



INTEGRACIÓN DE MINERÍA DE TEXTO Y ANÁLISIS DE REDES SOCIALES COMPLEJAS Para la Caracterización Cuantitativa de la Evolución Discursiva en Inteligencia Artificial a través de Plataformas de Streaming

FRED TORRES-CRUZ¹, YUDI JANEH YUCRA-MAMANI¹, VLADIMIRO IBANEZ-QUISPE¹, MARCO FIDEL MAYTA-
QUISPE¹

¹ Universidad Nacional del Altiplano Puno, P.O. Box, 291, Puno-Perú

PALABRAS CLAVE

*Inteligencia Artificial
Evolución Discursiva
Minería de Texto
Análisis de Redes Sociales
Plataformas de Streaming*

RESUMEN

Este estudio ofrece una caracterización cuantitativa de la evolución discursiva en Inteligencia Artificial a través de plataformas de streaming (Youtube), integrando minería de texto y análisis de redes sociales complejas. Mediante un corpus extenso obtenido, se implementaron algoritmos de procesamiento de lenguaje natural para analizar el contenido textual, identificando patrones, temáticas emergentes y cambios en los discursos sobre IA. Se realizó un análisis de redes sociales para examinar las estructuras de interacción y la influencia de distintos actores en la divulgación de información. Los resultados muestran tendencias significativas en la presentación y percepción de la IA, destacando la evolución de temas específicos, diferencias en la percepción entre diversos grupos, y la influencia de factores como avances tecnológicos y eventos globales. Este trabajo proporciona una comprensión profunda de la comunicación y percepción de la IA, ofreciendo perspectivas valiosas para académicos, comunicadores y formuladores de políticas en el campo de la IA.

Recibido: 20 / 03 / 2024

Aceptado: 18 / 07 / 2024

1. Introducción

La Inteligencia Artificial (IA) ha emergido como una de las fuerzas más transformadoras en la era digital, remodelando industrias, procesos y la manera en que interactuamos con la tecnología. A medida que su aplicación se expande, también lo hace el discurso público y académico que la rodea, evolucionando con cada avance tecnológico y su integración en la vida cotidiana. Sin embargo, la caracterización de esta evolución discursiva, especialmente en plataformas digitales de amplio alcance como las de *streaming*, ha presentado desafíos significativos debido a la vastedad y dinamismo de los datos generados (Morchid et al., 2013). Este estudio se enfoca en abordar dichos desafíos, ofreciendo una visión cuantitativa y analítica de cómo ha evolucionado el discurso sobre la IA en estas plataformas.

Con la proliferación de plataformas de *streaming* como YouTube, Twitch, y otras, el volumen de contenido relacionado con la IA y la manera en que se presenta y se percibe ha sufrido cambios notables. Estas plataformas no solo sirven como medios para la divulgación de información sino también como espacios para el debate público, la construcción de comunidades y la formación de opiniones (Campis, 2023). Por ende, una comprensión detallada de la evolución discursiva en estas plataformas puede proporcionar perspectivas valiosas sobre la percepción pública y la comunicación en torno a la IA.

Para abordar este objetivo, el presente estudio integra metodologías de minería de texto y análisis de redes sociales complejas, aprovechando el poder computacional y los avances en el procesamiento del lenguaje natural (PLN) para analizar extensos conjuntos de datos textuales (Li et al., 2024). La minería de texto permite identificar patrones, temas emergentes y cambios en el discurso, mientras que el análisis de redes sociales complejas ofrece una vista sobre las estructuras de interacción y la influencia de diversos actores en la diseminación de información sobre la IA.

Este enfoque metodológico multidisciplinario facilita una caracterización cuantitativa profunda de la evolución discursiva de la IA en plataformas de *streaming* (Bernstein et al., 2019). A través de la recopilación y análisis de un corpus comprensivo de contenido textual, el estudio identifica tendencias significativas en la presentación y percepción de la IA, la evolución de temas específicos, y cómo diferentes grupos perciben estos cambios (Gonzales et al., 2020). Este análisis no solo resalta la dinámica del discurso sobre la IA sino también cómo factores externos como avances tecnológicos y eventos globales pueden influir en esta evolución discursiva, al examinar la intersección entre tecnología, comunicación y sociedad, este trabajo no solo enriquece el discurso académico sobre la IA sino que también provee herramientas críticas para navegar su futuro desarrollo e integración en el tejido social.

2. Metodología

La metodología adoptada en este estudio descriptivo para analizar la evolución discursiva en Inteligencia Artificial (IA) a través de plataformas de *streaming*, específicamente en YouTube, combina técnicas avanzadas de minería de texto y análisis de redes sociales complejas (Lee, 2023). Este enfoque metodológico interdisciplinario permite una exploración detallada y cuantitativa de las dinámicas discursivas y la influencia de distintos actores en el ámbito de la IA (Lau et al., 2014).

2.1. Participantes

Para la fase de recolección de datos, en la cual se utilizaron 5301 videos publicados en YouTube los cuales se seleccionaron 16 canales de YouTube relevantes en el ámbito de la Inteligencia Artificial (Tabla 1). Se extrajo un corpus comprensivo de videos y sus descripciones, comentarios y metadatos relacionados, como la fecha de publicación y el número de visualizaciones. Esta selección se realizó con el objetivo de abarcar una amplia gama de perspectivas dentro de la comunidad de IA, incluyendo tanto a académicos como a entusiastas y profesionales de la industria.

Tabla1. Descripción de los participantes del studio

N	Canal de Youtube	Número de Suscriptores	Número de Videos
1	IBM Technology	685000	852
2	Lex Fridman	3770000	783
3	Ringa Tech	259000	48

N	Canal de Youtube	Número de Suscriptores	Número de Videos
4	Data School	236000	140
5	Arxiv Insights	91800	13
6	Tech Dummies Narendra L	155000	60
7	Inteligencia Artificial	77700	110
8	Two Minute Papers	1530000	868
9	Machine Learning 101	2540	4
10	StatQuest with Josh Starmer	1120000	268
11	Tech With Tim	1460000	974
12	Code Bullet	3110000	59
13	Siraj Raval	765000	466
14	3Blue1Brown	5970000	172
15	Dot CSV	843000	208
16	AprendeIA con Ligdi Gonzalez	65800	276
Total		20140840	5301

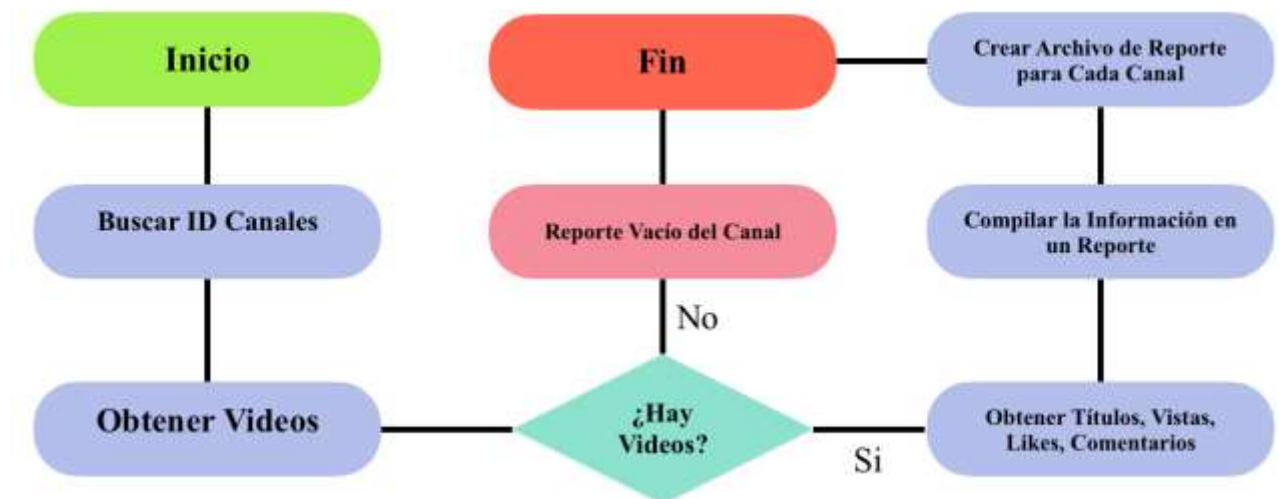
Fuente: Elaboración propia, 2024.

2.2. Técnicas e instrumentos

Obtención de Información

Se realizó la búsqueda y selección de canales de YouTube específicamente relacionados con la Inteligencia Artificial. Posteriormente, se realiza la obtención de los identificadores de canal, seguido de la extracción de videos de cada uno (Li et al., 2024). La decisión de si hay contenido disponible en el canal conduce a dos rutas: la recolección de identificadores de videos, títulos, visualizaciones, «me gusta» y comentarios, o la creación de un reporte vacío en caso de ausencia de videos (Srikumar & Srikumar, 2021). La información recolectada se compila entonces en un reporte, y se crea un archivo para cada canal, concluyendo así la fase de recopilación de datos del estudio como se muestra en la figura 1.

Figura 1. Proceso de Recopilación de Información



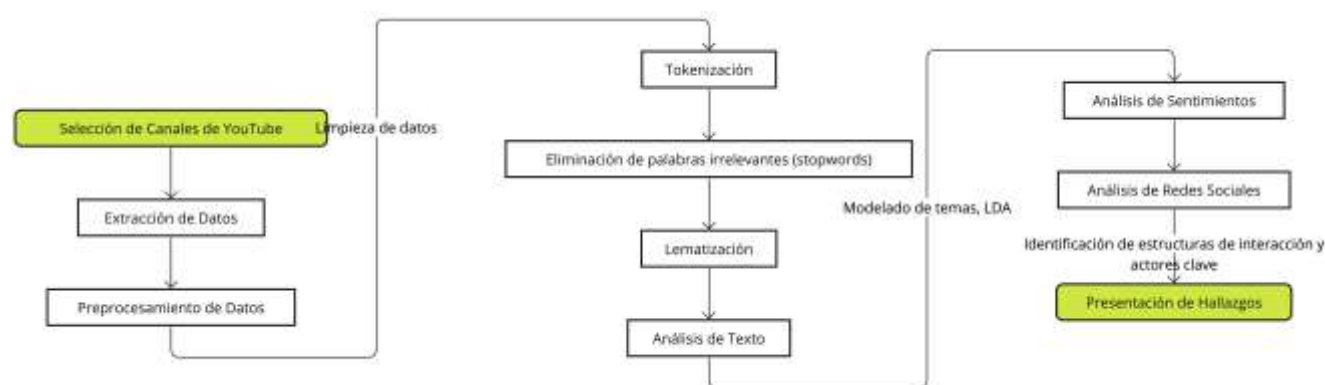
Fuente: Elaboración Propia

Procesamiento y Análisis de Texto

Para el procesamiento y análisis del texto extraído, se utilizó *Python* como lenguaje de programación principal, aprovechando su robusto ecosistema de librerías para el análisis de datos y procesamiento del lenguaje natural (PLN) (Torres-Cruz & Mamani, 2022). Las siguientes librerías fueron esenciales en nuestro análisis:

- **Pandas:** Utilizada para la manipulación y análisis de datos. Facilitó la organización del corpus de texto y metadatos asociados en estructuras de datos eficientes para su posterior análisis.
- **Gensim:** Especialmente su módulo de corpora y *LdaModel*, permitió la construcción de modelos de distribución latente de Dirichlet (LDA) para identificar temas dominantes y su evolución en el tiempo dentro del corpus.
- **NLTK (Natural Language Toolkit):** Esta librería se empleó para tareas de preprocesamiento del texto, incluyendo la *tokenización*, eliminación de *stopwords*, y lematización mediante el *WordNetLemmatizer*, preparando así el texto para un análisis más efectivo.
- **Matplotlib:** Facilitó la visualización de los resultados del análisis, permitiendo la generación de gráficos y diagramas para una mejor interpretación de los datos.

Figura 2. Diagrama de flujo del proceso de Información



Fuente: Elaboración Propia

Análisis de Redes Sociales

Conjuntamente con el análisis de texto, se realizó un análisis de las estructuras de interacción y la influencia de diferentes actores dentro de la comunidad de IA en YouTube (Campis, 2023). Este análisis implicó la identificación de redes de co-ocurrencia de temas y la influencia de canales específicos en la diseminación de información y formación de discursos.

Presentación de Resultados

Para la presentación de los resultados obtenidos, se utilizó la librería *reportlab* de Python para generar informes en formato PDF (Gonzales et al., 2023). Este proceso incluyó la creación de tablas y gráficos estilizados, así como la redacción de párrafos interpretativos de los resultados, utilizando *SimpleDocTemplate*, *Table*, *TableStyle*, y estilos personalizados de *Paragraph* para estructurar el documento final de manera profesional.

2.3. Contexto de estudio

El contexto de este estudio se sitúa en un momento histórico en el que la Inteligencia Artificial (IA) no solo ha alcanzado avances tecnológicos significativos, sino que también ha permeado el discurso público y académico de manera profunda. A medida que la IA se integra cada vez más en diversos aspectos de la vida cotidiana y profesional, las plataformas de *streaming* emergen como espacios cruciales para la diseminación y discusión de conocimientos, opiniones y predicciones relacionadas con esta tecnología.

(Campbell et al., 2000). Estas plataformas, representadas principalmente por YouTube, no solo actúan como canales de comunicación entre expertos y entusiastas, sino que también juegan un papel fundamental en la configuración de la percepción pública sobre la IA. Este fenómeno ha generado una rica amalgama de contenidos que reflejan, informan y a su vez moldean la evolución discursiva sobre la IA. En este contexto, el estudio se propone cuantificar y analizar esta evolución discursiva, empleando un enfoque interdisciplinario que integra la minería de texto y el análisis de redes sociales complejas, con el objetivo de ofrecer una perspectiva amplia y detallada sobre cómo las discusiones sobre IA han cambiado a lo largo del tiempo y cuál ha sido el impacto de estos cambios en la percepción pública (Torres & Moreta, 2023).

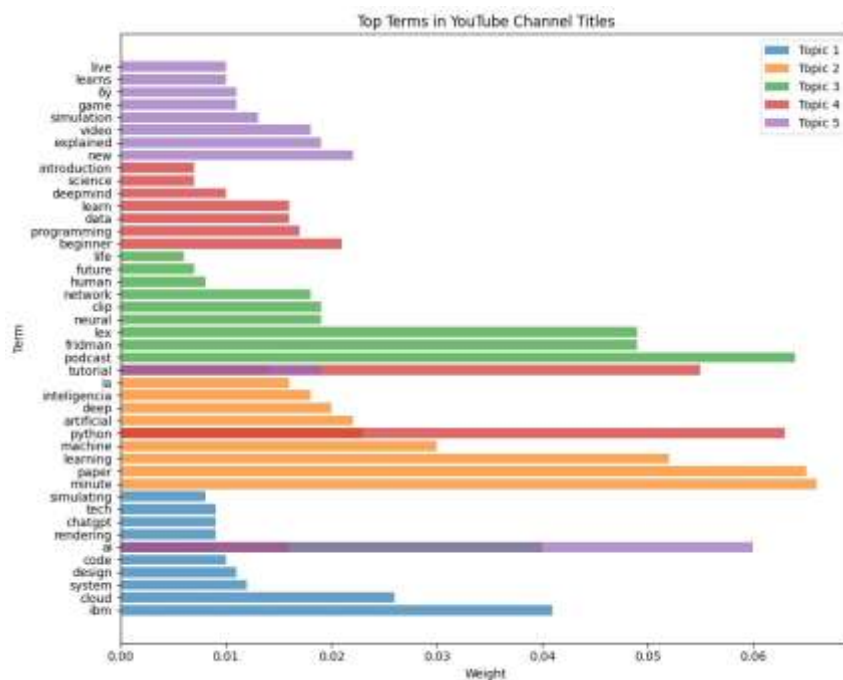
3. Objetivos

El objetivo general de este estudio es caracterizar cuantitativamente la evolución discursiva en el campo de la Inteligencia Artificial (IA) a través de plataformas de *streaming*, con un enfoque particular en YouTube. Esto implica el uso integrado de metodologías de minería de texto y análisis de redes sociales complejas para examinar cómo se presentan y perciben los temas relacionados con la IA en el discurso digital. Mediante el análisis de un corpus comprensivo de contenido generado en estas plataformas, el estudio busca identificar patrones, temas emergentes y cambios en el discurso sobre la IA. Además, se pretende explorar las estructuras de interacción y la influencia de diferentes actores en la diseminación de información sobre la IA. Este enfoque proporcionará *insights* valiosos sobre la evolución de la comunicación y la percepción pública de la IA, ofreciendo una base empírica para futuras investigaciones, estrategias de comunicación, y políticas en el ámbito de la inteligencia artificial

4. Resultados y discusión

La figura 2 muestra un análisis de términos clave dentro de los títulos de canales de YouTube, relacionados con el campo de la Inteligencia Artificial y distribuidos en cinco temas distintos, identificados como Tema 1 a Tema 5. Los términos están organizados en orden descendente de peso dentro de cada tema, el análisis de texto y la modelización de temas o *topic modeling* fueron los mecanismos para obtener los resultados.

Figura 2. Análisis Discursivo de los Canales de YouTube



Fuente: Elaboración propia, 2024.

La figura 2 muestra un análisis de términos clave dentro de los títulos de canales de YouTube, relacionados con inteligencia artificial IA, por lo que el modelado de tópicos nos permite hacer una distinción sobre los tópicos más abordados en los videos de IA, en las plataformas de *streaming*:

Tema 1 (Azul): Tecnología y Plataformas de IA

Los términos más relevantes son «tech», «chatgpt», «rendering», «ai», «code», «design», «system», «cloud» e «ibm», lo que sugiere que este tema podría estar relacionado con aspectos técnicos de la Inteligencia Artificial, incluyendo plataformas de computación en la nube, programación y empresas como IBM.

Tema 2 (Naranja): Educación y Tutoriales en Aprendizaje Automático

Los términos con más peso incluyen «simulating», «tech», «minute», «paper», «machine learning», «python», «artificial» y «deep inteligencia», que podrían indicar un enfoque en tutoriales, presentaciones breves y publicaciones académicas, posiblemente relacionadas con el aprendizaje profundo y la programación en Python.

Tema 3 (Verde): Programación Neural y sus Implicaciones Sociales

Aquí destacan «programming», «beginner», «life», «future», «human network», «neural», «clip» y «max», términos que pueden señalar una discusión sobre la programación neural y sus implicaciones futuras en la vida humana y la sociedad, con un enfoque en contenido accesible para principiantes.

Tema 4 (Rojo): Fundamentos Educativos en Ciencia de Datos e IA

Incluye «learn», «data», «programming», «beginner», «science», «deepmind», «introduction» y «new», sugiriendo un énfasis en recursos educativos para aprender sobre ciencia de datos y AI, con una mención a DeepMind, conocido por sus avances en la IA.

Tema 5 (Morado): Interactividad y Aprendizaje Lúdico en IA

Con términos como «live», «learns», «game», «simulation», «video» y «explained», este tema podría referirse al contenido interactivo y educativo que explora conceptos de IA a través de simulaciones y juegos, posiblemente dirigidos a un aprendizaje en vivo o transmitido.

5. Conclusiones

Las conclusiones de nuestro estudio proporcionan una visión integral de la presencia y evolución de la IA en las plataformas de *streaming*:

Evolución de la Discusión sobre IA: Se ha identificado una evolución significativa en los temas discutidos en plataformas como YouTube. La IA ha pasado de ser un tema especializado a uno de interés general, ampliando su alcance y la diversidad de su audiencia. Los patrones de discusión reflejan un aumento en la conciencia sobre los aspectos éticos, el impacto en la privacidad y las implicaciones socioeconómicas de la IA.

Percepción Pública y Comunicación de la IA: Los resultados indican que la forma en que los canales de YouTube presentan la IA influye en la percepción pública. Los creadores de contenido que enfatizan aplicaciones prácticas de la IA y sus beneficios tienden a atraer una audiencia más amplia y fomentan una percepción positiva de la tecnología.

Influencia de Actores en la Diseminación de la IA: El análisis de redes sociales ha revelado que ciertos canales de YouTube actúan como influenciadores clave en la comunidad de IA, con la capacidad de dar forma significativamente al discurso. La interacción entre canales pequeños y grandes puede crear un ecosistema de contenido rico y variado.

Impacto de Eventos y Avances Tecnológicos: Se ha observado que eventos globales y avances tecnológicos actúan como catalizadores de cambio en el discurso de la IA. Estos eventos no solo generan picos de actividad discursiva sino que también introducen nuevos temas y perspectivas en la conversación.

Metodología Innovadora para Análisis de Discursos: La integración exitosa de la minería de texto y el análisis de redes sociales complejas ha demostrado ser una metodología robusta para el estudio de

grandes conjuntos de datos de plataformas de *streaming*, proporcionando *insights* detallados que serían difíciles de obtener a través de métodos tradicionales.

Direcciones para Futuras Investigaciones: Este estudio abre múltiples caminos para futuras investigaciones, incluyendo el análisis longitudinal del impacto de la IA en diversas industrias, la influencia de la IA en la toma de decisiones políticas y la gestión de la comunicación estratégica en plataformas de *streaming*.

Este estudio ha proporcionado una visión cuantitativa y cualitativa de la evolución discursiva de la IA en plataformas de *streaming*, resaltando la importancia de una comunicación efectiva y la comprensión de las dinámicas de influencia dentro de la comunidad en línea. Los hallazgos enfatizan la necesidad de una consideración cuidadosa de cómo se presenta la IA al público, y el papel que juegan los creadores de contenido en la formación de la comprensión y la actitud del público hacia esta tecnología transformadora.

Referencias

- Bernstein, R., Osadchy, M., Keren, D., & Schuster, A. (2019). LDA classifier monitoring in distributed streaming systems. *Journal of Parallel and Distributed Computing*, 123, 156-167. <https://doi.org/10.1016/j.jpdc.2018.09.017>
- Campbell, M., Cosgrove, J. A., Greated, C. A., Jack, S., & Rockliff, D. (2000). Review of LDA and PIV applied to the measurement of sound and acoustic streaming. *Optics & Laser Technology*, 32(7), 629-639. [https://doi.org/10.1016/S0030-3992\(00\)00091-8](https://doi.org/10.1016/S0030-3992(00)00091-8)
- Campis, E. S. (2023). Aplicación del Análisis de Redes Sociales para el estudio de las redes de comunicación en línea: Evidencia empírica de Twitter. *EMPIRIA: Revista de Metodología de Ciencias Sociales*. <https://doi.org/10.5944/empiria.57.2023.36434>
- Gonzales, E. P. Z., Gallegos, W. L. A., Gallegos, W. L. A., Toia, A. M. del C., Rivera, R., & Condori, M. L. (2020). Depresión postparto en mujeres víctimas de violencia doméstica en la ciudad de Arequipa. *null*. <https://doi.org/null>
- Gonzales, J. E., Ortega, Á. C., Blanco, R. R., & Diez, E. B. (2023). ANÁLISIS DE SENTIMIENTOS Y EMOCIONES SOBRE GASTRONOMÍA PERUANA USANDO MINERÍA DE TEXTO CON PYTHON. *Hatun Yachay Wasi*. <https://doi.org/10.57107/hyw.v3i1.63>
- Lau, J. H., Lau, J. H., Newman, D. J., Newman, D., Baldwin, T., & Baldwin, T. (2014). Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. *null*. <https://doi.org/10.3115/v1/e14-1056>
- Lee, S. (2023). An Analysis on Major Issues of Medical Information on YouTube Platform Using LDA Topic Modeling. *The Journal of the Korea Contents Association*. <https://doi.org/10.5392/jkca.2023.23.08.198>
- Li, D., Bi, Z., Han, X., Qu, Z., & Lyu, C. (2024). Extracting key topics from massive COVID-19 information on social networks: An integrated deep learning and LDA framework. *High-Confidence Computing*, 100213. <https://doi.org/10.1016/j.hcc.2024.100213>
- Li, M., Lu, M., Akram, U., & Cheng, S. (2024). Understanding how customer social capital accumulation in brand communities: A gamification affordance perspective. *Journal of Retailing and Consumer Services*, 78, 103761. <https://doi.org/10.1016/j.jretconser.2024.103761>
- Morchid, M., Morchid, M., Linarès, G., & Linarès, G. (2013). A LDA-based method for automatic tagging of Youtube videos. *Workshop on Image Analysis for Multimedia Interactive Services*. <https://doi.org/10.1109/wiamis.2013.6616126>
- Srikumar, N., & Srikumar, N. (2021). Feature Augmentation for Improved Topic Modeling of Youtube Lecture Videos using Latent Dirichlet Allocation. *null*. <https://doi.org/10.21427/4ey6-qg08>
- Torres, V., & Moreta, O. E. (2023). Detección de ansiedad mediante minería de texto en la era de redes sociales: Revisión bibliográfica. *Ciencia Huasteca Boletín Científico de la Escuela Superior de Huejutla*. <https://doi.org/10.29057/esh.v11i22.10879>
- Torres-Cruz, F., & Mamani, Y. J. Y. (2022). Técnicas de inteligencia artificial en la valoración de la enseñanza virtual por estudiantes de nivel universitario. *HUMAN Review International Humanities Review / Revista Internacional de Humanidades*. <https://doi.org/10.37467/revhuman.v11.3853>