



Identidad lingüística en declive: riesgos de la homogeneización discursiva mediante inteligencia artificial

ANDRÉ RUNÉE CONTRERAS  
Universidad de Pamplona, Pamplona, Colombia

Historia del artículo / Article Info

Recibido/Received

28 de enero de 2026

Aprobado/Accepted

10 de abril de 2026

Publicado/Published online

13 de mayo de 2026

ARTÍCULO DE INVESTIGACIÓN

<https://doi.org/10.19053/uptc.0121053X.n48.2026.20147>



Citación/Citation: Contreras, A. (2026). Identidad lingüística en declive: riesgos de la homogeneización discursiva mediante inteligencia artificial. *Cuadernos de Lingüística Hispánica*, 48 2026, e20147 <https://doi.org/10.19053/uptc.0121053X.n48.2026.20147>

✉ Correspondencia/Correspondence:

Km 1 Vía Bucaramanga, Ciudad Universitaria
andre.contreras@unipamplona.edu.co

Resumen

Este artículo es un resultado de la tesis doctoral “La inteligencia artificial y su incidencia en el desarrollo de las habilidades comunicativas en estudiantes de educación superior”, que analiza los riesgos de homogeneización discursiva generados por el uso de inteligencia artificial (IA) en la producción académica universitaria. Se enfatiza el sesgo de uso en la adopción de fórmulas discursivas sugeridas por IA, que limita la creatividad y pone en riesgo la diversidad dialectal. El estudio siguió la metodología mixta, incluyendo una encuesta aplicada a 682 estudiantes de la Universidad de Pamplona, Colombia, junto con un análisis comparativo de textos producidos con y sin asistencia algorítmica en la cátedra de Habilidades Comunicativas. Los datos evidencian alta dependencia de estas herramientas, como también una reducción del pensamiento crítico. Los textos muestran patrones homogéneos y un tono impersonal; aunque la IA facilita la producción académica, plantea un desafío para preservar la identidad lingüística y la riqueza dialectológica en la educación superior.

Palabras clave: sociolingüística, variación lingüística, discurso, pensamiento crítico, inteligencia artificial, educación superior.

Linguistic Identity in Decline: Risks of Discursive Homogenization through Artificial Intelligence

Abstract

The article has been derived from the doctoral thesis entitled “Artificial intelligence and its impact on the development of communicative skills in higher education students”, in which the risks of discursive homogenization that are generated using artificial intelligence (AI) in university academic production are analyzed. The study emphasizes the bias in adopting AI-suggested discourse formulas, which limits creativity and endangers dialectal diversity. A mixed-methods approach was employed for the research, which included a survey of 682 students at the University of Pamplona in Colombia and a comparative analysis of texts produced with and without algorithmic assistance on the Communication Skills course. The data reveals a high level of dependence on these tools, as well as a decline in critical thinking. Although AI facilitates academic production, the texts show homogeneous patterns and an impersonal tone, posing a challenge to preserving linguistic identity and dialectal richness in higher education.

Keywords: sociolinguistics, linguistic variation, discourse, critical thinking, artificial intelligence, higher education.

Identité linguistique en déclin : les risques d'une homogénéisation discursive induite par l'intelligence artificielle

Résumé

Cet article est issu de la thèse de doctorat intitulée L'intelligence artificielle et son incidence sur le développement des compétences communicatives chez les étudiants de l'enseignement supérieur, qui analyse les risques d'homogénéisation discursive engendrés par l'utilisation de l'intelligence artificielle (IA) dans la production académique universitaire. L'accent est mis sur le biais d'utilisation lié à l'adoption de formules discursives suggérées par l'IA, qui limitent la créativité et mettent en péril la diversité dialectale. L'étude a été menée selon une méthodologie mixte, comprenant une enquête auprès de 682 étudiants de l'université de Pamplona, en Colombie, ainsi qu'une analyse comparative de textes produits avec et sans assistance algorithmique dans le cadre du cours de compétences communicatives. Les résultats montrent une forte dépendance à ces outils et une diminution de la pensée critique. Les textes présentent des schémas homogènes et un ton impersonnel. Si l'IA facilite la production académique, elle pose toutefois un défi pour la préservation de l'identité linguistique et de la richesse dialectologique dans l'enseignement supérieur.

Mots-clés: sociolinguistique, variation linguistique, discours, pensée critique, intelligence artificielle,

Identidade linguística em declínio: riscos da homogeneização discursiva por meio da inteligência artificial

Resumo

Este artigo é um resultado da tese de doutorado "A inteligência artificial e sua incidência no desenvolvimento das habilidades comunicativas em estudantes do ensino superior", que analisa os riscos de homogeneização discursiva gerados pelo uso da inteligência artificial (IA) na produção acadêmica universitária. Enfatiza-se o viés de uso na adoção de fórmulas discursivas sugeridas pela IA, o que limita a criatividade e coloca em risco a diversidade dialetal. O estudo seguiu uma metodologia mista, incluindo uma pesquisa aplicada a 682 estudantes da Universidade de Pamplona, Colômbia, juntamente com uma análise comparativa de textos produzidos com e sem assistência algorítmica na disciplina de Habilidades Comunicativas. Os dados evidenciam alta dependência dessas ferramentas, bem como uma redução do pensamento crítico. Os textos apresentam padrões homogêneos e um tom impessoal; embora a IA facilite a produção acadêmica, ela representa um desafio para a preservação da identidade linguística e da riqueza dialetológica no ensino superior.

Palavras-chave: sociolinguística, variação linguística, discurso, pensamento crítico, inteligência artificial, ensino superior.

Introducción

En la última década, la inteligencia artificial generativa (IAG) ha transformado de manera acelerada los escenarios de la educación superior. Modelos como *ChatGPT*, *Gemini* o *Copilot* son mediadores discursivos que pueden generar textos significativos y estilísticamente aceptables en cuestión de segundos (Floridi & Chiriatti, 2020).

Según Crompton *et al.* (2024), Cheung y Crosthwaite (2025), Lu (2025), y Lee *et al.* (2025), esta característica indudablemente hace crecer las posibilidades para el aprendizaje autodirigido y la creación académica, pero también conlleva peligros como la dependencia tecnológica y la escasez de juicio crítico. Esto sugiere, en un sentido bastante literal, que el lenguaje, considerado por muchos como el vehículo de la cultura y la identidad social *per se*, ha estado cada vez más bajo la influencia de algoritmos que promueven ciertos tipos o estandarizaciones de expresión (Crompton & Burke, 2013).

Un riesgo clave que surge de este proceso es el sesgo de uso, o el hecho de que los estudiantes imiten o adopten como estándar de corrección/legitimidad las fórmulas discursivas propuestas por la IA (Blodgett *et al.*, 2020; Hovy & Prabhumoye, 2021). Esta situación lleva a una homogeneización discursiva académica, que se caracteriza por patrones predecibles, conectores repetitivos y estilo impersonal (Curry, 2024; Gallegos *et al.*, 2024).

Este efecto no es puramente estilístico, ciertamente implica un elemento de control discursivo, en el sentido de que estos algoritmos actúan como tamizadores ideológicos que cortan de raíz las narrativas que las empresas tecnológicas prefieren no ver (Fairclough, 1995; Van Dijk, 2000, 2017).

De esta manera, debe considerarse la salvaguarda de la identidad lingüística como un principio particularmente importante, dado que la diversidad dialectal que existe en países como en Colombia, prolifera en los campus universitarios gracias a los estudiantes de diferentes regiones, lo cual constituye un capital cultural y sociolingüístico muy valioso (Bourdieu, 1991). Sin embargo, según Floridi y Chiriatti (2020), Crawford (2021) y Noble (2018), la tendencia hacia la uniformización algorítmica amenaza con invisibilizar esta riqueza y favorecer una escritura “neutral” que borra marcas de variación lingüística, rasgos dialectales o registros propios de las comunidades (Markl, 2022). Organismos como la UNESCO (2021, 2025) y World Economic Forum (2024) han advertido sobre la necesidad de defender la pluralidad lingüística frente a las fuerzas de estandarización impuestas por la tecnología y las políticas educativas.

Este artículo se inscribe en esa preocupación y se deriva de la tesis doctoral titulada “El u-learning y su incidencia en el desarrollo de las habilidades comunicativas de los estudiantes de Educación Superior”. En él se busca analizar cómo los estudiantes universitarios perciben y experimentan la influencia de la IA en sus prácticas de escritura académica, así como los riesgos de homogeneización discursiva que se derivan del sesgo de uso (Lovón & Parini, 2024; Muñoz-Basols *et al.*, 2024). La investigación parte de un

enfoque sociolingüístico y adopta un diseño mixto, combinando el análisis de una encuesta aplicada a 682 estudiantes de la Universidad de Pamplona con un estudio comparativo de textos producidos con y sin mediación algorítmica. Lo anterior, en búsqueda de una contribución conceptual a la polémica actual que atañe al valor del lenguaje; respecto a la automatización y el poder en su representación contextual, llama la atención sobre algunas de las tensiones entre la racionalidad tecnológica, por un lado, y la diversidad lingüística y la misión de la educación superior, por el otro.

En consecuencia, la problemática que guía esta investigación puede expresarse de la siguiente manera: ¿en qué medida afecta la inteligencia artificial generativa la construcción del discurso académico universitario a partir del procesamiento de la identidad lingüística de los estudiantes y de la diversidad dialectal en general?

Interrogante que se incrusta en una lucha más amplia sobre el lenguaje, los algoritmos y el poder, ya que el discurso académico se convierte en un espacio para la negociación de legitimidades, tipos de conocimiento y formas de representar la cultura (Crosthwaite, 2025; Fairclough, 1995; Van Dijk, 2000).

El objetivo principal fue analizar los riesgos de la homogeneización discursiva desencadenada por el sesgo de uso en la escritura académica mediada por máquina a través de la inteligencia artificial, con referencia particular a su influencia en la identidad lingüística y la variación dialectal en los marcos de la educación superior.

Para lograr este propósito, se investigó la percepción de los estudiantes, así como el *output* textual que representa patrones de adhesión discursiva. Además, este estudio se orientó hacia dos consultas principales:

1. ¿Cuáles son las ventajas y desventajas que los estudiantes perciben respecto al uso de herramientas de IA/IAG en la producción académica?
2. ¿De qué manera difieren los textos generados por IA, en relación con la originalidad de contenidos, la variación léxica y el estilo discursivo, de los textos no mediáticamente algorítmicos sobre los mismos temas?

Esto permitió vincular los hallazgos de datos cuantitativos de la encuesta que involucró a 682 estudiantes con los datos cualitativos y su enfoque en el estudio comparativo de textos, lo cual ayudó a obtener una imagen más completa de las tensiones entre la creatividad, la variedad lingüística y la homogeneización algorítmica.

1. Lenguaje, algoritmos y poder

El lenguaje no es solo una herramienta de comunicación, también es un escenario de lucha donde ocurre la producción y legitimación del conocimiento. El análisis crítico del discurso se ha centrado en las prácticas comunicativas que funcionan reproduciendo estructuras de poder e ideologías dominantes a través del desarrollo de marcos de interpretación que validan ciertas visiones del mundo (Fairclough, 1995; Van Dijk, 2000,

2017). De esta manera, se configuran nuevas formas de dominación simbólica y de estratificación social (Bourdieu, 1991).

Esta tensión se intensifica en la era posdigital, en donde la mediación algorítmica interviene directamente en la producción discursiva, pues los sistemas de inteligencia artificial, lejos de ser neutrales, se entrenan sobre corpus que incorporan criterios de selección y jerarquización, y determinan qué discursos adquieren visibilidad y cuáles quedan relegados (Noble, 2018; Blodgett et al., 2020; Crawford, 2021).

El surgimiento de modelos de lenguaje en el mercado ha añadido una nueva dinámica a este panorama, dado que su rol como nuevas “autoridades cognitivas” puede imponer modos legítimos de argumentación y escritura (Floridi & Chiriatti, 2020). La capacidad de generar textos decentes y coherentes a gran escala es impresionante y su intervención en la producción de discurso implícitamente establece un estándar de “corrección”, estilo y legitimidad. Por lo que quien controla el acceso al discurso controla el acceso al conocimiento (Van Dijk, 2017).

El algoritmo funciona como un filtro epistémico, es decir, a algunas perspectivas se les otorga visibilidad y a otras se las silencia (Markl, 2022), proceso que moldea la manera en que los sujetos producen y negocian significados tanto dentro de un contexto académico como en el campo social e impacta en la circulación del conocimiento. En el contexto académico, esto crea tipos específicos de tensiones: las instituciones pueden apoyar las actividades de escritura proporcionando acceso a herramientas algorítmicas, así como estructuras para organizar contenidos textuales; mientras que dicho *software* tiende simultáneamente a estandarizar estilos discursivos, lo cual hace que los registros y las voces sean menos dispares entre sí (Von Stecher, 2025). Por ello, la relación lenguaje/algoritmos/poder está inscrita en una controversia sobre diversidad expresiva versus estandarización tecnológica, que invita a la lingüística y la educación a realizar contribuciones críticas.

2. Sesgo algorítmico y sesgo de uso

El concepto de sesgo algorítmico en la tecnología del lenguaje ha recibido más atención en el campo de la IA aplicada al lenguaje. Por ejemplo, los problemas de sesgo en el procesamiento del lenguaje natural (PLN) se entienden como distorsiones sistémicas que engloban en el resultado principalmente los idiomas aprendidos de los datos de entrenamiento o decisiones algorítmicas implementadas, tomadas tanto al modelar como al usar algoritmos (Gallegos et al., 2024; Blodgett et al., 2020; Hovy & Prabhumoye, 2021).

Estos sesgos ven los moldes del lenguaje que favorecen variedades particulares del habla como el uso de los conectores “por lo tanto” y “en consecuencia” en lugar de “pues” o de “entonces” por “de esta manera” y “por consiguiente”. Por ejemplo, Curry *et al.* (2024) mencionan que los sesgos ideológicos integran narrativas dominantes, como el nacionalismo duro, lo que significa que esos países occidentales que se han llamado a sí mismos democracias occidentales durante algún tiempo, ahora están siendo cada

vez más inquisitivos promoviendo sesgos culturales contra la alienación de idiomas y comunidades no estandarizadas (Bender et al., 2021; Crawford, 2021; Noble, 2018).

Un ejemplo familiar son los sistemas de reconocimiento automático de voz, que funcionan mejor con formas estándar del inglés con altos niveles de precisión, pero no son capaces de reconocer texto hablado de dialectos regionales o minoritarios (Markl, 2022; Östling & Kurfali, 2023). Cabe destacar que esta lengua es el recurso de programación principal de estos modelos de inteligencia. En este sentido, el texto que producen estas inteligencias está conformado por estilos académicos claramente anglocéntricos, cuya forma de construcción del discurso no aprecia la diversidad cultural y lingüística de los usuarios (Gallegos et al., 2024).

El sesgo de uso surge dentro de este marco como un tema que merece una exploración adicional, dado que, a diferencia de aquellos en el entrenamiento que se originan naturalmente del modelo, este se manifiesta durante las interacciones cotidianas entre sistemas algorítmicos y seres humanos. Sin embargo, la mayoría de los académicos y estudiantes, cuando utilizan modelos como *ChatGPT* para escribir, pasan mucho tiempo replanteando estructuras prefabricadas, lo cual resulta “crucial” para sus escritos, dado que estos modelos de inteligencia utilizan conectores y estructuras discursivas estandarizadas que hacen predecibles sus producciones (Bender et al., 2021).

Este fenómeno tiene implicaciones directas para la producción de textos académicos. Por un lado, ofrece un medio eficaz para producir rápidamente texto bien organizado con temas claramente definidos; pero, por otro lado, también dificulta la creatividad y la diversidad en la expresión, lo cual da lugar a un estándar discursivo uniforme. Esta práctica erosiona la capacidad crítica de los estudiantes, ya que, como afirma Lu (2025), se cede más autoridad en el proceso de escritura a la “voz algorítmica” y menos a la exploración personal del propio discurso. Así, en términos psicociolingüísticos, el sesgo de uso se desarrolla en un proceso de dependencia donde la autonomía expresiva se ve disminuida y el modelo de escritura común se solidifica.

Vale la pena señalar que el sesgo de uso afecta no solo a individuos aislados, sino también a toda la comunidad académica. A medida que los textos modelados según algoritmos proliferan peligrosamente, se introduce un tipo de escritura que valora más la corrección formal sobre la creatividad estilística (Okhremenko et al., 2025). Esto va de la mano con una reducción tanto en el número como en la variedad de lenguajes utilizados en las universidades. Por lo tanto, identificar y formular preguntas sobre el sesgo de uso es clave si los algoritmos deben moldear no solo lo que la gente dice, sino cómo aprenden a decir las cosas.

3. Identidad lingüística y diversidad dialectal

La lengua es un marcador central de identidad, pues vehicula pertenencias sociales, históricas y culturales. Desde la sociolingüística crítica se ha subrayado que toda práctica lingüística está vinculada con relaciones de poder, legitimidad y exclusión (Bourdieu, 1991;

Budiarsa, 2017). En consecuencia, la variación lingüística no es un mero fenómeno descriptivo, sino un espacio donde se disputan jerarquías sociales y se configuran representaciones simbólicas de las comunidades (Labov, 1972; Bleortu & Cuevas-Alonso, 2013).

En el contexto colombiano, esta diversidad se manifiesta en la coexistencia de múltiples dialectos y registros que reflejan la pluralidad regional del país. Los estudiantes universitarios, provenientes de regiones como el Caribe, el Pacífico, los Andes o la Orinoquía, aportan a los espacios académicos una riqueza dialectológica que mejora las prácticas discursivas y promueve un aprendizaje intercultural (Contreras et al., 2022). Preservar esta pluralidad supone reconocer que las formas locales de hablar y escribir son igualmente legítimas frente a los modelos estandarizados promovidos por la globalización y, ahora, por la inteligencia artificial (UNESCO, 2025).

Sin embargo, esta diferencia está siendo socavada por el uso creciente de instrumentos algorítmicos en el trabajo académico, al favorecer un tratamiento discursivo «no acentuado» y globalizado, puesto que la acción de los modelos de inteligencia artificial parece invisibilizar las marcas de variación dialectal y producir una homogeneidad en el discurso que se cuela como universal pero que, de hecho, simplemente reproduce con ligeras modificaciones algunas formas preeminentes conocidas por ellos (Markl, 2022; Gallegos et al., 2024). Esta práctica restringe tanto la expresión como la transmisión del conocimiento indígena y el patrimonio lingüístico.

Este respeto, como ha señalado la UNESCO (2015, 2021), se refleja en el reconocimiento de la diversidad lingüística como un derecho cultural y como una dimensión imperativa dentro de la educación inclusiva y de calidad. La homogeneización discursiva facilitada algorítmicamente no puede interpretarse como un problema técnico exclusivo, sino que es un fenómeno político que tiene implicaciones epistémicas para la pluralidad y la equidad en la representación comunitaria (Mezyed, 2024; Rasmey, 2018).

Por consiguiente, examinar la relación entre identidad lingüística y uso de IA en la escritura académica (Okhremenko et al., 2025) implica reconocer una tensión evidente entre la eficiencia tecnológica y la diversidad cultural (Contreras, 2025; Contreras et al., 2022), que solo puede ser tratada mediante estrategias pedagógico-didácticas con sentido crítico, que no repriman ninguna variación lingüística y que, por el contrario, resistan ante la estandarización algorítmica que proponen los nuevos paradigmas tecnológicos en ámbitos educativos.

4. Metodología

Esta investigación parte del enfoque mixto como proponen Creswell y Creswell (2018). En primer lugar, justifica el propósito de análisis documental a través de diferentes fuentes bibliográficas; en segundo lugar, para explorar y comparar los diferentes fenómenos implicados en diversos niveles de producción textual mediante IA. En este caso, no solo se deseaba medir la prevalencia y las actitudes hacia el uso de IAG, sino también interpretar el impacto de tales prácticas en los procesos de producción textual en estudiantes universitarios.

La muestra estuvo compuesta por 682 estudiantes de diferentes facultades de la Universidad de Pamplona, seleccionados mediante muestreo incidental no probabilístico, debido a que nuestra intención fue abordar estudiantes de diferentes regiones de Colombia, por lo que se pudo dar cuenta de un rango representativo de diversidad dialectológica y sociolingüística.

Además, dentro de este estudio se consideró aplicar un matiz de análisis sociolingüístico y determinar aspectos esenciales sobre variación y cambio lingüístico (Labov, 1972; Bleortu & Cuevas-Alonso, 2013), lo que resulta pertinente dado los numerosos aportes que la investigación sociolingüística ha proporcionado respecto a la identificación de variaciones y cambios lingüísticos en contexto que promueven la identidad.

Se emplearon dos instrumentos clave en esta investigación:

1. Encuesta digital: preparada en *Google Forms* con preguntas cerradas y de opción múltiple sobre el uso de IA/IAG, frecuencia de uso, razones para usarlas, percepciones sobre su impacto en las competencias comunicativas y desventajas percibidas (dependencia, restricción de las habilidades de pensamiento crítico y homogeneización). Esta versión del instrumento, muy popular en la investigación educativa y sociolingüística, puede tener como objetivo la recopilación sistémica de datos adecuadamente caracterizados (Arias, 2023).
2. Corpus académico: para el análisis cualitativo del corpus académico se seleccionó un conjunto comparativo de alrededor de 64 textos producidos por estudiantes en contextos evaluativos equivalentes y correspondientes a actividades de escritura descriptiva y expositiva desarrolladas en cursos de habilidades comunicativas. El corpus estuvo conformado por textos redactados sin mediación algorítmica y por versiones elaboradas con asistencia de IA, lo que permitió un análisis contrastivo (McEnery & Hardie, 2012).

La recopilación de datos tuvo lugar en el segundo semestre de 2024 y primer semestre de 2025. Durante la etapa cuantitativa, los datos de la encuesta fueron categorizados y presentados junto a estadísticas descriptivas (frecuencias, porcentajes).

Los criterios de análisis incluyeron: (i) riqueza léxica, comprendida como la diversidad y variabilidad del vocabulario utilizado, así como el uso de sinónimos y expresiones que no se encuentran estandarizadas; (ii) variación dialectal, correspondiente a la presencia de regionalismos, marcas culturales y giros lingüísticos propios de la identidad y variación sociolingüística de los estudiantes; y (iii) recurrencia de fórmulas discursivas, relacionada con aspectos de repetición, conectores, estructuras sintácticas y secuencias argumentativas propias de la escritura académica generada por IA. Estos criterios permitieron sistematizar el análisis y fortalecer la replicabilidad del estudio (Moreira et al., 2021).

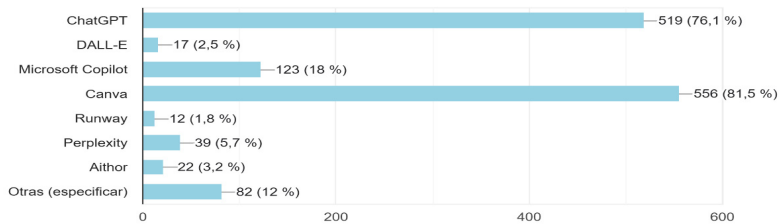
La triangulación de métodos, como señalan Creswell y Creswell (2018), ofrece mayor validez a los hallazgos al contrastar distintas fuentes de evidencia. De esta manera, se buscó identificar no solo patrones de uso, sino también tensiones entre la eficiencia algorítmica y la preservación de la identidad lingüística.

4. Resultados

Los resultados de la encuesta revelan un uso extendido de herramientas de inteligencia artificial generativa: el 96,5 % de los estudiantes afirmaron emplearlas en su proceso de formación académica. En cuanto a la frecuencia, la mayoría indicó hacerlo de manera semanal (69,9 %), mientras que un 9,4 % las utiliza diariamente y un 17,6 % de forma ocasional.

En relación con los procesos de producción escrita, destacan modelos de IA como *ChatGPT* (76,1 %), seguida por *Microsoft Copilot* (18 %) y otras en menor proporción como *Gemini*, *Claude* y *Meta AI*. Hallazgos que evidencian una fuerte penetración de los sistemas algorítmicos en las prácticas académicas cotidianas y que destacan cierta dependencia lingüística como se revela en la Figura 1.

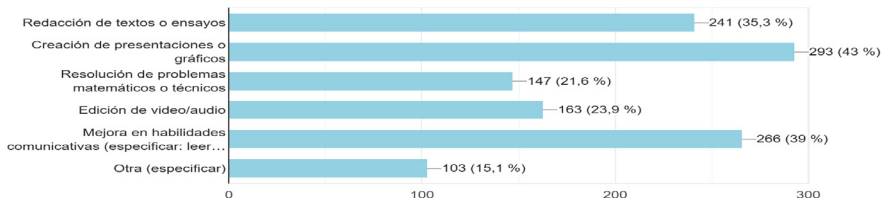
¿Qué herramientas de IA/IAG utiliza con mayor frecuencia? (Seleccione todas las que apliquen)
682 respuestas



Fuente: elaboración propia

Figura 1. Modelos de IA y frecuencia de uso

¿Con qué finalidades usa herramientas de IA/IAG? (Seleccione todas las que apliquen)
682 respuestas



Fuente: elaboración propia

Como se pudo apreciar en la Figura 2, los estudiantes usan en gran medida cualquier modelo de IA para la mejora de sus habilidades comunicativas (leer, escribir, hablar, escuchar) con 266 respuestas (39 %). Este dato es relevante porque muestra que, en cierta manera, no se limitan a la automatización de tareas, sino que buscan en la IA un aliado para desarrollar competencias lingüísticas.

Adicionalmente, se resalta que la redacción de textos o ensayos ocupa el tercer lugar con 241 respuestas (35,3 %), lo que lleva a asumir que existe una fuerte dependencia de la IA para la producción textual, hecho que puede facilitar la escritura, pero también promover la homogeneización discursiva.

Se considera un hallazgo relevante que el 69,8 % de los estudiantes percibe que la IA ha facilitado significativamente su trabajo académico, mientras que un 26,1 % considera que el apoyo ha sido solo parcial. Sin embargo, cuando se indaga sobre las desventajas, emergen preocupaciones consistentes: el 63,5 % reconoció una limitación en el uso del pensamiento crítico y el 59,7 % señaló un riesgo de dependencia tecnológica.

En términos de escritura, el 61 % de los encuestados indicó que el impacto de la IA ha sido mínimo o inexistente, mientras que apenas un 19,6 % reportó mejoras significativas. Algo similar ocurre con la lectura y el análisis de textos complejos, pues el 56,2 % no ha notado cambios sustanciales. Estas cifras reflejan una tensión entre la percepción de utilidad general y el impacto real en el desarrollo de competencias comunicativas.

El contraste entre textos elaborados con y sin mediación de IA demuestra una clara tendencia a la homogeneización discursiva, tal como se observa en la Tabla 1.

Tabla 1. Textos redactados frente a su versión mediante IA

<i>Ejemplo</i>	<i>Texto en papel (rasgos)</i>	<i>Texto digital con mediación en IA (rasgos)</i>	<i>Observaciones clave</i>
<i>Describe en qué consiste el festival de la Feria de las Flores (Medellín)</i>	Expresiones coloquiales: paisas, caldenses, criollo, campesinos... y comparaciones culturales espontáneas (“similar a la feria de los tulipanes en Turquía”).	Lenguaje académico y descriptivo, con adjetivos formales: los locales, ciudadanos, autóctonos... Uso de conectores como: cabe destacar que, en otras palabras.	Pérdida de comparaciones originales; homogeneización hacia tono académico.

Describe el Festival de la Leyenda Vallenata (Valledupar)	Referencias globales creativas (“Tomorrowland”, “Viña del Mar”). Terminología específica: parranda, piquería, verseadores, parrandón, rey vallenato...).	Terminología técnica y precisa (“versadores”, “acordeonero” otros como: festividad, evento, actividades...). Tono institucional y conectores como, por lo tanto, en consecuencia, o conclusión	Se eliminan referentes culturales espontáneos; prevalece un discurso estándar.
Describe en qué consiste la Semana Santa (en Pamplona o Popayán)	Explicaciones básicas (“una ciudad llamada Popayán o Pamplona”). Repeticiones. Narrativa sencilla y riqueza léxica como: procesiones, nazarenos, pasos, desande...	Cohesión formal, lenguaje valorativo (“procesiones solemnes”, “relevantes”). Reconocimiento patrimonial reforzado.	La IA ordena y jerarquiza, pero se pierde espontaneidad y naturalidad estudiantil.
Describe en qué consiste la Feria de Cali (Cali)	Estilo narrativo cercano, con marcas locales (salsódromo, folclore, salsa choque, brava, caleña, rumba, trapiches...). Redacción con poca cohesión.	Discurso turístico-académico, con precisión temporal (“cada diciembre”). Uso de adjetivos institucionales y definiciones estandarizadas: estilos de salsa, festividad, atuendos llamativos, tradiciones, folclore local...	Se reemplaza la expresividad local por fórmulas homogéneas de promoción cultural.

Fuente: elaboración propia

Como se pudo apreciar, los textos originales conllevan grandes cargas léxicas propias de la variación lingüística colombiana, mientras que el uso de IA parte de una adaptación semántica de los enunciados y se apoya en el uso de conectores estándar (“por lo tanto”, “al parecer”), además, elimina la calidez y cercanía del sujeto dentro del mensaje, lo cual convierte el contenido en un tono impersonal y generalizante, y en general se percibe escasez de presencia de marcas dialectales o expresiones propias de la identidad regional.

Se asume que los textos producidos sin apoyo algorítmico mostraron una mayor diversidad léxica, un uso más flexible de los conectores y, en algunos casos, la aparición de giros lingüísticos propios de la oralidad o de la variedad dialectal de los estudiantes.

Este contraste confirma la existencia del sesgo de uso, en el cual los usuarios adoptan sin cuestionamiento las fórmulas ofrecidas por la IA, situación que homogeneiza su escritura y reduce la creatividad discursiva (Elastic, 2024; Morales, 1997).

5. Discusión

Los resultados obtenidos permiten evidenciar cómo la inteligencia artificial generativa se ha convertido en un mediador central de las prácticas académicas universitarias. El alto nivel de uso reportado (96,5 %) indica que los estudiantes integran estas herramientas en su cotidianidad académica, lo que concuerda con estudios recientes que muestran la penetración global de la IA en los procesos de escritura y aprendizaje (Lu, 2025; Chen et al., 2024). Sin embargo, dado que las tareas más comunes son escribir textos, preparar presentaciones y estructurar ideas, podemos distinguir que la IA es menos un acólito mediático que un artilugio discursivo (Ciesla, 2024). Este resultado está en relación directa con lo que afirmaron Fairclough (1995) y Van Dijk (1998) respecto a la función del discurso en la reproducción de las relaciones de poder.

Al aceptar las fórmulas de salida de la IA como un «modelo», los estudiantes reflejan lo que se ha denominado el sesgo de uso, que resulta en un aplanamiento de los textos. Los datos cualitativos corroboran esta observación al revelar estructuras impersonales y predecibles en el texto apoyado por IA (Amaratunga, 2023), con un agotamiento resultante de la variedad léxica y la creatividad expresiva. En este sentido, los algoritmos actúan como nuevas “autoridades cognitivas” (Floridi & Chiriatti, 2020), capaces de fijar criterios de corrección y legitimidad en la escritura académica.

Desde una perspectiva sociolingüística, el riesgo más preocupante a causa de los modelos tecnológicos es el declive de la identidad lingüística de los usuarios de habla hispánica (Iglesias, 2025; Flores, 2023). Mientras los textos producidos sin mediación algorítmica muestran marcas de oralidad y variación dialectal, los generados con IA tienden a borrar dichas huellas e imponen una escritura “neutral” que responde a estándares globalizados y desconocen su procedencia (Huang et al., 2025; De La Rosa et al., 2022). Esto confirma lo advertido por Muñoz-Basols *et al.* (2010) y por la UNESCO (2021, 2025) sobre la amenaza de la estandarización tecnológica frente a la pluralidad lingüística. La dependencia señalada por los estudiantes (59,7 %) y la percepción de disminución del pensamiento crítico (63,5 %) refuerzan esta idea: el sesgo de uso no solo limita la autonomía discursiva, sino que también empobrece la diversidad cultural que se expresa en la lengua.

En términos sociolingüísticos, estos resultados sugieren que la IA no es simplemente un apoyo para el aprendizaje, sino un agente de socialización discursiva que moldea los procesos de aprendizaje de los sujetos para escribir y argumentar. Aunque esta eficiencia algorítmica implica beneficios obvios para la rapidez y el ordenamiento, también representa una paradoja: ¿cómo puede la IA integrarse en las universidades sin estandarizar los lenguajes o la capacidad crítica en los estudiantes? Como subrayan

Crawford (2021) y Noble (2018), la solución reside en una pedagogía crítica que aborde el sesgo, promueva el uso consciente de estas herramientas y defienda la diversidad epistémica (Contreras, 2026).

En este sentido, el análisis de los resultados revela que una homogeneización discursiva basada en el sesgo de uso sí representa una amenaza real para el mantenimiento de la identidad lingüística dentro de estos diversos entornos universitarios, como señalan Bender *et al.* (2021), Curry *et al.* (2024), Blodgett *et al.* (2020). Dados estos hallazgos, argumentamos que, según exploran Crompton y Burke (2023), y Crompton *et al.* (2024), el dilema que enfrenta la educación superior es tanto una cuestión de prohibir el uso de la IA como una educación en la crítica y el uso reflexivo que permita aprovechar su potencial sin sacrificar la riqueza dialectológica y la autonomía expresiva de los estudiantes (Lovón & Parini, 2024; Muñoz-Basols *et al.*, 2024).

6. Conclusiones

La investigación confirmó que la inteligencia artificial generativa se está convirtiendo en una parte integral de la vida académica en la universidad, puesto que la mayoría de los estudiantes (96,5 %) la usan regularmente al hacer tareas de escritura, crear presentaciones y estructurar ideas. Desde el punto de vista de los estudiantes, estos instrumentos sin duda apoyarán sus estudios y favorecerán un mayor grado de caminos de aprendizaje autodirigidos y autogestionados.

Al mismo tiempo, se determinó que también se ha planteado una considerable preocupación con respecto a la dependencia tecnológica y la disminución de la reflexión crítica, dado que puede haber una necesidad de esfuerzos pedagógicos dirigidos a fomentar un uso crítico y reflexivo de la IA.

En cuanto a la comparación de textos, se aprecia una gran diferencia. La escritura mediada algorítmicamente muestra estructuras similares, menos variación en el tono y el vocabulario, mientras que los textos no escritos por IA son más diversos y creativos (y a veces incluyen palabras dialectales que expresan las identidades lingüísticas de los estudiantes). Esto atestigua la existencia de un sesgo de uso que homogeniza el discurso académico y suprime la diversidad lingüística.

En este sentido, los hallazgos de la investigación trascienden el contexto institucional específico y permiten comprender el sesgo de uso como una categoría analítica extrapolable a otros escenarios de educación superior, particularmente en contextos latinoamericanos y multilingües. La evidencia empírica sugiere que la inteligencia artificial generativa no actúa únicamente como una herramienta técnica de apoyo, sino como un agente de socialización discursiva que contribuye a la normalización de estilos, estructuras y registros académicos. Desde una perspectiva sociolingüística crítica, este fenómeno plantea implicaciones teóricas relevantes para el estudio de la relación entre lenguaje, poder y tecnología, al mostrar cómo la mediación algorítmica

incide en la producción del conocimiento y en la configuración de identidades lingüísticas en la escritura académica contemporánea.

Este estudio supone que la idea no es excluir la IA de la práctica educativa, sino incluirla de manera crítica y reflexiva. El rescate de la diversidad dialectal y la identidad lingüística en la educación superior implica la construcción de tácticas pedagógicas que aumenten la conciencia sobre los prejuicios algorítmicos y capaciten a los estudiantes para mejorar su autonomía discursiva.

Finalmente, se considera que la introducción de la inteligencia artificial puede ofrecer una oportunidad para enriquecer las prácticas comunicativas y culturales, en lugar de empobrecerlas.

Conflictos de intereses:

El autor declara que no tiene ningún conflicto de intereses que pueda influir en la imparcialidad, objetividad o en la información presentada en este artículo de investigación.

Financiamiento:

La investigación fue financiada con recursos propios del autor.

Implicaciones éticas:

Esta investigación garantizó el consentimiento informado de todos los participantes, protegió su privacidad y confidencialidad mediante el manejo anónimo de los datos.

Declaración de datos:

Este artículo no tiene datos abiertos asociados a la investigación.

Referencias

- Amaratunga, T. (2023). *Understanding Large Language Models: Learning Their Underlying Concepts and Technologies*. Apress. <https://doi.org/10.1007/979-8-8688-0017-7>
- Arias, F. (2023). Investigación documental, investigación bibliométrica y revisiones sistemáticas. *REDHECS: Revista Electrónica de Humanidades, Educación y Comunicación Social*, 31(22), 9-28.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be too Big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM Digital Library. <https://doi.org/10.1145/3442188.3445922>

- Bleortu, C., & Cuevas-Alonso, M. (2013). Sali A. Tagliamonte: Variationist Sociolinguistics. Change, Observation, Interpretation. *Onomázein*, (28), 42-49. <https://doi.org/10.7764/onomazein.28.21>
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (Technology) Is Power: A Critical Survey of “Bias” in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454-5476). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.485>
- Bourdieu, P. (1991). *Language and Symbolic Power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, Trans.). Polity Press and Basil Blackwell.
- Budiarsa, I.M. (2017). Language, Dialect and Register Sociolinguistic Perspective. *RETORIKA: Jurnal Ilmu Babasa*, 1(2), 379-387. <https://doi.org/10.22225/jr.1.2.42.379-387>
- Chen, Z., Xu, L., Zheng, H., Chen, L., Tolba, A. et al. (2024). Evolution and Prospects of Foundation Models: From Large Language Models to Large Multimodal Models. *Computers, Materials & Continua*, 80(2), 1753-1808. <https://doi.org/10.32604/cmc.2024.052618>
- Cheung, L., & Crosthwaite, P. (2025). *CorpusChat: Integrating Corpus Linguistics and Generative AI for Academic Writing Development*. *Computer Assisted Language Learning*, 1-27. <https://doi.org/10.1080/09588221.2025.2506480>
- Giesla, R. (2024). Developments in Artificial Intelligence and Linguistics. In *The Book of Chatbots* (pp. 11-39). Springer Nature. https://doi.org/10.1007/978-3-031-51004-5_2
- Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press. <https://doi.org/10.2307/j.ctv1ghv45t>
- Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). Sage Publications.
- Contreras, A. R. (2024). Didáctica de la fraseología en la educación universitaria mexicana y el contraste desde la cultura colombiana. En A. R. Contreras, J. F. Díaz & Gómez (eds.), *Fraseología, oralidad y programación neurolingüística* (pp. 23-57). Unipamplona. <https://doi.org/10.24054/seu.53>
- Contreras, A. R. (2025). La construcción discursiva en estudiantes de educación superior a través del u-learning. En S.L. Dykinson (ed.), *Educación y humanidades como ejes de investigación e innovación* (pp. 87-97). Dykinson EBook. <https://doi.org/10.14679/4092>

- Contreras, A. R. (2026). La tecnofobia como narrativa emergente frente a la regulación de la inteligencia artificial. *Palabra Clave* 28(3), e2836. <https://doi.org/10.5294/pa-cla.2025.28.3.6>
- Contreras, A. R., Díaz González, S. P., & Díaz González, J. F. (2022). Digital Texts as a Teaching Alternative of the Mother Language. *Cuadernos de Lingüística Hispánica*, (39), 1-18. <https://doi.org/10.19053/0121053X.n39.2022.13474>
- Crompton, H., & Burke, D. (2023). Artificial Intelligence in Higher Education: The State of the field. *International Journal of Educational Technology in Higher Education*, 20(1), 22. <https://doi.org/10.1186/s41239-023-00392-8>
- Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English Language Teaching: Affordances and Challenges. *British Journal of Educational Technology*, 55(6), 2503-2529. <https://doi.org/10.1111/bjet.13460>
- Curry, N., Baker, P., & Brookes, G. (2024). Generative AI for Corpus Approaches to Discourse Studies: A Critical Evaluation of ChatGPT. *Applied Corpus Linguistics*, 4(1), 100082. <https://doi.org/10.1016/j.acorp.2023.100082>
- De la Rosa, J., Ponferrada, E. G., Villegas, P., González de Prado Salas, P., Romero, M., & Grandury, M. (2022). BERTIN: Efficient Pre-training of a Spanish Language Model Using Perplexity Sampling. *Procesamiento del Lenguaje Natural*, 68, 13-23. <https://doi.org/10.48550/arXiv.2207.06814>
- Elastic. (2024). ¿Qué es el procesamiento de lenguaje natural? *Elastic*. <https://www.elastic.co/es/what-is/natural-language-processing>
- Fairclough, N. (1995). Critical Discourse Analysis: *The Critical Study of Language*. Longman.
- Flores Borjabad, S. A. (2023). La inteligencia artificial y el lenguaje natural: hacia el desarrollo de un lenguaje autónomo. En C. G. Gallego Real (ed.), *Nuevas teorías y aproximaciones a estudios sobre lengua, lingüística y traducción* (pp. 55-70). Dykinson.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 681-694. <https://doi.org/10.1007/s11023-020-09548-1>
- Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics*, 50(3), 1097-1179. https://doi.org/10.1162/coli_a_00524

- Hovy, D., & Prabhumoye, S. (2021). Five Sources of Bias in Natural Language Processing. *Language and Linguistics Compass*, 15(8), e12432. <https://doi.org/10.1111/lnc3.12432>
- Huang, L., Lin, Z., Zhang, H., & Zhang, Y. (2025). Melting Pot: Linguistic Diversity and Collaboration. *Journal of Economic Behavior & Organization*. Advance online publication. <https://doi.org/10.1016/j.jebo.2025.1726>
- Iglesias Sánchez, A. B. (2025). Regulación del español en la inteligencia artificial: un nuevo giro en la discursividad panhispánica. *Lexis*, 49(1), 135-164. <https://doi.org/10.18800/lexis.202501.005>
- Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.
- Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the Language Classroom: A Systematic Review. *Interactive Learning Environments*, 34(1), 1-25. <https://doi.org/10.1080/10494820.2025.2498537>
- Lovón, M., & Parini, A. (2024). Inteligencia artificial, discurso digital y lenguaje: la lingüística en la era de las nuevas tecnologías. *Lengua y Sociedad*, 23(2), 577-598. <https://doi.org/10.15381/lengsoc.v23i2.29773>
- Lu, C. (2025). AI-Generated Corpus Learning and EFL Learners' Learning of Grammatical Structures, Lexical Bundles, and Willingness to Write. *PLoS one*, 20(7), e0321544. <https://doi.org/10.1371/journal.pone.0321544>
- Markl, N. (2022). Language Variation and Algorithmic Bias: Understanding Algorithmic Bias in British English Automatic Speech Recognition Systems. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1118-1132). <https://doi.org/10.1145/3531146.3533117>
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- Morales Ascencio, B. (1997). La lingüística en el contexto de la inteligencia artificial. *Forma y Función*, (10), 25-50. <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/17072>
- Moreira, D., Cruz, I., Gonzalez, K., Quirumbay, A., Magallan, C., Guarda, T., Andrade, A., & Castillo, C. (2021). Análisis del estado actual de procesamiento de lenguaje natural. *RISTI: Revista Ibérica de Sistemas e Tecnologías de Informação*, E42, 126-136. <https://www.risti.xyz/issues/ristie42.pdf>
- Mezyed Rivas, R. (2024). *La variedad lingüística de la inteligencia artificial conversacional en el aula de ELE* [Trabajo de grado]. Lund University. <http://lup.lub.lu.se/student-papers/record/9147577>

- Muñoz-Basols, J., Palomares Marín, M. del M., & Moreno Fernández, F. (2024). El Sesgo Lingüístico Digital (SLD) en la inteligencia artificial: implicaciones para los modelos de lenguaje masivos en español. *Lengua y Sociedad*, 23(2), 623-647. <https://doi.org/10.15381/lengsoc.v23i2.28665>
- Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. <https://doi.org/10.2307/j.ctt1pwt9w5>
- Okhremenko, I., Vasyukov, V., Shapauov, A., Kochetkov, E., & Panova, E. (2024). Reducir las diferencias sociales con la IA: mejorar el acceso a la terminología lingüística especializada. *Interacción y Perspectiva: Revista de Trabajo Social*, 15(1), 98-108. <https://doi.org/10.5281/zenodo.14031159>
- Östling, R., & Kurfalt, M. (2023). Language Embeddings sometimes Contain Typological Generalizations. *Computational Linguistics*, 49(4), 1003-1051. https://doi.org/10.1162/coli_a_00491
- Rasmy, M. (2018). The World Bank's Support to Higher Education Reform in Egypt: Educators' Perspective on its Impact for Quality and Equality [Master's thesis], The American University in Cairo, Egypt.
- UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
- UNESCO. (2025). *Languages Matter: Global Guidance on Multilingual Education*. UNESCO Publishing. <https://doi.org/10.54675/MLIO7101>
- Van Dijk, T. A. (2000). *Ideology: A Multidisciplinary Approach*. Sage. <https://doi.org/10.4135/9781446217856>
- Van Dijk, T. A. (2017). *Discourse and Power*. Bloomsbury Publishing
- Von Stecher, P. (2025). Las nuevas irregularidades del lenguaje. Desafíos de interpretación y mecanismos de simplificación discursiva de la inteligencia artificial. *Forma y Función*, 38(1). <https://doi.org/10.15446/fyf.v38n1.114947>
- World Economic Forum. (2024, 10 de octubre). *La oportunidad perdida con la brecha de diversidad lingüística de la IA*. Foro Económico Mundial. <https://es.weforum.org/stories/2024/10/la-oportunidad-perdida-con-la-brecha-de-diversidad-linguistica-de-la-ia/>