24215/24690090e168>  
<https://revistas.unlp.edu.ar/TrayectoriasUniversitarias>  
Universidad Nacional de La Plata  
La Plata | Buenos Aires | Argentina

## Alfabetización crítica en IA. Recursos educativos para una pedagogía de la descajanegrización

### Critical literacy in AI. Educational resources for a pedagogy of de-blackboxing

Alejandro Artopoulos

<https://orcid.org/0000-0003-4554-2412>

[alepoulos@udesa.edu.ar](mailto:alepoulos@udesa.edu.ar)

Universidad de San Andrés, Argentina

Alejandra Lliteras

<https://orcid.org/0000-0002-4148-1299>

[alejandra.lliteras@lifa.info.unlp.edu.ar](mailto:alejandra.lliteras@lifa.info.unlp.edu.ar)

Universidad Nacional de la Plata, Argentina

#### RESUMEN

Este artículo explora el impacto de la inteligencia artificial (IA) generativa en la educación, con foco en los recursos educativos necesarios para la formación crítica en la educación superior. Se analiza la adopción de la IA generativa, particularmente ChatGPT, por parte de estudiantes. Los autores argumentan que, si bien la IA generativa tiene el potencial de acelerar el aprendizaje, también presenta riesgos, como la distorsión cognitiva de la realidad en estudiantes sin competencias críticas en IA debido a la dilución epistemológica del control de fuentes y consistencia conceptual que ocasionan “alucinaciones” y sesgos. Se propone la necesidad de una “alfabetización crítica en IA” para los docentes, que implica la comprensión técnica, la evaluación crítica, y la aplicación práctica en la enseñanza de los aspectos críticos y creativos de la IA en sus disciplinas. Finalmente, se argumenta que la formación crítica en IA requiere la descajanegrización de recursos educativos, tanto de los ya disponibles como de los nuevos recursos especialmente diseñados para este fin.

#### ABSTRACT

This article explores the impact of generative artificial intelligence (AI) in education, with a focus on the educational resources needed for critical training in higher education. It analyses the adoption of generative AI, particularly ChatGPT, by students. The authors argue that while generative AI has the potential to accelerate learning, it also presents risks, such as cognitive distortion of reality in students without critical AI skills due to epistemological dilution of source control and conceptual consistency leading to ‘hallucinations’ and biases. The need for ‘critical AI literacy’ for teachers is proposed, involving technical understanding, critical evaluation, and practical application in teaching the critical and creative aspects of AI in their disciplines. Finally, it is argued that critical AI literacy requires the de-blackboxing of educational resources, both those already available and new resources specially designed for this purpose.

#### PALABRAS CLAVE

descajanegrización,  
IA Generativa,  
alucinación,  
formación crítica

#### KEYWORDS

deblackboxing,  
Generative AI,  
hallucination,  
critical training



## INTRODUCCIÓN

Luego de la pandemia, la emergencia de la IA generativa ha introducido una nueva ola de desafíos a la relación tecnología y educación. Herramientas como ChatGPT<sup>1</sup>, Gemini<sup>2</sup>, Copilot<sup>3</sup> (entre otras) han puesto a disposición la capacidad automatizada para generar textos que resultan coherentes. El evento generó un gran entusiasmo tanto del público como de estudiantes, pero en particular en quienes trabajan desde una mirada tecnopedagógica (Gurukkal, 2021).

Uno de los principales riesgos asociados a la IA generativa es la propagación de sesgos algorítmicos. Los modelos de lenguaje se entrenan con grandes cantidades de datos, que pueden reflejar y amplificar las desigualdades sociales y culturales existentes. Esto puede llevar a la generación de contenidos discriminatorios o estereotipados, lo que podría tener un impacto negativo en la formación de los estudiantes, sumado a los ajustes finos de los modelos que son realizados por humanos (Schneider, 2022).

Además, la IA generativa no está exenta de generar lo que se conoce como alucinaciones, es decir, información falsa o engañosa que puede ser difícil de distinguir de la información que se considera verdadera. Estas alucinaciones, dada la dificultad de su detección, plantean desafíos para la evaluación de la calidad de los textos generados y problematiza aún más los potenciales riesgos en el proceso de aprendizaje ya que, detectar estas alucinaciones, requiere, entre otras cosas, de habilidades de pensamiento crítico (Chelli et al., 2024).

Por último y en línea con lo anterior, otro riesgo surge al considerar la posibilidad de que los estudiantes dependan (en gran medida) de la IA generativa para realizar sus producciones, lo que podría socavar el desarrollo de habilidades fundamentales como la lectura

1 <https://chatgpt.com/>

2 <https://gemini.google.com/app>

3 <https://copilot.microsoft.com>

comprensiva, la escritura creativa y el pensamiento crítico, ya que al generar este tipo de IA contenidos automáticamente, el estudiante podría perder la capacidad de buscar información a partir de diferentes fuentes, analizarla, relacionarla y generar, a partir de lo anterior argumentaciones sólidas.

Por lo anterior es importante un abordaje que considere como estas tecnologías interfieren y muchas veces moldean al individuo más allá de un enfoque instrumentalista (Jandrić y Knox, 2022).

En este trabajo nos ocupamos de la necesidad de desarrollar la alfabetización crítica en IA mediante el cultivo de capacidades de identificación de alucinaciones y sesgos algorítmicos, tanto para docentes como para estudiantes. Para lo cual argumentamos que, dada la dificultad de llevar a cabo esta “pedagogía de las alucinaciones”, es necesario desarrollar dichas capacidades con recursos educativos descajanegrizados.

La propuesta de descajanegrizar recursos educativos se origina en el concepto de cajanegrizar. Del inglés *blackboxing*, se trata de un término que se emplea en los estudios de ciencia, tecnología y sociedad para denominar al proceso por el cual una tecnología se vuelve invisible cuando se estabiliza y es aceptada por la sociedad. Se descomplejiza y solo se utiliza como un objeto tecnológico sencillo. Paradójicamente, cuanto más éxito tiene la tecnología más opaca se vuelve (Latour, 2008).

**La propuesta de descajanegrizar recursos educativos se origina en el concepto de cajanegrizar. Del inglés *blackboxing*, se trata de un término que se emplea en los estudios de ciencia, tecnología y sociedad para denominar al proceso por el cual una tecnología se vuelve invisible cuando se estabiliza y es aceptada por la sociedad.**

Por lo tanto la propuesta indica que para develar los aspectos opacos de la IA no solo es necesario formar la mirada para detectar alucinaciones, también son requeridos recursos educativos que permitan experimentar procesos cajanegrizados de la IA. La alfabetización crítica en IA requiere la creación de nuevos materiales didácticos que son posibles transparentar de lo que se encuentra opaco de la IA.

## **IA GENERATIVA EN EDUCACIÓN**

La IA generativa permite, en base a sus datos de entrenamiento (fuentes de datos), generar contenido (respuestas) como, por ejemplo,

texto, imagen y video. El modelo de lenguaje se diseña y luego que se implementa pasa por una etapa de entrenamiento automático y puede pasar por un entrenamiento por parte de personas (entrenamiento humano). La Figura 1 muestra un diagrama simplificado de lo mencionado.

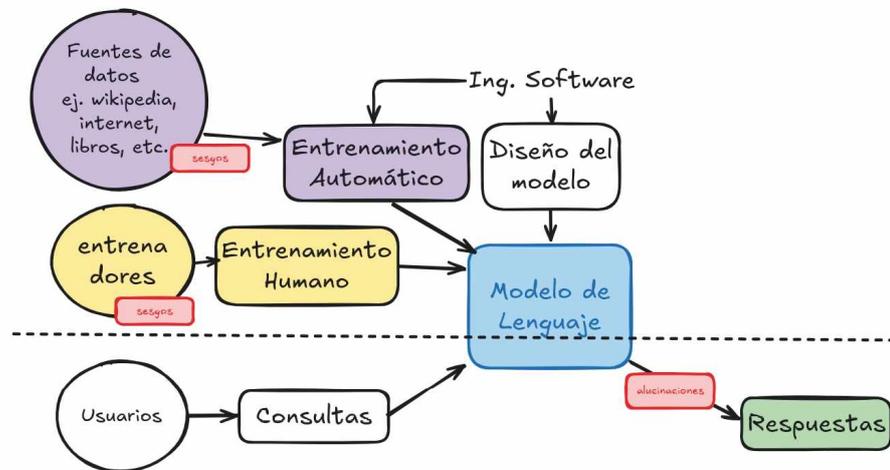


Figura 1: Esquema de ciclo de vida de un modelo de lenguaje [Fuente Propia]

En la Figura 1 se muestra que los sesgos pueden ser introducidos con los datos al momento de entrenar automáticamente o bien por los humanos, mientras que las alucinaciones, se pueden producir cuando el modelo genera una respuesta en base a una consulta de un usuario, tanto los sesgos como las alucinaciones se impregnan en la respuesta que recibe el usuario. La línea punteada en la Figura 1 separa lo que es accesible por un usuario de lo que se encuentra cajanegrizado.

A fines del 2022 la IA generativa se conoce por su representante más popular, ChatGPT, a los pocos meses ya fue adoptado por los estudiantes. En general, viendo en ella, un aliado para su apoyo en el aprendizaje que le permitía pedir explicaciones y ejemplos adicionales sobre un tema, resumir textos, generar ideas, resolver actividades educativas y además, como una posibilidad de soporte para favorecer la educación inclusiva (Fütterer et al., 2023).

Es necesario notar que la introducción de la IA se dió en el contexto de la continua y pronunciada caída de la inversión de riesgo en el sector Edtech. La inversión global EdTech luego de alcanzar un pico de u\$ 22,73 billones en 2021 ha caído continuamente de forma acelerada en los últimos tres años, llegando en 2024 a niveles prepandémicos. Se trata de una paradoja. En tanto los inversores se retiran de la educación, la tecnología de vanguardia requiere de la educación para su legitimación (Glasner, 2024).

Desde el 2023, la IA generativa constituye la cuarta disrupción en el sistema educativo, luego de la Web 2.0 (2006-2012), teléfonos inteligentes, plataformas y nube (2013-2019), la pandemia con el ecosistema de plataformas (2020-2022). Se trata de imaginarios sociotécnicos de la educación que han dominado la escena, y que están siendo sometidos a fuertes transformaciones en la actualidad (Artopoulos y Lliteras, 2024; Dussel y Williams, 2023).

**Desde el 2023, la IA generativa constituye la cuarta disrupción en el sistema educativo, luego de la Web 2.0 (2006-2012), teléfonos inteligentes, plataformas y nube (2013-2019), la pandemia con el ecosistema de plataformas (2020-2022).**

En este contexto el reto de acompañar a los estudiantes en el “desarrollo de pensamiento crítico y su capacidad creativa, especialmente en lo que se refiere a la comprensión de los procesos y mecanismos computacionales”, redundaría en que “las habilidades informáticas son un requisito previo para el pensamiento crítico”. La emergencia del fenómeno de la IA es la relación sólida entre la formación de capacidades de pensamiento crítico y la condición sociomaterial de su práctica en los entornos computacionales (Silander et al., 2022, p. 106).

Heinz analiza cómo debería cultivarse la sinergia entre la alfabetización en IA y el pensamiento computacional en contextos cotidianos de resolución de problemas (Heintz, 2022). Desde la mirada de Dennig y Tedre no sería un solo tipo genérico de pensamiento computacional, sino que un pensamiento computacional para cada tipo de contexto de aplicación y de los niveles educativos o la profesión en la cual se aplica (Denning y Tedre, 2019).

Por lo cual si consideramos la transformación digital de los procesos de prácticas de lectoescritura desde la perspectiva de las humanidades digitales, los saberes computacionales críticos que entrelazan la formación en lectoescritura con el pensamiento computacional, incluirían la lectura lejana o distante y el uso responsable de la IA como los asistentes de escritura, traducción y aprendizaje de idiomas (Artopoulos y Lliteras, 2024).

Para superar las limitaciones de la experiencia cajanegrizada de ChatGPT en la que no es posible saber cuál o cuáles fuentes de datos consulta para generar contenido, se puede recurrir a herramientas como NotebookLM donde utilizar el modelo de lenguaje curando un

corpus específico (un conjunto de archivo en formato pdf desde el cual se quiere genere el contenido) y así interactuar con el mismo (Huffman y Hutson, 2024).

Desde la visión tecnopedagógica, sería aún más didáctico contar con recursos abiertos que permitan superar las limitaciones de los entornos cerrados. Por ejemplo, para experimentar en un entorno seguro el procesamiento de lenguaje natural en la lectura distante.

La lectura distante o lejana es una práctica de la teoría literaria por la cual se experimentó la exploración de corpus de textos mediante algoritmos. Con la proliferación de herramientas de búsqueda en procesadores de textos, esta práctica se incorporó también en las actividades cotidianas que muchos usuarios realizan con textos. Tareas que realizan la mayoría de las veces sin una formación específica (Moretti, 2013).

La propuesta tecnopedagógica apunta a continuar un sendero similar a lo que sucedió con el desarrollo de lenguajes de programación para niños. Scratch, como material didáctico, permitió desarrollar el pensamiento computacional y las narrativas digitales, basado en un modelo de recurso abierto que está en la nube y que cuenta con una comunidad que comparte sus creaciones y tutoriales. De manera análoga, se podría pensar en contar con recursos educativos abiertos que permitan trabajar los saberes computacionales críticos, como por ejemplo la comprensión lectora mediante la lectura lejana.

## **ADOPCIÓN DE LA IA**

Como se ha mencionado anteriormente, los estudiantes adoptaron en gran medida a la IA para resumir textos o resolver actividades de aprendizaje. Lo que puede conducir a la disminución de pensamiento crítico y de habilidades de resolución de problemas. A raíz de lo anterior, los autores indican la necesidad de mejorar la alfabetización digital de los estudiantes para garantizar, entre otras cosas, un uso ético de IA (Lo et al., 2024).

**Para conocer datos preliminares acerca de la adopción de esta tecnología por parte de estudiantes universitarios, se diseñó una encuesta semiestructurada para estudiantes que cursan alguna de las carreras disponibles en la Facultad de Informática, UNLP, en particular la encuesta se disponibilizó vía un formulario de Google Forms los dos primeros días hábiles de diciembre 2024 e indagaba acerca del uso e impacto de la IA en su quehacer como estudiante.**

Para conocer datos preliminares acerca de la adopción de esta tecnología por parte de estudiantes universitarios, se diseñó una encuesta semiestructurada<sup>4</sup> para estudiantes que cursan alguna de las carreras disponibles en la Facultad de Informática, UNLP, en particular la encuesta se disponibilizó vía un formulario de Google Forms los dos primeros días hábiles de diciembre 2024 e indagaba acerca del uso e impacto de la IA en su quehacer como estudiante. El grupo destinatario, hace que los resultados obtenidos, puedan estar sesgados ya que los estudiantes están más familiarizados con algunos conceptos en relación a otras carreras.

Se indagó respecto a las herramientas de IA que habían utilizado, su percepción de la IA respecto con su aprendizaje, el concepto de sesgo y alucinaciones, qué rol deberían adoptar los docentes frente a la IA, su percepción respecto a la comprensión lectora y a la forma de leer textos.

La encuesta fue respondida por 62 estudiantes (15 mujeres, 47 varones, 2 prefieren no decir). De los estudiantes que respondieron la encuesta 67% (41 estudiantes) cuentan con más del 50% de alguna de las carreras cursadas. El 62% de los encuestados tienen entre 21 y 25 años. Las herramientas de IA que mencionan usar los estudiantes se presentan

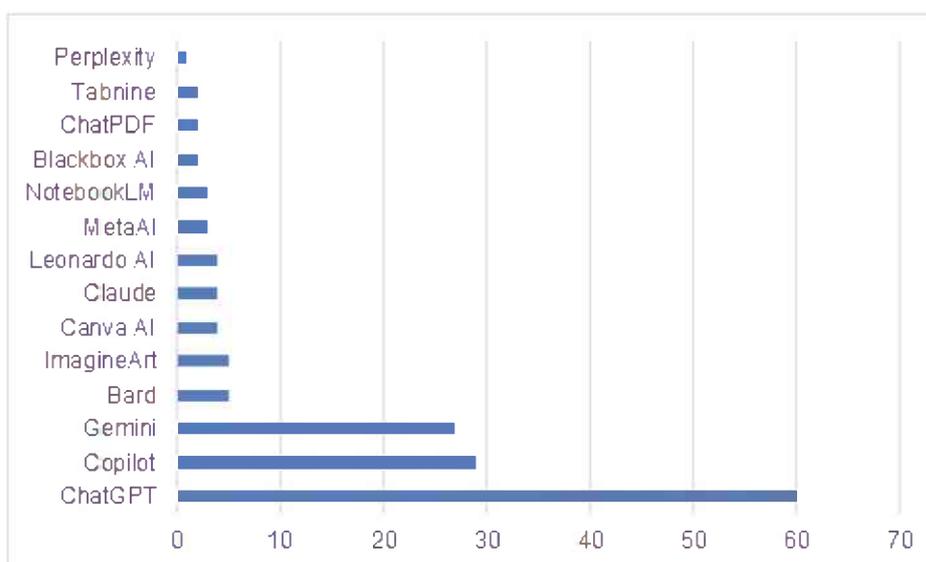


Figura 2: Herramientas de IA generativas usadas por los estudiantes [Fuente Propia]

De la Figura 2 se desprende que el 98% (61 estudiantes) usaron ChatGPT, y sólo un estudiante no lo usó indicando que en su lugar usó

<sup>4</sup> Encuesta sobre el Uso de la IA Generativa en el Nivel Universitario (1.0). Zenodo. <https://doi.org/10.5281/zenodo.14292073>

Gemini. Aparecen en la lista de herramientas algunas que se usan para generar código para algún lenguaje de programación en particular, lo que se relaciona directamente con el perfil de las personas que completaron esta encuesta. De los encuestados, sólo dos personas usaron herramientas más específicas como NotebookLM.

Para conocer la opinión de los estudiantes, se empleó una escala likert para la respuesta (1:Totalmente en desacuerdo; 5:Totalmente en desacuerdo) y se presentó la afirmación “Considero que el uso de herramientas de IA generativas ha mejorado mi experiencia de aprendizaje en general” para lo cual el 61.3 % selecciona las opciones 4 y 5, mientras que en la media se ubica el 24.2% y los restantes en los extremos 1 y 2. Lo anterior muestra que la percepción de los estudiantes es que este tipo de IA mejoran su experiencia de aprendizaje.

Respecto a los sesgos, el 78.2% conoce el concepto y de su posible presencia en los contenidos que genera esta IA, sin embargo el 45.2% de los encuestados indica no saber cómo detectarlos o bien se sienten que poco pueden hacer para detectarlos. Respecto a las alucinaciones, el 53.6% conoce del tema y sus consecuencias, y el 74.2% sabe que el contenido generado puede ser falso o inventado, respondiendo el 93,5% de los encuestados que es necesario verificar el contenido obtenido.

**Respecto a los sesgos, el 78.2% conoce el concepto y de su posible presencia en los contenidos que genera esta IA, sin embargo el 45.2% de los encuestados indica no saber cómo detectarlos o bien se sienten que poco pueden hacer para detectarlos.**

El 85% de los encuestados indican la importancia del rol de docente para un uso crítico, ético y responsable de la IA en educación, el 72.6% indica que es necesario que los docentes diseñen actividades que complementen y potencien el uso de IA, mientras que el 93.5% coincide en la necesidad de que los docentes enseñen acerca de las limitaciones de este tipo de herramientas y a no depender de las mismas en el ámbito educativo.

### **DESCAJANEGRIZACIÓN DE RECURSOS EDUCATIVOS**

Los Recursos Educativos Abiertos (REA) son materiales de aprendizaje, enseñanza e investigación en cualquier formato y soporte que existen en el dominio público o están bajo derechos de autor y fueron liberados bajo

una licencia abierta, que permiten su acceso sin costo, su reutilización, reorientación, adaptación y redistribución por parte de terceros.

Dentro de estos materiales de aprendizaje podemos encontrar software para la enseñanza de disciplinas específicas como GeoGebra<sup>5</sup>. Software matemático gratuito y de código abierto que se puede utilizar para crear actividades didácticas interactivas. Para su utilización es necesario que dicho software se ofrezca como un servicio en la nube. Algo similar ocurre con Zotero<sup>6</sup>, CODAP<sup>7</sup> o Wikipedia<sup>8</sup>.

Siguiendo el diseño de REA en la nube proponemos desarrollar software para la enseñanza de la alfabetización crítica en IA. Para lo cual es necesario develar el funcionamiento de los modelos de lenguaje mediante la descajanegrización de algunos de sus aspectos. Por ejemplo desarrollando actividades de lectura lejana con el procesamiento de lenguaje natural.

La propuesta apunta a develar los aspectos opacos de la IA formando la mirada para detectar alucinaciones, y también mediante recursos educativos que permitan experimentar procesos descajanegrizados de la IA.

Como parte de la descajanegrización de los recursos educativos, Lliteras et al., proponen Alfadatizando<sup>9</sup> como una plataforma para crear, usar y compartir recursos educativos que involucren actividades para el desarrollo de pensamiento computacional a través de la visualización de datos, en particular en áreas de humanidades y ciencias sociales. Desde esta plataforma es posible crear, usar, y resignificar actividades de visualización así como fuentes de datos.

En el trabajo citado, los autores presentan una actividad usando un grafo de personajes de la obra literaria “La casa de Bernarda Alba” para trabajar aspectos de pensamiento computacional en el área de literatura (Lliteras et al., 2024).

La actividad propuesta por los autores se presenta como complementaria a la lectura tradicional en el aula llevando a cabo lo que se conoce como lectura distante (Moretti, 2013).

Este recurso educativo se encuentra en una versión prototípica disponible en la web y es de código abierto, lo que permite no sólo conocer cómo está implementado, sino que además extenderlo y personalizarlo para contemplar otro tipo de visualizaciones acordes al nivel educativo en el que se desee trabajar.

5 <https://www.geogebra.org>

6 <https://www.zotero.org/>

7 <https://codap.concord.org/>

8 <https://es.wikipedia.org>

9 <https://alfadatizando.okd.lifia.info.unlp.edu.ar/>

## CONCLUSIONES

La paradoja de que justo en el momento en que cae la inversión de riesgo en tecnología educativa de vanguardia las grandes empresas de tecnología que lideran el desarrollo de IA generativa necesitan casos de aplicación en educación nos llama a reflexionar. En tanto los inversores se retiran de la educación, la alfabetización en IA requiere el desarrollo de nuevos recursos educativos.

La alfabetización crítica en IA además de la pedagogía de las alucinaciones requiere el desarrollo de nuevos materiales didácticos que favorezcan el desarrollo del pensamiento crítico y la resolución de problemas en entornos tecnológicos menos opacos. Es decir que descajanegricen en lo posible la experiencia del trabajo asistido con IA.

El resultado de la encuesta preliminar con estudiantes presentada muestra que los estudiantes si bien declaran conocimiento teórico sobre sesgos y alucinaciones, admiten que tiene muchas dificultades para su detección en situaciones de práctica. Este es el punto neurálgico de la alfabetización crítica en IA. La práctica de la detección de sesgos y alucinaciones.

Si bien usan este tipo de herramientas esperan que los docentes les brinden soporte para su uso, aun siendo estudiantes de temas relacionados a la informática

Por lo tanto se verifica que nuestra propuesta de alfabetización crítica en IA requiere tanto de la pedagogía de la detección de alucinaciones como de el diseño y construcción de recursos educativos abiertos que permitan experimentar procesos descajanegrizados de la IA.

## REFERENCIAS BIBLIOGRÁFICAS

Artopoulos, A. y Lliteras, A. (2024). La emergencia de la alfabetización crítica en IA : La reconstrucción social de la ciudadanía en democracias bajo acecho digital. *Revista Diálogo Educativo*, 24(80).  
<https://doi.org/10.7213/1981-416X.24.083.DS01>

Chelli, M., Descamps, J., Lavoué, V., Trojani, C., Azar, M., Deckert, M., Raynier, J.-L., Clowez, G., Boileau, P. y Ruetsch-Chelli, C. (2024). Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis. *Journal of Medical Internet Research*, 26, e53164.  
<https://doi.org/10.2196/53164>

Denning, P. J. y Tedre, M. (2019). *Computational thinking*. The MIT Press.

- Dussel, I. y Williams, F. (2023). Los Imaginarios sociotécnicos de la política educativa digital en México (2012-2022). *Profesorado, Revista de Currículum y Formación del Profesorado*, 27(1), 39-60.  
<https://doi.org/10.30827/profesorado.v27i1.26247>
- Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M. y Gerjets, P. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, 13(1), 15310.  
<https://doi.org/10.1038/s41598-023-42227-6>
- Glasner, J. (11 de junio de 2024). Edtech Funding Has Not Hit Bottom. *Crunchbase News*.  
<https://news.crunchbase.com/edtech/startup-funding-down-global-us-2024/>
- Gurukkal, R. (2021). Techno-Pedagogy Needs Mavericks. *Higher Education for the Future*, 8(1), 7-19.  
<https://doi.org/10.1177/2347631121989478>
- Heintz, F. (2022). The Computational Thinking and Artificial Intelligence Duality. En S.-C. Kong y H. Abelson (Eds.), *Computational Thinking Education in K-12* (pp. 143-152). The MIT Press.  
<https://doi.org/10.7551/mitpress/13375.003.0012>
- Huffman, P. y Hutson, J. (2024). Enhancing History Education with Google NotebookLM: Case Study of Mary Easton Sibley's Diary for Multimedia Content and Podcast Creation. *ISRG Journal of Arts Humanities & Social Sciences*, 2(5), 246-253.  
<https://doi.org/10.5281/ZENODO.13860338>
- Jandrić, P. y Knox, J. (2022). The postdigital turn: Philosophy, education, research. *Policy Futures in Education*, 20(7), 780-795.  
<https://doi.org/10.1177/14782103211062713>
- Latour, B. (2008). *Re-ensamblar lo social. Una introducción a la teoría del actor-red*. Manantial.
- Lliteras, A., Artopoulos, A., Ger, J. y Boza, J. (2024). *Alfadatizando 2.0 applied to data visualization at high school level and for digital humanities. Empowering digital citizens* [Ponencia]. 19th Latin American Conference on Learning Technologies (LACLO 2024). Universidad de la República. Montevideo, Uruguay.
- Lo, C. K., Hew, K. F. y Jong, M. S. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 219, 105100.  
<https://doi.org/10.1016/j.compedu.2024.105100>

Moretti, F. (2013). *Distant reading*. Verso.

Schneider, B. (2022). Multilingualism and AI: The regimentation of language in the age of digital capitalism. *Signs and Society*, 10(3), 362-387.

Silander, P., Riikonen, S., Seitamaa-Hakkarainen, P. y Hakkarainen, K. (2022). Learning Computational Thinking in Phenomena-Based Co-creation Projects: Perspectives from Finland. En S.-C. Kong y H. Abelson (Eds.), *Computational Thinking Education in K-12* (pp. 103-120). The MIT Press. <https://doi.org/10.7551/mitpress/13375.003.0008>