



Los Datos de X en la Investigación Científica: Tendencias y Desafíos

X Data-Based Research: A Review of Trends and Challenges

Lucía Rivadeneira

Universidad Técnica de Manabí. Ecuador.

lucia.rivadeneira@utm.edu.ec



Ignacio Loor

Universidad Técnica de Manabí. Ecuador.

ignacio.loor@utm.edu.ec



Cómo citar este artículo / Referencia normalizada:

Rivadeneira, Lucía y Loor, Ignacio (2025). Los Datos de X en la Investigación Científica: Tendencias y Desafíos [X Data-Based Research: A Review of Trends and Challenges]. *Revista Latina de Comunicación Social*, 83, 1-14. <https://www.doi.org/10.4185/RLCS-2025-2418>

Fecha de Recepción: 03/10/2024

Fecha de Aceptación: 18/12/2024

Fecha de Publicación: 20/03/2025

RESUMEN

Introducción: El crecimiento de las redes sociales, especialmente X (antes Twitter), ha impulsado la investigación científica, destacándose como fuente valiosa de datos. Esta revisión analiza los factores clave que han favorecido su uso, las tendencias futuras y los desafíos para los investigadores. **Metodología:** La revisión, basada en una búsqueda sistemática en Scopus, adoptó un mapeo temático para identificar aplicaciones interdisciplinarias, innovaciones metodológicas y el impacto de eventos globales, con énfasis en el procesamiento del lenguaje natural (PLN) para el análisis de datos. **Resultados:** El PLN creció un 268% entre 2019 y 2023, consolidándose como herramienta clave. Sin embargo, entre 2021 y 2023 se observó una desaceleración en publicaciones basadas en X, mientras que Instagram y TikTok crecieron. X sigue siendo la plataforma más usada, aunque las restricciones de datos y el auge de nuevas plataformas podrían haber influido. **Conclusiones:** La investigación señala la necesidad de desarrollar técnicas de análisis más sofisticadas, integrar estándares éticos sobre privacidad y consentimiento, y fomentar enfoques interdisciplinarios en el uso de datos de redes sociales.

Palabras clave: redes sociales; Twitter; X; investigación; tendencias.

ABSTRACT

Introduction: The growth of social media, especially X (formerly Twitter), has become a key resource for scientific research. This literature review identifies the factors driving its use, forecasts trends, and addresses

challenges faced by researchers. **Methodology:** The review, based on a systematic search in Scopus, employed thematic mapping to identify interdisciplinary applications, methodological innovations, and the impact of global events. Key among these innovations was natural language processing (NLP) for data analysis, which grew 268% from 2019 to 2023. **Results:** NLP has established itself as a vital tool. However, publications based on X data showed a slowdown between 2021 and 2023, while Instagram and TikTok-based publications accelerated, signaling increased interest in these platforms. X remains the most used platform, followed by Facebook. **Conclusions:** The review highlights the need for more advanced analysis methods, stronger ethical standards concerning privacy and consent, and interdisciplinary approaches in social media research.

Keywords: social media; Twitter; X; research; trends.

1. INTRODUCCIÓN

La incorporación de datos de X, anteriormente conocida como Twitter, en la investigación científica constituye una intersección fundamental entre el análisis de redes sociales y diversas disciplinas académicas. Este artículo, basado en una revisión de literatura, examina el uso de datos generados por Twitter entre 2019 y 2023, periodo en el que se produjo su *rebranding* como X. El estudio identifica los factores que han impulsado el uso de los datos de esta plataforma en investigaciones científicas, anticipa tendencias futuras y aborda las principales barreras que enfrentan los investigadores, como las restricciones en el acceso a datos, los problemas de representatividad y el surgimiento de plataformas emergentes.

Por lo menos desde Asur y Huberman (2010), Twitter se ha venido consolidando como una plataforma esencial para recopilar datos para el análisis de los fenómenos sociales, discursivos y comunicativos, gracias a su capacidad para proporcionar datos en tiempo real sobre la opinión pública, movilización social y difusión de información. Sin embargo, el auge de plataformas como TikTok e Instagram, junto con los cambios en la estructura demográfica de los usuarios de X, plantea nuevos retos para su uso como fuente principal de datos en investigaciones interdisciplinarias.

A lo largo de la trayectoria de Twitter en los manuscritos indexados en Scopus, se han registrado 377 artículos de revisión de literatura sobre el análisis de contenido de esta plataforma. Con un volumen tan considerable de revisiones, puede resultar difícil identificar aspectos que aún no han sido explorados en profundidad. Esta revisión se distingue por abordar las dificultades metodológicas y éticas señaladas en estudios previos y por adoptar una perspectiva prospectiva que destaca tanto las oportunidades como los desafíos para la investigación en redes sociales en un entorno digital en constante transformación. Su importancia radica en ofrecer un panorama actualizado del estado del arte, destacando tendencias actuales, retos emergentes y recomendaciones prácticas para fortalecer la investigación en redes sociales.

2. OBJETIVOS

Tres objetivos guiaron el alcance de esta investigación. Primero, buscamos desentrañar los factores que han contribuido al aumento de la investigación basada en datos de Twitter, centrándonos en el análisis del uso de datos de redes sociales en el ámbito académico y su impacto en la generación de conocimiento. Segundo, pretendemos anticipar las tendencias futuras en este campo, identificando temas emergentes y áreas potenciales de crecimiento en la investigación con datos de X. Finalmente, aspiramos a delinear las principales dificultades que los investigadores han reportado al trabajar con datos de Twitter, abarcando preocupaciones relativas al acceso, el análisis y los aspectos éticos.

3. METODOLOGÍA

El estudio consistió en una revisión sistemática de la literatura centrada en el uso de datos de Twitter en la investigación científica. El proceso comenzó con una búsqueda en la base de datos Scopus, seleccionada por su

amplia cobertura de publicaciones académicas internacionales y multidisciplinarias. La estrategia de búsqueda incluyó los términos Twitter AND data OR content AND analysis, limitando los resultados a los años 2019-2023 y a tipos de documentos como artículos, actas de congresos y capítulos de libros. Esta búsqueda inicial arrojó un total de 11.058 publicaciones.

Para refinar los resultados, se aplicaron filtros adicionales que incluyeron exclusivamente publicaciones en inglés y español, además de restringir la búsqueda a áreas temáticas relacionadas con las ciencias sociales. Estos criterios redujeron el número de publicaciones a 3.227. Posteriormente, se exportaron los títulos y resúmenes de estas publicaciones a un archivo de Excel para un análisis más detallado. En esta etapa, se implementó un proceso sistemático basado en palabras clave para alinear los resultados con los objetivos del estudio. Las palabras clave para este proceso de filtro se asociaron con cada objetivo: para identificar factores que contribuyen al aumento de la investigación basada en Twitter, se utilizaron términos como *growth*, *adoption*, *popularity*, *increase*, *impact* y *use of Twitter*; para anticipar tendencias futuras, se incluyeron términos como *future*, *emerging topics*, *opportunities*, *evolution* y *advancements*; mientras que para examinar las principales dificultades, se emplearon palabras como *challenges*, *limitations*, *barriers*, *access*, *privacy*, *ethics* y *bias*.

Con el apoyo de fórmulas en Excel, cada resumen fue evaluado automáticamente para identificar la presencia de estos términos clave, y aquellos que abordaban uno o más objetivos fueron marcados para su inclusión. A continuación, los resúmenes fueron clasificados en función de un puntaje asignado según el número de objetivos que cumplían, priorizando aquellos relevantes a por lo menos dos de los objetivos de este estudio. Este paso permitió refinar el conjunto a 80 resúmenes. Finalmente, se realizó una revisión manual de estos resúmenes para verificar su relevancia y alineación con los objetivos del estudio, lo que resultó en la selección final de 43 manuscritos. A continuación, la Tabla 1 presenta un detalle de los pasos de búsqueda y selección de artículos.

Tabla 1. Pasos para la búsqueda y selección de publicaciones relevantes

	Detalles	Código / fórmula
Búsqueda en Scopus	Estrategia inicial de búsqueda en Scopus, centrada en Twitter y el análisis de datos entre 2019 y 2023.	Twitter AND data OR content AND analysis
Filtro por idioma	Publicaciones limitadas a inglés y español.	Aplicado directamente en Scopus
Filtro por área temática	Restringido a áreas relacionadas con ciencias sociales.	Aplicado directamente en Scopus
Filtro por palabras clave (Excel)	Identificación de resúmenes relevantes según los objetivos: Factores, Tendencias y Dificultades.	=IF(OR(ISNUMBER(SEARCH("growth"; C2)); ISNUMBER(SEARCH("adoption"; C2)); ...), "Yes"; "No") (para cada objetivo)
Puntuación	Priorización de resúmenes que aborden múltiples objetivos.	=COUNTIF(D2:F2; "Yes")
Revisión manual	Revisión de los 80 resúmenes más relevantes para garantizar su alineación con los objetivos de investigación.	Sin código—evaluación subjetiva de los resúmenes marcados durante el filtrado.

Fuente: Elaboración propia (2024).

Los 43 manuscritos seleccionados fueron examinados mediante análisis temático siguiendo la metodología de Soaita *et al.* (2020), que implica la codificación sistemática de los datos y la identificación de patrones y temas recurrentes. Para garantizar la validez del análisis temático, el proceso fue colaborativo entre los autores del estudio, quienes discutieron y revisaron las codificaciones iniciales en varias rondas iterativas. Esto permitió identificar y refinar las categorías emergentes y agruparlas en temas más amplios de manera consensuada, reduciendo posibles sesgos individuales. Adicionalmente, las decisiones tomadas durante el análisis fueron evaluadas con base en su coherencia con los objetivos del estudio, de modo que los temas finales estuvieran directamente fundamentados en los datos de los manuscritos.

4. RESULTADOS

La investigación basada en datos de redes sociales ha experimentado un crecimiento importante desde el 2010, especialmente en el caso de Twitter, que se ha consolidado como una plataforma clave para el análisis de fenómenos sociales, políticos y económicos. Este auge responde tanto a su capacidad para captar interacciones sociales en tiempo real como a los avances tecnológicos que han permitido manejar grandes volúmenes de datos no estructurados con mayor precisión. No obstante, la creciente dependencia de estas plataformas también introduce desafíos críticos, como la necesidad de garantizar la representatividad de las muestras, superar restricciones de acceso a los datos y abordar cuestiones éticas emergentes. Los resultados presentados a continuación examinan los factores detrás del crecimiento de la investigación con datos de Twitter, las tendencias y oportunidades que definen el futuro del campo, y los desafíos que los investigadores enfrentan en su acceso y análisis, configurando un panorama en constante evolución.

4.1. Factores detrás del crecimiento de la investigación con datos de Twitter

Durante los últimos cinco años, la investigación basada en datos de Twitter ha crecido aceleradamente, impulsada por una combinación de factores tecnológicos, sociales y académicos. El número de publicaciones en Scopus que analizan contenido de Twitter aumentó de 1.827 en 2019 a 2.723 en 2023, reflejando un incremento del 49% que evidencia la creciente relevancia de esta plataforma en la investigación académica. Este fenómeno ha consolidado a Twitter como una herramienta clave para abordar preguntas de investigación contemporáneas, ofreciendo una profundidad metodológica sin precedentes en disciplinas diversas. Entre los factores más destacados se encuentran la expansión del alcance de la plataforma, definido por su número de usuarios, los avances tecnológicos, las preferencias emergentes en el ámbito académico y la integración del análisis de *big data* en las ciencias sociales.

El alcance global y su capacidad para reflejar en tiempo real dinámicas sociales y globales han sido esenciales para su amplia adopción en la investigación científica. Nellore *et al.* (2023) destacan que la accesibilidad de esta plataforma ha permitido explorar fenómenos sociales complejos y diversos, convirtiéndola en un recurso indispensable en disciplinas que abarcan desde la salud pública hasta los estudios de comunicación. Ejemplos de estas aplicaciones incluyen el análisis de respuestas sociales ante eventos críticos, como pandemias, disturbios políticos y desastres naturales (Cvetojevic y Hochmair, 2018; Raja *et al.*, 2016). Estos estudios han demostrado cómo los datos de Twitter captan el pulso de las interacciones sociales en momentos clave, ofreciendo una ventana única para estudiar cómo las sociedades responden a eventos de gran impacto. Además, Henry *et al.* (2018) destacan que esta plataforma también actúa como un canal de difusión transnacional de información, reforzando su relevancia global en la investigación.

No obstante, la capacidad de Twitter para reflejar dinámicas sociales en tiempo real plantea importantes retos. La naturaleza efímera y fragmentada de las interacciones puede llevar a un sesgo en los análisis, priorizando narrativas emergentes mientras se desatienden aspectos estructurales más profundos. Por ejemplo, aunque Cvetojevic y Hochmair (2018) y Raja *et al.* (2016) han evidenciado el potencial de la plataforma para analizar respuestas sociales, el riesgo de sobre-representar las perspectivas de usuarios activos o digitalmente conectados podría excluir a comunidades menos visibles.

Twitter también se ha convertido en un espacio donde las voces pueden amplificarse y las opiniones públicas pueden movilizarse, desafiando las formas tradicionales de mediación de información. Mao *et al.* (2023), en el contexto de la pandemia de COVID-19, exploran cómo la interacción de Twitter con los medios tradicionales contribuye a moldear flujos de noticias, destacando el papel de la plataforma en la difusión de información crítica en tiempo real. Yuan (2017), en el marco de las elecciones presidenciales en Estados Unidos, resalta su papel en el análisis de patrones de comunicación, demostrando cómo Twitter puede ser utilizado para identificar dinámicas en la propagación de mensajes políticos y predecir tendencias futuras. En ese mismo orden de ideas, Dersan Orhan (2020), al analizar las reacciones a los tuits de Donald Trump mientras era presidente de los Estados Unidos en relación con Irán, demuestra cómo la plataforma puede actuar como un barómetro en tiempo real para informar decisiones estratégicas, proporcionando una visión detallada de las percepciones públicas sobre temas de política exterior. Al respecto, Efanova (2023), en un estudio sobre la diplomacia digital de Estados Unidos, destaca el uso de Twitter como una herramienta diplomática, evidenciando su papel en la construcción de narrativas y la negociación simbólica en contextos internacionales complejos.

Sin embargo, estas aparentes prelações no están exentas de limitaciones. El aumento en el uso de *bots* (Ghosh *et al.*, 2023) y cuentas automatizadas plantea retos adicionales, introduciendo dudas sobre la autenticidad de las interacciones y, por consiguiente, sobre la validez de los análisis basados en datos de Twitter. Esto destaca la necesidad de metodologías más robustas que permitan distinguir entre el comportamiento humano y la actividad automatizada, asegurando la fiabilidad de los resultados.

Por otro lado, los avances tecnológicos, desde dispositivos con mayor capacidad de almacenamiento y procesamiento hasta algoritmos más sofisticados, han fortalecido la capacidad de los investigadores para analizar los datos generados en Twitter. Herramientas avanzadas como el aprendizaje automático (ML) y el procesamiento del lenguaje natural (PLN) han sido fundamentales para manejar grandes volúmenes de datos no estructurados (Yu y Muñoz-Justicia, 2022). Los datos no estructurados se refieren a información que no sigue un modelo predefinido, como ocurre con grandes volúmenes de texto. El PLN, en particular, se ha consolidado como un método clave para interpretar el contenido textual de la plataforma, ofreciendo una comprensión más matizada del discurso digital y su contexto. Estas innovaciones han ampliado el alcance del análisis académico, permitiendo explorar dinámicas sociales y patrones de comunicación con mayor profundidad.

Simultáneamente, los científicos de datos, en su esfuerzo por avanzar en los métodos de análisis, han recurrido intensamente a los datos de Twitter para probar y validar modelos y algoritmos, lo que ha impulsado considerablemente el crecimiento de investigaciones basadas en esta plataforma. Por ejemplo, Budiharto y Meiliana (2018) emplearon datos de Twitter relacionados con elecciones presidenciales en Indonesia para evaluar la capacidad predictiva y explicativa de sus modelos. De manera similar, Rangel *et al.* (2020) utilizaron estos datos para desarrollar modelos enfocados en el estudio de la desinformación y las noticias falsas. Estas aplicaciones reflejan no solo el interés creciente en fenómenos digitales, sino también un cambio en el paradigma de investigación hacia abordajes más interdisciplinarios y adaptados al entorno digital.

Por último, la integración del análisis de *big data* ha redefinido las posibilidades de la investigación en ciencias sociales. Schroeder y Cowls (2019) y Steinert-Threlkeld (2018) destacan cómo el uso de grandes conjuntos de datos de Twitter ha permitido una comprensión más detallada del comportamiento humano y ha fomentado el desarrollo de nuevos modelos teóricos. Aplicaciones como las de Pagolu *et al.* (2017) y Zou y Herremans (2023) en el análisis financiero ilustran la capacidad de estos datos para integrar teorías económicas con métodos computacionales avanzados. De manera similar, Wang *et al.* (2019) destacan cómo los datos de Twitter han vinculado respuestas sociales y medioambientales en la gestión de desastres, reforzando el valor de enfoques interdisciplinarios.

En síntesis, los factores detrás del crecimiento de la investigación basada en datos de Twitter reflejan un entorno académico dinámico donde la plataforma no solo facilita la exploración de fenómenos sociales complejos, sino que también redefine las prácticas investigativas. Aunque los avances tecnológicos y el interés académico han consolidado su relevancia, persisten desafíos relacionados con la representatividad y la autenticidad de los datos. Esto invita a la comunidad académica a continuar innovando tanto en el desarrollo de métodos como en el enfoque crítico hacia el uso de estos datos, asegurando así que las oportunidades ofrecidas por Twitter se utilicen de manera responsable y eficaz.

4.2. Tendencias y oportunidades en la investigación con datos de X

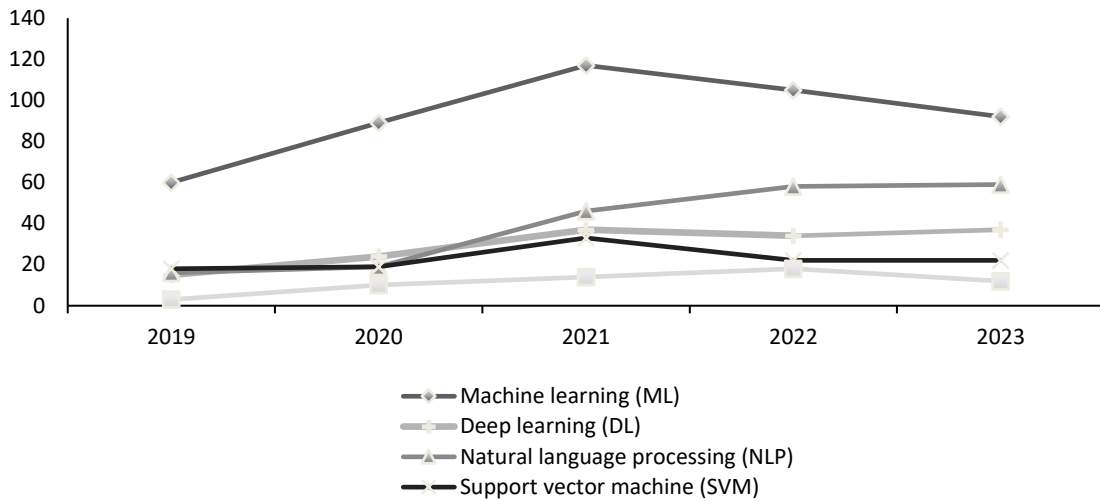
La investigación con datos de Twitter ha experimentado una expansión multidimensional en los últimos años, abarcando nuevas áreas de aplicación y avances metodológicos. Por un lado, estos datos se han integrado progresivamente en un número creciente de disciplinas, encontrando aplicaciones en campos como la salud pública, la gestión de desastres, los estudios medioambientales y la analítica predictiva. Por otro lado, los métodos de extracción y análisis de datos han alcanzado un nivel de sofisticación creciente, permitiendo predicciones más precisas y representaciones más transparentes. Técnicas de aprendizaje automatizado se han consolidado como herramientas esenciales para analizar grandes volúmenes de datos, tanto estructurados como no estructurados, con una precisión notable. Paralelamente, su alcance conceptual ha evolucionado, dejando atrás enfoques meramente descriptivos para dar paso a la construcción de modelos teóricos más complejos que ofrecen perspectivas innovadoras sobre fenómenos sociales y globales.

Un ejemplo claro de esta multidisciplinariedad es el análisis del discurso público. En este ámbito, Fütterer *et al.* (2023) y Hu *et al.* (2021) han demostrado cómo los métodos computacionales facilitan la exploración de actitudes y percepciones colectivas en temas de interés común. Estas aplicaciones no solo reflejan la versatilidad de la plataforma, sino también el creciente reconocimiento de su potencial para abordar problemas complejos desde múltiples disciplinas de forma integrada.

El auge de las aplicaciones interdisciplinarias de los datos de X resalta su potencial para abordar problemas complejos desde múltiples perspectivas. Sin embargo, también plantea el desafío de integrar hallazgos provenientes de metodologías y prioridades diversas. Por ejemplo, mientras que la gestión de desastres utiliza los datos de X para monitorear respuestas inmediatas y coordinar acciones, estudios de salud pública pueden enfocarse en los impactos a largo plazo sobre el bienestar colectivo. Esta diversidad metodológica, lejos de ser un obstáculo, enfatiza la necesidad de desarrollar marcos analíticos que permitan conectar hallazgos aparentemente dispares, construyendo así narrativas más completas que reflejen la complejidad inherente de los fenómenos sociales y globales.

Paralelamente, los desarrollos tecnológicos han redefinido las metodologías para analizar datos de X, priorizando técnicas avanzadas como el ML y PLN. Ioannides *et al.* (2023) destacan que el PLN ha mostrado un crecimiento sostenido en su adopción, debido a su capacidad para extraer patrones del lenguaje dinámico y a menudo informal de X. Este método ha demostrado ser esencial para el análisis de sentimientos, predicciones y comprensión de dinámicas sociales complejas. La Figura 1 muestra cómo el PLN ha superado a otros métodos, como las máquinas de vectores de soporte (SVM) y los árboles de decisión, los cuales han experimentado una disminución notable en los últimos años. Nguyen *et al.* (2024) señalan que el aprendizaje profundo (DL) ha alcanzado su máxima adopción en 2022, aunque su uso ha mostrado una leve disminución en 2023, lo que sugiere una posible maduración de estas metodologías.

Figura 1. Tendencias en el uso de métodos analíticos en datos de Twitter (2019-2023)

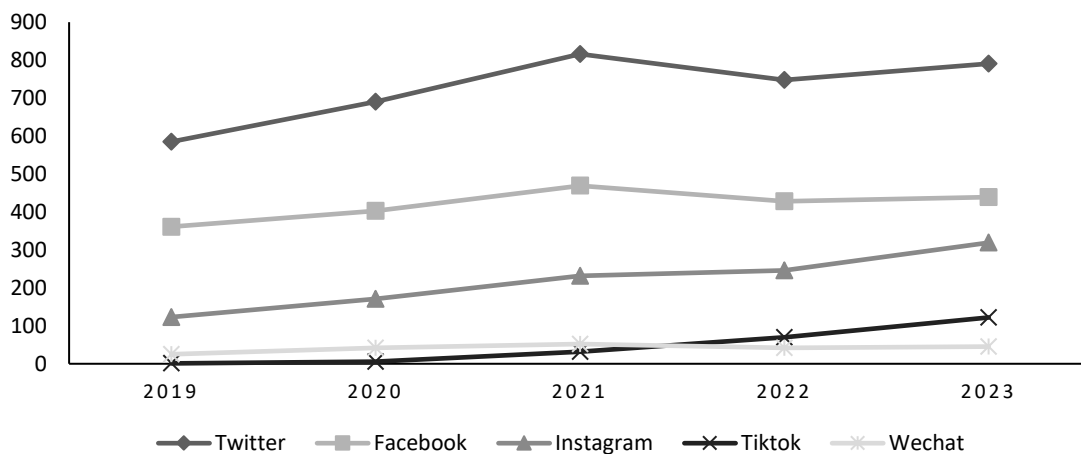


Fuente: Base de datos de Scopus.

Aunque las herramientas avanzadas como el PLN y el DL han permitido un manejo más eficaz de los datos generados en X, su creciente dependencia en la investigación plantea desafíos relacionados con la opacidad de estos métodos. La "caja negra" de los algoritmos empleados en aprendizaje automático y DL puede dificultar la interpretación y replicación de los resultados, comprometiendo la transparencia científica (Mazhar y Dwivedi, 2024). En consecuencia, al momento de seleccionar una herramienta para el análisis del contenido de X, es importante que los investigadores consideren no solo la eficacia de estas herramientas, sino también cómo comunicar sus limitaciones y decisiones metodológicas.

Fuera de los avances metodológicos, las tendencias hacia un análisis multiplataforma están configurando un panorama más diverso para la investigación académica. La Figura 2 compara la frecuencia de publicaciones basadas en datos de X, Instagram, TikTok y WeChat entre 2019 y 2023. Mientras que X alcanzó su pico en 2021, en 2022 se observa una clara desaceleración en su uso en investigaciones, lo que podría señalar una etapa de madurez en su adopción. En contraste, TikTok e Instagram han mostrado un crecimiento sostenido, con TikTok destacándose por su capacidad para captar dinámicas sociales en audiencias más jóvenes a través de formatos visuales y participativos. Sinnenberg *et al.* (2017) sugieren que, aunque X sigue siendo insustituible para ciertos tipos de estudios, la diversificación hacia estas plataformas ofrece nuevas oportunidades para explorar fenómenos sociales desde perspectivas comparativas.

Figura 2. Comparación de publicaciones basadas en datos de redes sociales (2019-2023)



Fuente: Base de datos de Scopus.

Mirando hacia el futuro, la integración de plataformas avanzadas de inteligencia artificial como GPT de OpenAI, IBM Watson y Julius.ai promete revolucionar la investigación con datos de X. Estas herramientas no solo automatizan procesos complejos, sino que también permiten un análisis más refinado de grandes volúmenes de datos textuales. Según Nguyen *et al.* (2024), la combinación de PLN y DL con modelos avanzados de IA podría transformar el análisis de sentimientos y la predicción de tendencias sociales. Sin embargo, esta transición hacia la inteligencia artificial plantea interrogantes éticas y epistemológicas que requieren una reflexión crítica sobre la transparencia y el impacto de estas tecnologías en la investigación social.

Por último, las oportunidades futuras en la investigación con datos de X dependerán de la capacidad de los investigadores para desarrollar metodologías más sofisticadas y éticamente responsables. La naturaleza dinámica y efímera de los datos de X plantea retos únicos en términos de recolección y análisis, especialmente frente a restricciones como las impuestas por la API de la plataforma y los sesgos demográficos en su base de usuarios. Dixon (2023) advierte que la disminución de la popularidad de X entre usuarios jóvenes podría limitar su representatividad en estudios centrados en la cultura juvenil, lo que subraya la necesidad de diversificar las fuentes de datos y adaptar las metodologías a nuevas plataformas.

En suma, las tendencias y oportunidades en la investigación basada en datos de X reflejan un campo en constante evolución, caracterizado por un balance entre innovación metodológica y adaptaciones a un ecosistema digital diversificado. Mientras que X sigue siendo una fuente valiosa para la generación de conocimiento, el crecimiento de plataformas alternativas y los avances en inteligencia artificial prometen expandir los horizontes de la investigación académica hacia territorios más inclusivos e interdisciplinarios. Las Figuras 1 y 2 evidencian cómo estas dinámicas están remodelando el panorama, subrayando la importancia de un enfoque crítico y flexible que permita maximizar el potencial de los datos de redes sociales en un contexto global.

4.3. Desafíos en el acceso y análisis de datos de X

A pesar de las numerosas oportunidades que ofrece el análisis de datos de X para la investigación científica, este campo enfrenta desafíos relacionados con el acceso a los datos, la representatividad de las muestras y las complejidades metodológicas y éticas. Estos desafíos no solo han condicionado la evolución de las investigaciones, sino que también han exigido una reflexión crítica sobre las metodologías empleadas y las implicaciones epistemológicas de trabajar con datos de redes sociales. Estas implicaciones se refieren a cómo las características intrínsecas de las redes sociales —como la naturaleza no estructurada de los datos, los sesgos inherentes en la representación de ciertos grupos sociales y la volatilidad de las interacciones digitales— pueden influir en las preguntas de investigación, en la interpretación de los resultados y en la construcción del conocimiento, planteando preguntas sobre la validez, la generalización y el alcance teórico de los hallazgos obtenidos a partir de estos datos.

Uno de los principales obstáculos es el acceso restringido a los datos. Desde las primeras etapas de la investigación basada en Twitter, las limitaciones impuestas por su API han dificultado la obtención de conjuntos de datos completos y profundos. Weller (2014) destaca que estas restricciones obligan a los investigadores a depender de herramientas de terceros, como *CrowdTangle* o *NodeXL*, que pueden introducir sesgos y afectar la integridad de los resultados. De Vreese y Tromble (2023) señalan que estas herramientas externas, aunque permiten sortear algunas limitaciones, presentan sus propios desafíos, como inconsistencias en la calidad de los datos recolectados, costos adicionales y dificultades en el procesamiento. Estas barreras, además de complicar la recopilación, generan discrepancias entre estudios que utilizan diferentes enfoques y herramientas, reduciendo la comparabilidad de los resultados (Maci *et al.*, 2024). Además, las políticas de acceso cada vez más restrictivas de X no solo limitan la posibilidad de obtener datos en tiempo real y de manera integral, sino que también desincentivan a la comunidad académica a considerarla como una fuente viable para investigaciones futuras. Esta tendencia amenaza con erosionar la relevancia de X como herramienta de análisis

en favor de plataformas más accesibles y transparentes.

El acceso a los datos no es el único problema. La representatividad de las muestras también plantea preocupaciones importantes. La base de usuarios de Twitter, tradicionalmente con mayores ingresos y niveles educativos más altos que la población general, introduce un sesgo demográfico significativo (Blank, 2017; McLachlan, 2024). Este sesgo afecta la validez de los estudios que intentan extrapolar resultados a poblaciones más amplias o representativas. Además, los cambios demográficos recientes en la composición de los usuarios de Twitter agravan este problema, limitando su aplicabilidad en investigaciones que dependen de una representación diversa. Dixon (2023) advierte que la disminución en el uso de Twitter entre audiencias más jóvenes podría afectar su utilidad en estudios centrados en la cultura juvenil, requiriendo una reevaluación de la plataforma como fuente de datos.

La disminución de usuarios jóvenes en Twitter no solo plantea desafíos prácticos, sino también interrogantes teóricas sobre cómo interpretar los datos generados por una plataforma cuyo perfil demográfico está cambiando rápidamente. Si bien los usuarios jóvenes tienden a ser un grupo clave para estudios relacionados con cultura, tendencias y consumo digital, su migración a otras plataformas como TikTok e Instagram sugiere que Twitter puede estar perdiendo relevancia como espacio de interacción social para estos segmentos. Este fenómeno no solo limita la representatividad de los datos, sino que también exige un replanteamiento sobre qué tipos de dinámicas sociales pueden capturarse de manera efectiva en la plataforma y cuáles podrían estar subrepresentadas o ausentes en estudios futuros.

A estos problemas se suman los desafíos éticos asociados con la privacidad y el consentimiento. Nellore *et al.* (2023) destacan que el uso de datos de redes sociales plantea dilemas complejos, especialmente en lo que respecta a la protección de los datos personales y el cumplimiento de normativas en evolución. Este panorama ético se ha vuelto más intrincado debido a la creciente concienciación pública sobre los riesgos asociados con el manejo de datos personales y la necesidad de estándares más estrictos. Bluemke *et al.* (2023) sugieren que la presión regulatoria y el escrutinio público obligarán a los investigadores a implementar estrategias más rigurosas para garantizar el cumplimiento ético y la transparencia en sus métodos.

Además, las metodologías empleadas para la recopilación y análisis de datos presentan limitaciones adicionales. La limpieza y organización de grandes volúmenes de datos no estructurados de Twitter requieren procesos exigentes y a menudo costosos, que consumen tiempo y recursos significativos. Kim *et al.* (2013) destacan que estas etapas pueden introducir sesgos adicionales si no se manejan con cuidado, afectando la precisión y fiabilidad de los análisis. La creciente presencia de *bots* y cuentas automatizadas en la plataforma complica aún más este panorama, ya que estos actores generan contenido que no refleja comportamientos humanos reales, distorsionando las conclusiones de los estudios.

Por último, los desafíos relacionados con la integridad de los datos y las estrategias de muestreo también son significativos. La abundancia de datos en Twitter no garantiza necesariamente su calidad o relevancia. Los investigadores deben ser críticos con respecto a las limitaciones inherentes a los datos y las metodologías disponibles. Rivadeneira (2023) argumenta que las preocupaciones sobre la representatividad y la ética exigen una revisión continua de las prácticas de investigación para adaptarse a los cambios en el ecosistema digital.

En conjunto, estos desafíos resaltan la necesidad de adoptar metodologías más sofisticadas y éticamente responsables para maximizar el potencial de los datos de Twitter en la investigación científica. Si bien las barreras son significativas, también representan oportunidades para innovar en el diseño de estudios y en el desarrollo de herramientas analíticas que permitan superar estas limitaciones. En un contexto digital en constante cambio, la capacidad de los investigadores para adaptarse y anticipar nuevos desafíos será fundamental para garantizar la relevancia y el impacto de sus contribuciones académicas.

5. DISCUSIÓN Y CONCLUSIONES

Esta revisión de la literatura aborda de manera integral los factores que han impulsado el auge de la investigación basada en datos de Twitter, las tendencias emergentes en este campo y las principales dificultades que enfrenta la comunidad académica al trabajar con este tipo de datos. A partir de los objetivos planteados, los resultados reflejan cómo la plataforma ha transformado la generación de conocimiento académico, destacándose por su capacidad para capturar en tiempo real dinámicas sociales y ofrecer un acceso sin precedentes a discursos públicos, aunque no sin limitaciones.

El análisis de los factores que han contribuido al crecimiento de la investigación con datos de Twitter muestra que la combinación de accesibilidad, avances tecnológicos y un interés académico en comprender fenómenos sociales contemporáneos ha fortalecido su adopción. Sin embargo, los resultados también revelan que el acceso restringido a los datos, las limitaciones impuestas por la API y la dependencia de herramientas de terceros han condicionado el alcance de estas investigaciones. Estas barreras afectan no solo la calidad y consistencia de los estudios, sino también su capacidad para capturar representaciones completas y no sesgadas de la realidad social. Este panorama plantea la necesidad de desarrollar enfoques que reduzcan estas limitaciones, garantizando análisis más robustos y comparables.

En cuanto a las tendencias futuras, los resultados apuntan a una diversificación de las fuentes de datos, con un interés creciente en plataformas como TikTok e Instagram. Este fenómeno responde, en parte, a los cambios demográficos en la base de usuarios de Twitter y a las diferencias en las dinámicas de interacción que ofrecen otras plataformas. La exploración de un enfoque multiplataforma, identificado como una tendencia emergente, promete ampliar la comprensión de las dinámicas sociales y permite comparar patrones de participación pública en distintos entornos digitales. Esta perspectiva ofrece una oportunidad para enriquecer el análisis, pero también demanda herramientas metodológicas consistentes que permitan integrar datos de fuentes heterogéneas.

Finalmente, en lo que respecta a las dificultades reportadas por los investigadores, los resultados destacan que, además de los desafíos en el acceso a los datos, existen preocupaciones éticas asociadas con la privacidad y el consentimiento. El uso creciente de *bots* y cuentas automatizadas complica la autenticidad de los datos recolectados y plantea interrogantes sobre su validez para representar comportamientos humanos reales. Estas dificultades, junto con la necesidad de garantizar la transparencia y la interpretabilidad en el uso de herramientas avanzadas como el aprendizaje automático, subrayan la importancia de establecer estándares claros en el uso ético de datos de redes sociales.

En este contexto, se propone el diseño de metodologías específicas que aborden los sesgos demográficos identificados en los datos de Twitter. Estas metodologías podrían incluir estrategias de muestreo que consideren explícitamente las características demográficas y de comportamiento de los usuarios, así como técnicas analíticas que compensen la subrepresentación de ciertos grupos. Asimismo, se recomienda la promoción de estándares éticos claros y ampliamente aceptados para guiar la investigación basada en datos de redes sociales. Esto incluye la implementación de protocolos que garanticen el manejo responsable de la privacidad, el consentimiento informado y la minimización de posibles daños a las comunidades estudiadas.

En síntesis, los objetivos de esta revisión se reflejan en los hallazgos que destacan tanto las oportunidades como las limitaciones en el uso de datos de X para la investigación académica. Este ejercicio, además de evidenciar la relevancia de la plataforma en la producción de conocimiento interdisciplinario, invita a reflexionar sobre la necesidad de adaptar las prácticas investigativas a un panorama digital en constante cambio. Conectar los resultados con tendencias futuras, dificultades reportadas y recomendaciones prácticas permite delinear un camino que priorice enfoques metodológicos funcionales, reflexiones éticas y una comprensión crítica de las dinámicas sociales en un entorno digital en constante cambio.

6. REFERENCIAS

- Asur, S. y Huberman, B. A. (2010). Predicting the future with social media. HP Laboratories Technical Report, 53. Scopus. <https://acortar.link/295pNj>
- Blank, G. (2017). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review*, 35(6), 679-697. <https://doi.org/10.1177/0894439316671698>
- Bluemke, E., Collins, T., Garfinkel, B. y Trask, A. (2023). Exploring the relevance of data Privacy-Enhancing technologies for AI governance use cases. *arXiv preprint arXiv:2303.08956*. <https://doi.org/10.48550/ARXIV.2303.08956>
- Budiharto, W. y Meiliana, M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *Journal of Big Data*, 5(1). <https://doi.org/10.1186/s40537-018-0164-1>
- Cvetojevic, S. y Hochmair, H. H. (2018). Analyzing the spread of tweets in response to Paris attacks. *Computers, Environment and Urban Systems*, 71, 14-26. <https://doi.org/10.1016/j.compenvurbsys.2018.03.010>
- de Vreese, C. y Tromble, R. (2023). The Data Abyss: How Lack of Data Access Leaves Research and Society in the Dark. *Political Communication*, 40(3), 356-360. <https://doi.org/10.1080/10584609.2023.2207488>
- Dixon, S. J. (13 de septiembre de 2023). *U.S. annual X/Twitter growth 2025*. Statista. <https://www.statista.com/statistics/238729/twitters-annual-growth-rate-in-the-us/>
- Efanova, E. V. (2023). Twitter diplomacy in shaping the foreign policy agenda of the United States of America during the presidency of D. Trump. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 4. Istorija. Regionovedenie. Mezhdunarodnye otnosheniya*, 28(3), 97-103. <https://doi.org/10.15688/jvolsu4.2023.3.9>
- 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). <https://doi.org/10.1038/s41598-023-42227-6>
- Ghosh, S., Fernandez, J. M. Z., González, I. Z., Calle, A. M. y Shaghghi, N. (2023). Detecting Fake News Spreaders on Twitter Through Follower Networks. En R. Hou, H. Huang, D. Zeng, G. Xia, K. K. A. Ghany y H. M. Zawbaa (Eds.), *Big Data Technologies and Applications. BDTA BDTA 2022 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* (Vol. 480, pp. 181-195). Springer, Cham. https://doi.org/10.1007/978-3-031-33614-0_13
- Henry, D., Stattner, E. y Collard, M. (2018). Information Propagation Routes between Countries in Social Media. En Companion Proceedings of the The Web Conference 2018 (WWW '18) (pp. 1295-1298). International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/3184558.3191569>
- Hu, T., Wang, S., Luo, W., Zhang, M., Huang, X., Yan, Y., Liu, R., Ly, K., Kacker, V., She, B. y Li, Z. (2021). Revealing public opinion towards covid-19 vaccines with twitter data in the united states: Spatiotemporal perspective. *Journal of Medical Internet Research*, 23(9). <https://doi.org/10.2196/30854>
- Ioannides, G., Jadhav, A., Sharma, A., Navali, S. y Black, A. W. (2023). Compressed models for co-reference resolution: Enhancing efficiency with debiased word embeddings. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-45677-0>

- Kim, A. E., Hansen, H. M., Murphy, J., Richards, A. K., Duke, J. y Allen, J. A. (2013). Methodological considerations in analyzing twitter data. *Journal of the National Cancer Institute - Monographs*, 47, 140-146. <https://doi.org/10.1093/jncimonographs/lgt026>
- Maci, S., Demata, M., McGlashan, M. y Seargeant, P. (Eds.). (2024). *The Routledge handbook of discourse and disinformation*. Routledge.
- Mao, Y., Menchen-Trevino, E. y Cronin, J. (2023). Communicating environmental issues across media: An exploration of international news flows between twitter and traditional media. *Journal of International Communication*, 29(1), 39-61. <https://doi.org/10.1080/13216597.2022.2149605>
- Mazhar, K. y Dwivedi, P. (2024). Decoding the black box: LIME-assisted understanding of Convolutional Neural Network (CNN) in classification of social media tweets. *Social Network Analysis and Mining*, 14(1), 133. <https://doi.org/10.1007/s13278-024-01297-8>
- McLachlan, S. (7 de mayo de 2024). 19 X (Twitter) Demographics for Marketers in 2024. *Hootsuite*. https://blog.hootsuite.com/twitter-demographics/?utm_source=chatgpt.com
- Nellore, N., Zimmer, M. y Apostol, B. (2023). Updating the Topology of Twitter Research: Towards A Systematic Review of Twitter Research from 2013-2022. En Companion Proceedings of the 2023 ACM International Conference on Supporting Group Work (GROUP '23) (pp. 4-5). Association for Computing Machinery. <https://doi.org/10.1145/3565967.3570972>
- Nguyen, H.-D., Nguyen, D. Q., Nguyen, C.-D., To, P. T., Nguyen, D. H., Nguyen-Gia, H., Tran, L. H., Tran, A. Q., Dang-Hieu, A., Nguyen-Duc, A. y Quan, T. (2024). Supervised learning models for social bot detection: Literature review and benchmark. *Expert Systems with Applications*, 238. <https://doi.org/10.1016/j.eswa.2023.122217>
- Dersan Orhan, D. (2021). Making Foreign Policy Through Twitter: An Analysis of Trump's Tweets on Iran. En E. Esiyok (Ed.), *Handbook of Research on New Media Applications in Public Relations and Advertising* (pp. 380-394). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-7998-3201-0.ch022>
- Pagolu, V. S., Reddy, K. N., Panda, G. y Majhi, B. (2017). Sentiment analysis of Twitter data for predicting stock market movements. En *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPE5)* (pp. 1345-1350). IEEE. <https://doi.org/10.1109/SCOPE5.2016.7955659>
- Raja, H., Ilyas, M. U., Saleh, S., Liu, A. X. y Radha, H. (2016). Detecting national political unrest on Twitter. En *2016 IEEE International Conference on Communications (ICC)* (pp. 1-7). <https://doi.org/10.1109/ICC.2016.7511393>
- Rangel, F., Giachanou, A., Ghanem, B. y Rosso, P. (2020). Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter. En *CEUR workshop proceedings* (Vol. 2696, pp. 1-18). Sun SITE Central Europe. <https://acortar.link/RAgg5H>
- Rivadeneira, L. (2023). Análisis del comportamiento de decisión usando datos de Twitter: Una revisión de la literatura. *Revista Ibérica de Sistemas e Tecnologias de Informação*, E61, 307-321. <https://acortar.link/kzucGg>
- Schroeder, R. y Cows, J. (2019). Big Data Approaches to the Study of Digital Media. En J. Hunsinger, M. Allen y L. Klastrup (Eds.), *Second International Handbook of Internet Research* (pp. 957-977). https://doi.org/10.1007/978-94-024-1555-1_13

- Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L. y Merchant, R. M. (2017). Twitter as a tool for health research: A systematic review. *American Journal of Public Health*, 107(1), e1-e8. <https://doi.org/10.2105/AJPH.2016.303512>
- Soaita, A. M., Serin, B. y Preece, J. (2020). A methodological quest for systematic literature mapping. *International Journal of Housing Policy*, 20(3), 320-343. <https://doi.org/10.1080/19491247.2019.1649040>
- Steinert-Threlkeld, Z. C. (2018). *Twitter as data*. Cambridge University Press. <https://doi.org/10.1017/9781108529327>
- Wang, Z., Lam, N. S. N., Obradovich, N. y Ye, X. (2019). Are vulnerable communities digitally left behind in social responses to natural disasters? An evidence from Hurricane Sandy with Twitter data. *Applied Geography*, 108, 1-8. <https://doi.org/10.1016/j.apgeog.2019.05.001>
- Weller, K. (2014). What do we get from twitter- and what not? A close look at twitter research in the social sciencest. *Knowledge Organization*, 41(3), 238-248. <https://doi.org/10.5771/0943-7444-2014-3-238>
- Yu, J. y Muñoz-Justicia, J. (2022). Free and Low-Cost Twitter Research Software Tools for Social Science. *Social Science Computer Review*, 40(1), 124-149. <https://doi.org/10.1177/0894439320904318>
- Yuan, Y. (2017). Modeling inter-country connection from geotagged news reports: A time-series analysis. En Y. Tan, H. Takagi y Y. Shi (Eds.) *Data Mining and Big Data. DMBD 2017. Lecture Notes in Computer Science* (Vol. 10387, pp.183-190). Springer, Cham. https://doi.org/10.1007/978-3-319-61845-6_19
- Zou, Y. y Herremans, D. (2023). PreBit — A multimodal model with Twitter FinBERT embeddings for extreme price movement prediction of Bitcoin. *Expert Systems with Applications*, 233. <https://doi.org/10.1016/j.eswa.2023.120838>

CONTRIBUCIONES DE AUTORES/AS, FINANCIACIÓN Y AGRADECIMIENTOS

Contribuciones de los/as autores/as:

Conceptualización: Rivadeneira, Lucía. **Software:** Rivadeneira, Lucía. **Validación:** Loor, Ignacio. **Análisis formal:** Rivadeneira, Lucía y Loor, Ignacio. **Curación de datos:** Rivadeneira, Lucía y Loor, Ignacio. **Redacción-Preparación del borrador original:** Rivadeneira, Lucía. **Redacción-Revisión y Edición:** Loor, Ignacio. **Visualización:** Rivadeneira, Lucía. **Supervisión:** Rivadeneira, Lucía y Loor, Ignacio. **Administración de proyectos:** Rivadeneira, Lucía y Loor, Ignacio. **Todos los/as autores/as han leído y aceptado la versión publicada del manuscrito:** Rivadeneira, Lucía y Loor, Ignacio.

Financiación: Esta investigación no recibió financiamiento externo.

Conflicto de intereses: No existen.

AUTOR/A/ES/AS:

Lucía Rivadeneira

Universidad Técnica de Manabí.

Lucía Rivadeneira es doctora en Administración y Negocios por la Universidad de Manchester, Reino Unido; tiene una maestría en Sistemas de Información por la Universidad Tecnológica de Nanyang, Singapur, y es ingeniera de Sistemas por la Universidad Técnica de Manabí, Ecuador. Actualmente es docente e investigadora en la Facultad de Ciencias Informáticas de la Universidad Técnica de Manabí, donde se especializa en el análisis de datos de redes sociales para modelar fenómenos sociales. Sus áreas de interés incluyen inteligencia artificial, aprendizaje automático, análisis de datos no estructurados, análisis de sentimiento, análisis de redes sociales y modelos de clasificación y optimización.

lucia.rivadeneira@utm.edu.ec

Orcid ID: <https://orcid.org/0000-0001-5989-7703>

Scopus ID: <https://www.scopus.com/authid/detail.uri?authorId=57221283412>

Ignacio Loor

Universidad Técnica de Manabí.

Ignacio Loor es doctor en Geografía Humana por la Universidad de Manchester, Reino Unido; tiene una maestría en Negocios Internacionales por Nova Southeastern University, Estados Unidos, y es economista por la Universidad Católica Santiago de Guayaquil, Ecuador. Actualmente se desempeña como investigador en urbanismo y desarrollo sostenible y como Vicedecano de Investigación de la Facultad de Ciencias Humanísticas y Sociales en la Universidad Técnica de Manabí, Ecuador. Sus áreas de interés incluyen la infraestructura de asentamientos informales, la infraestructura verde, la organización social de comunidades informales, las prácticas informales y la transición hacia emisiones netas cero de carbono.

ignacio.loor@utm.edu.ec

Índice H: 6

Orcid ID: <https://orcid.org/0000-0003-4806-1032>

Scopus ID: <https://www.scopus.com/authid/detail.uri?authorId=57224464879>

ResearchGate: <https://www.researchgate.net/profile/Ignacio-Loor>

Artículos relacionados:

- Arce Garcia, S., Cano Garcinuño, M. I., Quiles Cano, C. y Cano Pérez, J. (2023). Vacunas anticovid y trombosis: el miedo en las redes sociales. *Revista de Comunicación y Salud*, 14, 1-19. <https://doi.org/10.35669/rcys.2024.14.e307>
- Demuner Flores, M. del R. (2021). Uso de redes sociales en microempresas ante efectos COVID-19. *Revista de Comunicación de la SEECI*, 54, 97-118. <https://doi.org/10.15198/seeci.2021.54.e660>
- Martínez-Fresneda Osorio, H. y Sánchez Rodríguez, G. (2022). La influencia de Twitter en la agenda *setting* de los medios de comunicación. *Revista de Ciencias de la Comunicación e Información*, 27, 1-21. <https://doi.org/10.35742/rcci.2022.27.e136>
- Moreno Cabanillas, A. y Castellero Ostio, E. (2023). Comunicación política y redes sociales: análisis de la comunicación en Instagram de la campaña electoral del 13-F. *Vivat Academia*, 156, 199-222. <https://doi.org/10.15178/va.2023.156.e1461>
- Pérez Altable, L. y Serrano-Tellería, A. (2021). Communications patterns and power dynamics in the digital public sphere: A case study of the conversation about Minimum Living Income on Twitter. *European Public & Social Innovation Review*, 6(1), 1-15. <https://epsir.net/index.php/epsir/article/view/148>