

Estado de la publicación: El preprint ha sido publicado como artículo en una revista
DOI del artículo publicado: <https://doi.org/10.54786/revistaepic.v27i1.20954>

Wikipedia, de los bots a la IA. Una historia de colaboración sociotécnica

Agustín Zanotti

<https://doi.org/10.1590/SciELOPreprints.11137>

Enviado en: 2025-01-27

Postado en: 2025-02-03 (versión 1)

(AAAA-MM-DD)

Wikipedia, de los bots a la IA. Una historia de colaboración sociotécnica
Wikipedia, from bots to AI. A story of sociotechnical collaboration
Wikipedia, dos bots à IA. Uma história de colaboração sociotécnica

Agustín Zanotti

CCONFINES Universidad Nacional de Villa María- CONICET, Argentina

 <https://orcid.org/0000-0002-7662-7593>

azanotti@unvm.edu.ar

Resumen

El artículo examina a Wikipedia y sus procesos de automatización. Esta enciclopedia sin ánimo de lucro se destaca por su modelo de gobernanza, que involucra a organizaciones y comunidades locales. Desde sus inicios, Wikipedia ha implementado bots para garantizar la sostenibilidad del proyecto y fomentar la participación de voluntarios. En la última década, ha promovido activamente el uso de técnicas de aprendizaje automático y aplicaciones de inteligencia artificial. El estudio se basa en fuentes secundarias y datos recopilados de la plataforma, utilizando un enfoque de investigación-acción. Las conclusiones abordan críticamente la integración de este proyecto en el actual ecosistema mediático.

Palabras clave: producción entre pares; sistema sociotécnico; ecosistema de medios; aprendizaje automático; inteligencia artificial

Abstract

The article examines Wikipedia and its automation processes. This non-profit encyclopedia stands out for its governance model, which involves organizations and local communities. Since its inception, Wikipedia has implemented bots to ensure project sustainability and encourage volunteer participation. Over the past decade, it has actively promoted the use of machine learning techniques and artificial intelligence applications. The study is based on secondary sources and data collected from the platform, employing a research-action approach. The conclusions critically address the integration of this project into the current media ecosystem.

Keywords: peer production; sociotechnical system; media ecosystem; machine learning; artificial intelligence

Resumo

O artigo analisa Wikipédia e seus processos de automação. Esta enciclopédia sem fins lucrativos destaca-se pelo seu modelo de governança, que envolve organizações e comunidades locais. Desde o seu início, a Wikipédia implementou bots para garantir a sustentabilidade do projeto e incentivar a participação voluntária. Na última década, tem promovido ativamente o uso de técnicas de aprendizado de máquina e aplicativos de inteligência artificial. O estudo baseia-se em fontes secundárias e dados coletados da plataforma, empregando uma abordagem de pesquisa-ação. As conclusões abordam criticamente a integração deste projeto no atual ecossistema midiático.

Palavras-chave: produção entre pares; sistema sociotécnico; ecossistema de mídia; aprendizado de máquina; inteligência artificial

Introducción

El artículo examina a Wikipedia y sus procesos de automatización. Este proyecto se basa en el trabajo colaborativo entre pares, con el objetivo inagotable de crear una enciclopedia que recopile “la suma del conocimiento humano” y en la que “todos puedan editar”¹. Es dirigido por una fundación sin fines de lucro a nivel mundial, en colaboración con organizaciones y comunidades locales.

Para comprender esta iniciativa, es esencial analizar un entramado complejo de formas de gobernanza. Encontramos una multiplicidad de instancias deliberativas que moldean un movimiento diverso y global. Basándonos en el constructivismo social de la tecnología, adoptamos la definición de trayectoria sociotécnica, entendida como un proceso de "co-construcción de productos, procesos productivos y organizacionales" (Thomas, 2008, p. 249). A través de un recorrido por distintos momentos y oleadas de automatización, el objetivo del artículo es comprender en qué medida las contribuciones humanas y automáticas constituyen partes complementarias de un sistema sociotécnico (Niederer & van Dijck, 2010).

A lo largo de su trayectoria, Wikipedia ha enfrentado desafíos y momentos de incertidumbre con respecto a su futuro. Desde sus inicios, surgió como un proyecto inesperado derivado de Nupedia, el proyecto principal. Durante sus primeros años, se enfrentó a dificultades para lograr una cobertura y calidad adecuadas en sus contenidos, mientras que movimientos radicales y actos de vandalismo ocuparon una parte significativa de su atención. Más recientemente, se ha visto confrontada por una crisis de escala, con un número limitado de editores para abordar una vasta cantidad de tareas de creación y mantenimiento en diversos formatos (Reagle, 2020).

Una parte importante de los desafíos señalados pudieron superarse gracias a la implementación de diferentes estrategias de automatización. Wikipedia destaca por la incorporación temprana de bots -piezas de software o scripts diseñados para realizar ediciones automatizadas- como parte de sus rutinas de creación de contenidos. La presencia de bots sirvió para garantizar la sostenibilidad del proyecto y facilitar la participación, acompañando las tareas humanas. Estos desarrollos se han ido nutriendo de avances en la inteligencia artificial (IA), que van desde el uso de aprendizaje automático hasta modelos generativos.

Metodológicamente, el trabajo se basa en el análisis de fuentes secundarias y datos de la plataforma wiki. Se incorporaron investigaciones especializadas en procesos de automatización en Wikimedia, así como coberturas periodísticas sobre temas recientes. Se seleccionaron entradas de la sección Tech del sitio de Wikimedia Foundation (WMF) y el portal Diff de noticias del movimiento. Se utilizaron herramientas analíticas como Wikimedia

¹ Son estos dos de los eslóganes principales del proyecto.

Statistics y Wikiscan para la elaboración de gráficos de tendencias generales. El estudio se enmarca en un abordaje de investigación-acción en curso, en el cual se realizaron unas 10.000 intervenciones sobre proyectos Wikimedia, entre ellas la escritura y revisión de artículos enciclopédicos, la carga de objetos digitales, la realización de editatones (Aimar, Pagola & Zanotti, 2021) y la actualización de portales temáticos (Zanotti, 2022a).

A lo largo de la presentación, repasamos algunos debates sobre tendencias macro del capitalismo actual. Nos detenemos luego en tres momentos de implementación de técnicas de automatización y sus consecuencias. Nos enfocamos en la discusión más reciente sobre los alcances de la IA generativa y sus impactos sobre los proyectos wiki, así como la articulación entre actores al interior del ecosistema de medios actual.

1. Capitalismo informacional y desarrollos de IA

Podemos definir al capitalismo actual como un capitalismo digital (Fuchs, 2020) o informacional (Zuckerfeld, 2020). Tres macrotendencias atraviesan el desarrollo informacional actual:

1. la informacionalización, donde cada vez más trabajadores tienen como principal actividad laboral la producción de alguna forma de información digital y como medio de trabajo una o varias tecnologías digitales;
2. la plataformización, esto es, la intermediación de todo tipo de actividades a través de plataformas digitales que se desarrollan a nivel de infraestructura y en diferentes nichos de mercado;
3. la automatización, es decir el reemplazo de trabajo humano por tecnologías digitales y software, por contenidos audiovisuales o combinaciones de software y sensores (Zuckerfeld, 2020).

Estos tres procesos se retroalimentan mutuamente. La informacionalización y plataformización se complementan para transformar los trabajos del sector de servicios, tensionando la noción de trabajo y empleo. La plataformización supone además un proceso creciente de automatización, en la medida en que las plataformas desarrollan cada vez más sofisticados algoritmos, que permiten la recopilación masiva de datos y su procesamiento en tiempo real, generando la personalización de contenidos a gran escala (Zuckerfeld, 2020).

En el artículo nos enfocamos en el tercero de estos procesos. La automatización ha sido una característica constante en el desarrollo industrial moderno, impulsada por la competencia por mayor productividad laboral, el control gerencial y la acumulación de capital (Smith & Fressoli, 2021). Sin embargo, a nivel global, diferentes comunidades están utilizando la automatización en escenarios alternativos. Entre ellas podemos mencionar los movimientos de

innovación de base, la producción entre pares, el campo de la producción colaborativa y el diseño abierto (Chierchie & Fressoli, 2022). El término post-automatización surge principalmente en estos espacios donde las tecnologías se reconstruyen en función de relaciones más plurales, arraigadas en la creatividad humana, la convivencia y el cuidado (Smith & Fressoli, 2021).

En particular, nos concentramos en los desarrollos de la última década con relación a la IA. Se considera que el término "inteligencia artificial" fue acuñado por John McCarthy en 1956, para resolver problemas mediante programas informáticos capaces de simular procesos cognitivos humanos. Impulsados por grandes cantidades de datos y mejoras en el hardware de computación se producen en las últimas décadas avances notables que dan lugar a una integración de la IA en diversas industrias y en la vida cotidiana (Preston, 2022).

Un modelo de IA es una representación matemática o computacional que se utiliza para resolver tareas y problemas específicos mediante el aprendizaje y la toma de decisiones automatizada. El aprendizaje en IA se refiere a la capacidad de un sistema para mejorar su rendimiento a medida que se le proporciona más información o experiencia. Un sistema de IA se entrena utilizando datos y ejemplos relevantes para aprender patrones y reglas que le permitan realizar tareas específicas o tomar decisiones² (Gómez-Zermeño, 2023).

Siguiendo a Dyer-Whiteford, Kjosen & Steinhoff (2019), podemos discernir entre tres niveles de IA: estrecha, general y superinteligencia. La IA estrecha es capaz de ejecutar tareas específicas dentro de un ámbito limitado. Por otro lado, la IA general puede abordar una amplia variedad de tareas en dominios diversos, mientras que la superinteligencia sobrepasará la capacidad humana en múltiples áreas. En la medida en que la IA ha sido impulsada mayormente por gigantes tecnológicos para fines comerciales, podemos definirla como una IA-capitalista.

Su desarrollo ha sido alimentado por la extracción masiva de datos, que pueden representar procesos de acumulación primitiva u originaria. Estos se basan en un modelo de profit from openness (Lund & Zukerfeld, 2020). Mediante la apropiación impaga (legal o no) de conocimientos digitales, se entrenan algoritmos cuyos productos y servicios luego comercializan con fines de lucro (Zukerfeld *et al.*, 2023). En los últimos años, la creciente

² Algunos de los marcos conceptuales más importantes en el campo de la IA son:

Aprendizaje Automático (*Machine Learning*): Creación de sistemas que pueden aprender de los datos y mejorar su rendimiento en tareas específicas a medida que se les proporciona más información. Incluye subáreas como el aprendizaje supervisado, no supervisado y por refuerzo.

Procesamiento del Lenguaje Natural: Creación de sistemas que comprenden y generan lenguaje humano. Se utiliza en aplicaciones como traducción automática, chatbots y análisis de sentimientos.

Redes Neuronales Artificiales: Modelización de sistemas de IA inspirados en la estructura y el funcionamiento del cerebro humano. Las redes neuronales artificiales son esenciales en el aprendizaje profundo y se utilizan en una amplia variedad de aplicaciones.

utilización de IA ha incrementado sus impactos sociales, políticos y ambientales. La toma de decisiones automatizada, los sesgos algorítmicos y la vigilancia digital, son sólo algunos de los aspectos que apuntan a la creación de marcos regulatorios asociados con estas tecnologías en constante evolución (Crawford, 2021).

2. Wikipedia y automatización

Wikipedia cumplió 20 años en el año 2021, ubicándose entre los proyectos más duraderos y un caso de éxito en un ecosistema de medios complejo y dinámico (Van Dijck, 2019). La enciclopedia logró desplegar un modelo de gobernanza que podemos entender como gobernanza de pares (Zanotti, 2022b). Siguiendo a Bauwens (2005), el término se refiere a un modo ascendente de toma de decisiones participativas. En ésta, no operan jerarquías estrictas de comando y control, sino un entorno más flexible que “permite la existencia de múltiples equipos de participantes que trabajan simultáneamente en una variedad de direcciones, posiblemente opuestas” (Bruns, 2008, p. 26).

Como anticipamos, proponemos definir a Wikipedia como un sistema sociotécnico en evolución, donde, en sintonía con los planteos de Latour (2008), la interrelación dinámica entre agentes humanos y no humanos es un aspecto crucial de su funcionamiento. El éxito de la enciclopedia se basa en un control sociotécnico, una combinación de su infraestructura técnica y el trabajo colaborativo de usuarios y editores (Niederer & van Dijck, 2010).

2.1 Bots

Los bots son piezas de software o scripts diseñados para realizar ediciones sin necesidad de toma de decisiones humanas. En los últimos años, ha habido un enorme aumento de bots en línea, que van desde rastreadores web para motores de búsqueda, hasta chatbots para servicio al cliente en línea, spambots y bots de edición de contenido (Tsvetkova *et al.*, 2017).

En el caso de Wikimedia, los bots son en general diseñados y mantenidos por usuarios voluntarios. En general existen dos tipos de bots, los de edición y los administrativos. Mientras los primeros participan en la creación y mantenimiento de entradas, los segundos son desplegados para detectar y revertir el vandalismo, monitorear artículos y bloquear usuarios, entre otras funciones. En general utilizan la infraestructura de Wikimedia y están alojados en el entorno de Toolforge (Wikipedia:Bots, 2024).

Cada edición idiomática puede tener su propia política de bots y estos deben pasar por un periodo de revisión comunitaria para su aprobación. En el caso de la Wikipedia en inglés, un comité de voluntarios revisa las nuevas solicitudes y ayuda a resolver conflictos relacionados con la automatización (Geiger, 2011). Una vez aprobados, los bots obtienen su

página de usuario, nivel de acceso y derechos administrativos (Wikipedia:Bot policy, 2024).

El primer bot en la Wikipedia en inglés se creó en 2002. Para 2006, el número de bots había aumentado a 151, y para 2008, llegó a 457 (Niederer y van Dijck, 2010). A medida que pasaba el tiempo, los bots asumieron un papel cada vez más significativo en la edición de esta Wikipedia. Entre 2005 y 2006, los bots fueron responsables de alrededor del 5% de las ediciones, mientras que en 2009 representaron el 16%, y en 2012 superaron el 20% (Clement, 2015). Para el año 2024, la Wikipedia en inglés cuenta con 2.682 tareas automatizadas. Hay 298 bots activos y más de 400 que han sido utilizados anteriormente. Junto con ello, se han desarrollado una variedad de herramientas que permiten la edición semiautomatizada de grandes cantidades de artículos (Wikipedia:Bots, 2024).

En los primeros años de Wikipedia, los bots desempeñaron un papel importante en la creación de artículos. En la edición en inglés, los bots generaron artículos sobre temas de tecnología, religión, ciudades, poblaciones y asteroides. Un caso conocido fue el de Rambot, que generó más de 30.000 artículos sobre ciudades y condados de los Estados Unidos, utilizando datos de la CIA y del censo (Niederer y van Dijck, 2010).

En la Wikipedia en italiano, Gacbot se utilizó desde 2004 para generar los municipios de Italia, Polonia, Botsuana, Estados Unidos, Alemania, Finlandia, Portugal y parcialmente Francia (Wikipedia:History of Wikipedia bots, 2024). En la Wikipedia Sueca, Lsjbot desempeñó un papel significativo al crear más de 450.000 artículos, lo que llegó a representar en 2013 casi la mitad de esta enciclopedia. El bot se enfocó en artículos sobre especies de animales y plantas. Para ello, recopiló datos de múltiples fuentes y luego adaptó la información al formato wiki (Guldbrandsson, 17 junio 2013).

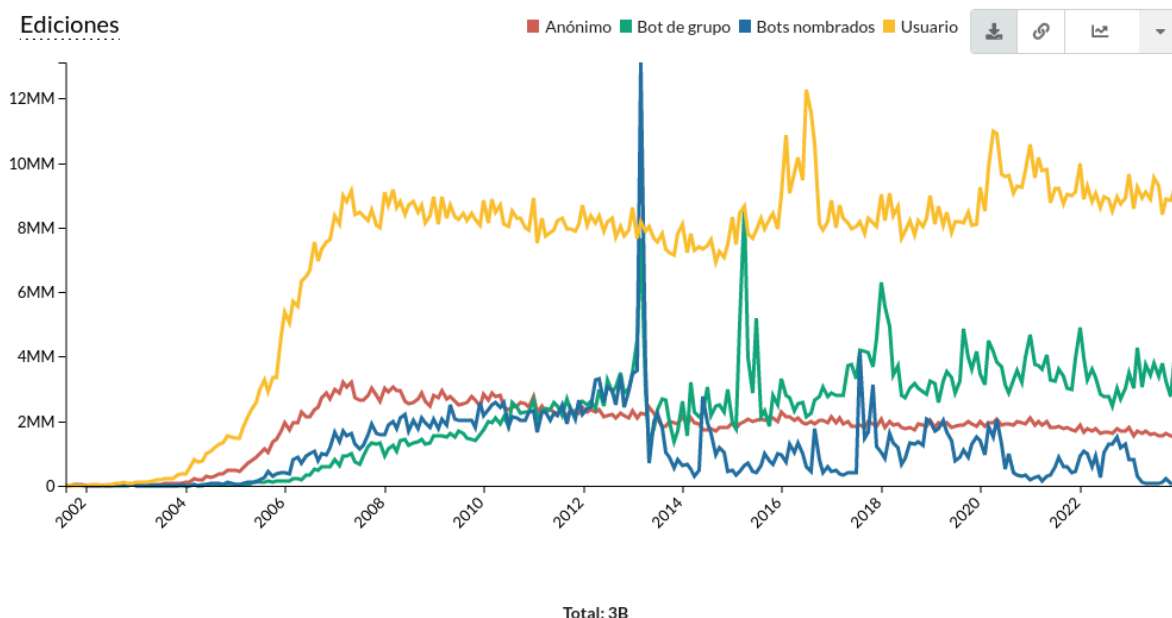
En los últimos años, Lsjbot tradujo unos 6 millones de artículos para la Wikipedia en cebuano, a partir de traducciones masivas de la Wikipedia en inglés. Gracias a este impulso, a pesar de ser una comunidad con solo 157 usuarios activos, la Wikipedia en cebuano se convirtió en el segundo idioma con mayor cantidad de artículos (Cebuano Wikipedia, 2024).

Con el transcurso del tiempo, Wikipedia comenzó a implementar el aprendizaje automático en sus bots con el fin de ayudar a detectar vandalismo. Uno de los bots más destacados es ClueBot NG, que ha estado activo desde 2012. Este utiliza un modelo de aprendizaje automático entrenado a partir de datos que han sido revisados por humanos. En otras palabras, los voluntarios utilizan una interfaz para reconocer manualmente una edición como vandalismo o no. Luego, el modelo es entrenado con esos datos para identificar nuevas ediciones sospechosas y revertirlas (Wikimedia Meta-wiki, 2024).

Las herramientas analíticas de Wikipedia nos permiten extraer consideraciones adicionales sobre el trabajo de los bots en la producción de la enciclopedia. El Gráfico 1

muestra la participación según el tipo de usuario³ para el conjunto de las ediciones idiomáticas.

Gráfico 1. Ediciones según tipo de usuario.
Todas las Wikipedias. Por año y miles de millones de ediciones



Fuente: Wikimedia Statistics (2024).

Los datos presentados permiten observar una presencia de bots que se incrementa en los primeros años del proyecto y se mantiene constante hasta el presente. Algunos bots se ejecutan de manera permanentemente, mientras otros se activan por periodos de tiempo o para tareas puntuales. De hecho algunos picos evidencian la activación de bots de edición masiva, que llegan a superar a los colaboradores humanos durante un periodo puntual.

La participación de usuarios registrados se presenta como la fuente principal de ediciones. Luego de un periodo de crecimiento, se mantiene constante, con algunas subas y bajas, durante el ciclo de vida del proyecto. Se observa además un leve decrecimiento de los usuarios anónimos a lo largo del tiempo.

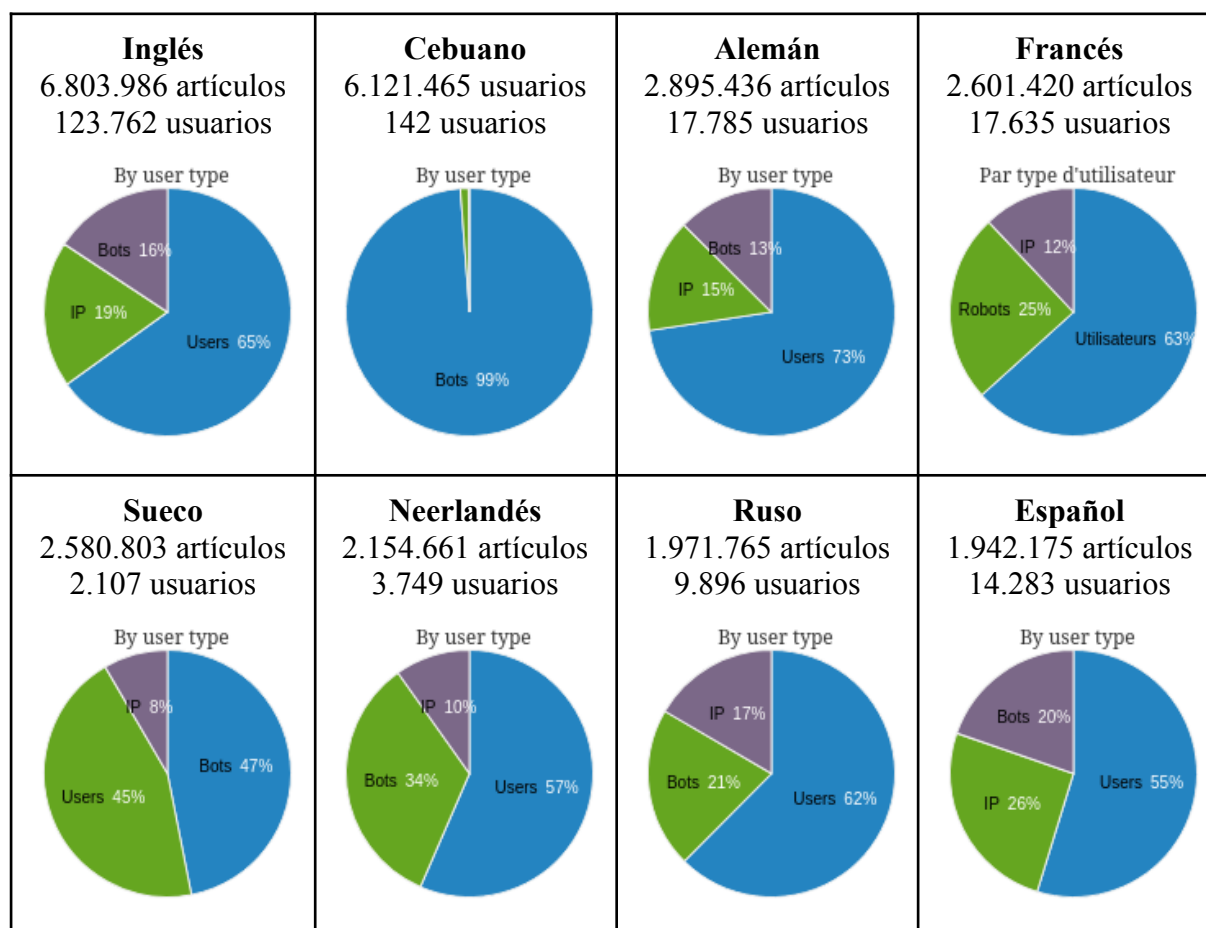
Las distintas ediciones idiomáticas de Wikipedia muestran una variación considerable en cuanto a la utilización de automatización. El Gráfico 2 permite comparar la actividad de los bots en los ocho principales idiomas. Mientras que en la Wikipedia en Cebuano el 99% de

³ Estos pueden ser: Anónimo, usuarios que no han iniciado sesión y son identificados por su número de IP; Bot de grupo, usuarios que han iniciado sesión y están en el grupo de usuarios registrados como bot; Bots nombrados, usuarios que han iniciado sesión cuyo nombre contiene la palabra "bot". Estos usuarios tienen una alta probabilidad de ser un bot, aunque puedan existir excepciones; Usuario, usuarios que han iniciado sesión y no son considerados un bot por nombre o grupo (Wikimedia Meta-wiki, 2024c).

las ediciones fueron realizadas por bots, la edición sueca muestra una participación del 47% y la neerlandesa del 34%. Las Wikipedias en francés, ruso y español presentan una participación de bots que oscila entre el 21% y el 26%. Por otro lado, la Wikipedia inglesa y la alemana son las de menor edición automatizada, con registros inferiores al 20%.

Gráfico 2. Cantidad de artículos, usuarios activos y ediciones según tipo de usuario.

Ocho mayores wikipedias



Fuentes: Wikiscan (29 marzo 2024) y Wikimedia Meta-wiki (2024a).

Los datos precedentes indican además que, a mayor tamaño de la comunidad editora, menor es la participación relativa de los bots. Sin embargo, esta se mantiene presente en todos los idiomas analizados.

La literatura disponible sostiene que la actividad de los bots es ampliamente aceptada, incluso cuando se trata de los más intrusivos. Estos son considerados valiosos para la viabilidad de Wikipedia y defendidos por los editores activos, al tiempo que su trabajo es auditado intensamente por parte de la comunidad (Clement, 2015). Los bots además interactúan entre sí y, a menudo, pueden incluso deshacer las ediciones de otros bots (Tsvetkova *et al.*, 2017).

2. 2 Aprendizaje automático

Además de los bots desarrollados por la comunidad, Wikimedia implementa y mantiene cientos de modelos de aprendizaje automático que cumplen diversos roles, como traducir artículos y mejorar la experiencia de los editores. Estos modelos incluyen aquellos que predicen temas, calidad de artículos y de ediciones. Muchos de estos modelos son específicos para cada idioma (Wikimedia Meta-wiki, 2024b).

La WMF está liderando una iniciativa para modernizar los sistemas de aprendizaje automático, permitiendo la implementación de una amplia gama de modelos. Es fundamental destacar que el desarrollo de herramientas de IA es solo una fase del proceso, que incluye su puesta en producción y el seguimiento de su desempeño. El área de operaciones se encarga de gestionar la implementación de estos avances.

Wikimedia se esfuerza por ser un ejemplo público de prácticas éticas en el aprendizaje automático, abogando por la transparencia y la responsabilidad en el desarrollo de modelos. Los principios que guían los sistemas de aprendizaje automático en Wikimedia se centran en la escalabilidad, agilidad, colaboración, participación humana, accesibilidad y gobernanza comunitaria. En este enfoque, es fundamental para garantizar que los modelos sean comprensibles y estén abiertos a los comentarios de los usuarios. Además, se busca hacer que el desarrollo y la gobernanza de los modelos sean accesibles, permitiendo la participación de personas con diferentes niveles de habilidad y experiencia, y garantizando que las comunidades tengan control sobre los modelos que las afectan (Wikimedia Meta-wiki, 2024b).

En los últimos años, Wikimedia ha fortalecido una comunidad técnica de desarrolladores con el objetivo de mejorar su infraestructura tecnológica. Como resultado de esta colaboración, han surgido una variedad de herramientas que van desde la edición y marcado de contenidos, protección de datos personales, trabajo colaborativo en línea, aplicaciones móviles, dashboards y herramientas de productividad, entre otras (Wikimedia Meta-wiki, 2024b). Al mismo tiempo, en 2023 se otorgaron los primeros subsidios del Fondo de Dotación para proyectos de innovación técnica, la mayoría de los cuales estaban vinculados con tareas de automatización (WMF, 13 abril 2023).

Recientemente, se lanzó un conjunto de funciones para apoyar a nuevos editores, que incluye un panel de control y un feed de ediciones sugeridas que utiliza el aprendizaje automático para ofrecer ediciones rápidas y fáciles (WMF, 6 diciembre 2022). También se han implementado técnicas de IA que permiten mejorar las búsquedas y realizar tareas sofisticadas en forma automatizada (Coleman, 9 enero 2018).

En 2023 se lanzó MinT, un servicio de traducción automática neuronal que ofrece soporte para más de 200 idiomas, incluyendo 44 lenguas menos difundidas que reciben traducción automática por primera vez. Los editores colaboran en la mejora continua de esta herramienta al traducir artículos y detectar posibles errores. MinT se basa en modelos de traducción desarrollados por empresas tecnológicas como Meta, en conjunto con universidades, los cuales están disponibles bajo licencias de código abierto. Este servicio busca posicionarse como una alternativa a otros modelos de traducción que también están disponibles en Wikipedia, algunos de ellos con licencias comerciales (Giner, 13 de junio de 2023).

Otro proyecto de gran envergadura es Abstract Wikipedia, lanzado en 2020. Se trata de una versión computable y multilingüe de la enciclopedia que, en lugar de presentar información en texto natural como la Wikipedia tradicional, representa conceptos de manera estructurada utilizando los datos de Wikidata. En Abstract Wikipedia, los usuarios pueden crear y mantener artículos independientemente del idioma, ya que la generación de traducciones se realiza automáticamente. Este proyecto se relaciona con Wikifunctions, una base de funciones computacionales abiertas lanzada en 2023 (Wikipedia:Abstract Wikipedia, 2024).

2. 3 IA generativa

La IA generativa emplea algoritmos y redes neuronales avanzadas para aprender de textos e imágenes, y posteriormente generar contenido nuevo y único, que se asemeja a algo creado por un ser humano (Gómez-Zermeño, 2023). Los modelos de lenguaje de gran tamaño (LLM, por sus siglas en inglés), como los sistemas GPT (*Generative Pre-trained Transformer*) desarrollados por OpenAI, son ejemplos destacados de esta tecnología. Estos modelos pueden ser muy complejos, con cientos de miles de millones de parámetros, como es el caso de GPT-3.

Los sistemas de IA generativa se entrenan a partir del raspado de sitios web⁴, que permite generar conjuntos de datos variados. Aunque las empresas tecnológicas generalmente no detallan las fuentes de datos utilizadas, sabemos que, en general los LLM se nutren con el contenido de Wikipedia, siendo en la mayoría de los casos una fuente principal para su entrenamiento (Deckelmann, 12 julio 2023). Esto se debe a que la enciclopedia es uno de los corpus abiertos más extensos en Internet, su contenido está altamente estructurado, validado tanto por fuentes como por los propios usuarios, y abarca más de 300 idiomas.

⁴ El raspado web, también conocido como *web scraping*, es el proceso de extracción de datos de sitios web. Utiliza bots de software para recorrer páginas web, analizar su estructura y extraer información de manera automatizada.

De acuerdo con Gertner, en base a consultas con referentes del sector tecnológico, existe una preocupación de que los nuevos desarrollos tecnológicos terminen canibalizando a Wikipedia (Gertner, 8 septiembre 2023). Su información fue siendo cada vez más «datificada» y extraída por motores de búsqueda y asistentes virtuales (Ford, 2020). Los enlaces de Wikipedia son muy frecuentes en los resultados de búsqueda, tanto en dispositivos móviles como en escritorio. Con frecuencia, su información se ubica en posiciones destacadas sobre diversas temáticas (Vincent & Hecht, 2021).

La otra cara de esta gran visibilidad es el riesgo de volverse dependiente de plataformas externas. Según datos de febrero de 2024, la principal vía de acceso a wikipedia.org es a través de motores de búsqueda, lo que representa el 84% del tráfico, mientras que el acceso directo al sitio solo cuenta por el 13.5% de las visitas (similarweb, 2024). Además, existen pocos incentivos para que los usuarios exploren temas dentro de la propia enciclopedia (Zanotti, 2022a).

McMahon, Johnson & Hecht (2017) lo plantean en términos de una “paradoja de la reutilización”. Si los resultados de buscadores no proporcionan enlaces a Wikipedia (u otras plataformas de producción colaborativa), pueden generar efectos adversos en estas comunidades, como una merma en la participación de nuevos miembros.

Este problema puede expandirse con la proliferación de la IA. Mientras los resúmenes de motores de búsqueda proporcionan enlaces y fuentes, las respuestas de IA generativa en general carecen de citas o referencias que respalden la información ofrecida. Esto se debe a la naturaleza probabilística inherente de estos modelos (Nmvg, 2 diciembre 2022).

Como un principio de respuesta a esta limitación, Wikipedia ha desarrollado un complemento experimental que permite a ChatGPT buscar y resumir información actualizada en Wikipedia en respuesta a consultas de conocimiento general. El complemento permite especificar la atribución que debe incluirse para este contenido y proporciona enlaces a los artículos para lecturas adicionales (Pinchuk, 13 julio 2023). Cabe destacar que, de momento, los complementos sólo están disponibles para la versión de pago.

Otra respuesta ha sido la creación de Wikimedia Enterprise en 2021. Se trata de un servicio pago para empresas que reutilizan el contenido de Wikimedia a gran escala, por medio de APIs⁵ para motores de búsqueda, LLM, aprendizaje profundo y más. El proyecto busca alcanzar un enfoque sostenible y que establezca una relación clara entre Wikimedia y el sector tecnológico. De este modo, los beneficios financieros obtenidos de grandes usuarios comerciales, se invertirán en el movimiento y sus comunidades (Wikimedia Meta-wiki

⁵ Una API (del inglés, *application programming interface*, en español, interfaz de programación de aplicaciones) es una pieza de código que permite a aplicaciones comunicarse entre sí, compartir información y funcionalidades.

(2024d)

Los intentos de utilizar LLM dentro de Wikipedia revelan asimismo desafíos significativos. Estos modelos han demostrado su capacidad para identificar problemas gramaticales y sintácticos. Además pueden sugerir cambios en el texto para mejorar la claridad o reescribir frases para garantizar la coherencia tonal. Los LLM también pueden ser utilizados para generar esquemas de artículos y texto en formato wiki en tareas que pueden resultar complejas para los usuarios (Thomas, 2023).

Sin embargo, encontramos un conjunto de riesgos. Una deficiencia proviene de la falta de verificación con hechos o fuentes de información que respalden el contenido. Además los LLM no pueden utilizar información en la que no han sido entrenados. Esto significa que los libros que no están disponibles en línea y la información en idiomas que no sean inglés, no forman parte del entrenamiento. Como resultado, pueden amplificar las desigualdades y los sesgos existentes (Deckelmann, 12 julio 2023).

Otro problema es la posible generación de investigación original. Los LLM, al solicitarles que generen texto sobre un tema específico, a menudo producen respuestas que carecen de fuentes confiables o resumen una variedad de fuentes para llegar a una conclusión novedosa que no ha sido publicada. Un problema relacionado es que estos modelos pueden producir información incorrecta o sesgada, y lo hacen con un lenguaje confiado y asertivo, lo que puede engañar a un usuario inadvertido. Este tipo de salidas son conocidas coloquialmente como "alucinaciones" (Thomas, 2023).

Finalmente, una tendencia que se incrementó en el último tiempo es el uso de LLM para entrenar otros LLM, en lugar de utilizar fuentes originales. Tales modelos pueden empeorar y olvidar información, fenómeno conocido como "colapso del modelo" (Deckelmann, 12 julio 2023).

El conjunto de estas consideraciones vuelven una y otra vez sobre la necesidad de contenido humano original, lo que hace que fuentes como Wikipedia adquieran una renovada relevancia en el ecosistema informacional. Mientras la comunidad wikipedista debate reglas y políticas para el uso de IA generativa en sus proyectos, desde la WMF se pronuncian en favor de la construcción humana del conocimiento (WMF, 2024).

Consideraciones finales

A lo largo de nuestro análisis, hemos definido a Wikipedia como una historia de colaboración, donde las aportaciones humanas y automáticas interactúan en un sistema sociotécnico. Este proyecto se distingue por su integración de bots en sus procesos de edición y corrección desde sus primeras etapas. Durante su segunda década, la implementación de

aprendizaje automático y herramientas de IA generativa ha fortalecido esta dinámica.

Wikimedia aspira a ser un ejemplo de prácticas éticas en el uso de la IA, promoviendo la transparencia y la participación humana, asegurando que puedan aprovecharse los beneficios de los avances tecnológicos a la vez que se garantice el control comunitario sobre los modelos. Ello requiere, a la vez, de un fortalecimiento de sus capacidades y su infraestructura, junto con un modo de gobernanza que posibilite tales objetivos. En este recorrido, se han establecido alianzas para el desarrollo y la implementación de tecnologías basadas en código abierto.

Nos centramos en el debate más reciente sobre las implicancias de la IA generativa sobre los proyectos wiki y su inserción en el ecosistema informacional. La extracción masiva de datos se ha convertido en una práctica común tanto en motores de búsqueda como en el entrenamiento de modelos de lenguaje a gran escala. Aunque el vasto acervo acumulado por Wikimedia adquiere una nueva relevancia en este contexto, es crucial explorar vías que impidan su uso parasitario o canibalizador, así como aquellas que puedan poner en riesgo la salud y la cohesión de sus comunidades.

Las innovaciones recientes posan nuevamente la mirada sobre Wikimedia y su modelo de producción colaborativa entre pares:

¿Podría una IA generativa reemplazar a Wikipedia? Podría intentarlo, pero el resultado sería un reemplazo que nadie realmente desea. No hay nada inevitable sobre la nueva tecnología. En cambio, depende de todos nosotros elegir lo que es más importante. Podemos priorizar la comprensión humana y la contribución del conocimiento al mundo de manera sostenible, equitativa y transparente, como un objetivo clave de los sistemas de IA generativa, no como una idea secundaria. (Deckelmann, 12 julio 2023, A vision for a trusted future)

A partir del fragmento previo, es posible concebir a Wikimedia en los términos delineados por la post-automatización. Esto es, un proceso de automatización reflexiva y centrada en los objetivos de un proyecto libre y abierto, sin finalidades comerciales, y no como un destino inexorable. Basándose en la dinámica comunitaria y la vasta experiencia acumulada durante sus más de 20 años de trayectoria, el proyecto se prepara para afrontar los retos de las nuevas oleadas tecnológicas.

Referencias

Aimar, L., Pagola, L. I., & Zanotti, A. (2021). Ediciones para el abordaje de sesgos en Wikipedia en español. Análisis de tres experiencias de edición colectiva y simultánea sobre la enciclopedia libre. *Virtualidad, Educación y Ciencia*, 12(22), 66–83. <https://revistas.unc.edu.ar/index.php/vesc/article/view/32118>

Bauwens, M. (2005). The political economy of peer production. *CTheory*, 12-1.

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. New York: Peter Lang.

Cebuano Wikipedia. (2024). Wikipedia. https://en.wikipedia.org/wiki/Cebuano_Wikipedia

Chierchie, L., & Fressoli, J. M. (2022). ¿Una automatización alternativa?: Prácticas con enfoques sociales y participativos. *Tableros+Bold*. 60 años de Diseño en la Facultad de Artes.

Clément, M. & Guitton, M. J. (2015). Interacting with bots online: Users' reactions to actions of automated programs in Wikipedia. *Computers in Human Behavior*. 50: 66–75.
<https://doi.org/10.1016/j.chb.2015.03.078>

Coleman, V. (9 enero 2108). "We keep the servers going ... and much more": Recent highlights from our Technology department. Wikimedia Foundation. Disponible en: <https://wikimediafoundation.org/news/2018/01/09/technology-department-highlights/>

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Deckelmann, S. (12 julio 2023). Wikipedia's value in the age of generative AI. By <https://wikimediafoundation.org/news/2023/07/12/wikipedias-value-in-the-age-of-generative-ai/>

Dyer-Whiteford, N. Kjøsen, A. and Steinhoff, J. (2019). *Inhuman Power: Artificial Intelligence and the Future of Capitalism*. London: Pluto Press

Fuchs, C. (2021). *Digital Capitalism: Media, Communication and Society Volume Three*. Routledge.

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., ... & Leahy, C. (2020). The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint*.
<https://doi.org/10.48550/arXiv.2101.00027>

Geiger, S. (2011). The Lives of Bots. In Lovink, G.; Tkacz, N. (eds.). *Critical Point of View: A Wikipedia Reader*. INC Reader. 7. Amsterdam: Institute of Network Cultures.
<https://arxiv.org/abs/1810.09590>

Gertner, J. (8 septiembre 2023). Wikipedia momento of truth.
<https://www.nytimes.com/2023/07/18/magazine/wikipedia-ai-chatgpt.html>

Giner, P. (13 junio 2023). MinT: Supporting underserved languages with open machine translation. Diff.
<https://diff.wikimedia.org/2023/06/13/mint-supporting-underserved-languages-with-open-machine-translation/>

Gómez-Zermeño, M. G. (2023). *Inteligencia Artificial Conceptos clave y tendencias para la innovación educativa*. Querétaro: Transdigital.

Guldbrandsson, L. (17 June 2013). Swedish Wikipedia surpasses 1 million articles with aid of

article creation bot.
<https://diff.wikimedia.org/2013/06/17/swedish-wikipedia-1-million-articles/>

Latour, B. (2008). Reensamblar lo social. *Una introducción a la teoría del actor-red*. Buenos Aires: Manantial.

Lund, A. & Zukerfeld, M. (2020). Corporate Capitalism's Use of Openness: Profit for Free? London: Palgrave Macmillan.

McMahon, C., Johnson, I., & Hecht, B. (2017). The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 142-151. <https://doi.org/10.1609/icwsm.v11i1.14883>

Niederer, S., & van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, 12(8), 1368-1387. <https://doi.org/10.1177/1461444810365297>

Nmvg. (2 diciembre 2022). The Paradox of Reuse, Language Models Edition. <https://nmvg.mataroa.blog/blog/the-paradox-of-reuse-language-models-edition/>

O'reilly, T. (2005). *What is web 2.0*. <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>

Pinchuk, M. (13 July 2023). Exploring paths for the future of free knowledge: New Wikipedia ChatGPT plugin, leveraging rich media social apps, and other experiments. <https://diff.wikimedia.org/2023/07/13/exploring-paths-for-the-future-of-free-knowledge-new-wikipedia-chatgpt-plugin-leveraging-rich-media-social-apps-and-other-experiments/>

Preston, J. (2022). *Artificial intelligence in the capitalist university: Academic labour, commodification, and value* (p. 182). Taylor & Francis.

Reagle, J. (2020). The Many (Reported) Deaths of Wikipedia. En *Wikipedia @ 20*. <https://wikipedia20.mitpress.mit.edu/pub/159g0pbd>

similarweb. (2024). wikipedia.org Top Traffic Sources. <https://www.similarweb.com/website/wikipedia.org/#traffic-sources>

Smith, A., & Fressoli, M. (2021). Post-automation. *Futures*, 132, 102778. <https://doi.org/10.1016/j.futures.2021.102778>

Tsvetkova, M., García-Gavilanes, R., Floridi, L., & Yasseri, T. (2017). Even good bots fight: The case of Wikipedia. *PloS one*, 12(2), e0171774.

Thomas, H. (2008). Estructuras cerradas vs. Procesos dinámicos: trayectorias y estilos de innovación y cambio tecnológico. In Thomas, H., & Buch, A. *Actos, actores y artefactos. Sociología de la tecnología*. Bernal: UNQ.

Thomas, P.A. (2023), "Wikipedia and large language models: perfect pairing or perfect storm?", *Library Hi Tech News*, Vol. 40 No. 10, pp. 6-8. <https://doi.org/10.1108/LHTN-03-2023-0056>

Van Dijck, J. (2019). *La cultura de la conectividad: una historia crítica de las redes sociales*. Siglo XXI editores.

Vincent, N., & Hecht, B. (2021). A deeper investigation of the importance of Wikipedia links to search engine results. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1-15. <https://dl.acm.org/doi/abs/10.1145/3449078>

WMF. (6 diciembre 2022). New Wikipedia editor features make it easy for everyone to contribute.

<https://wikimediafoundation.org/news/2022/12/06/new-wikipedia-editor-features-make-it-easy-for-everyone-to-contribute/>

WMF. (13 abril 2023). First grants announced from the Wikimedia Endowment to support technical innovation across Wikipedia and Wikimedia projects.

<https://wikimediafoundation.org/news/2023/04/13/first-grants-announced-from-the-wikimedia-endowment/>

WMF. (2024). Knowledge is human. Technology is important. Humanity is essential.

<https://wikimediafoundation.org/our-work/knowledge-is-human/>

Wikimedia Meta-wiki (2024a). List of Wikipedias by speakers per article.

https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_speakers_per_article

Wikimedia Meta-wiki (2024b). Machine Learning.

https://www.mediawiki.org/wiki/Machine_Learning/

Wikimedia Meta-wiki (2024c). Research:Wikistats metrics.

https://meta.wikimedia.org/wiki/Research:Wikistats_metrics#Common_Terms

Wikimedia Meta-wiki (2024d). Wikimedia Enterprise.

https://meta.wikimedia.org/wiki/Wikimedia_Enterprise

Wikimedia Statistics (2024). Wikimedia Statistics.

https://stats.wikimedia.org/#/all-wikipedia-projects/contributing/edits/normal|line|all|editor_type~anonymous*group-bot*name-bot*user|monthly

Wikipedia:Abstract Wikipedia. (2024). Wikipedia.

https://en.wikipedia.org/wiki/Abstract_Wikipedia

Wikipedia:Bots. (2024). Wikipedia. <https://en.wikipedia.org/wiki/Wikipedia:Bots>

Wikipedia:Bot policy. (2024). Wikipedia. https://en.wikipedia.org/wiki/Wikipedia:Bot_policy

Wikipedia:History of Wikipedia bots. (2024). Wikipedia.

https://en.wikipedia.org/wiki/Wikipedia:History_of_Wikipedia_bots

Wikiscan. (29 marzo 2024). Wikiscan. <https://wikiscan.org/>

Zanotti, A. (2022a). Rediseño de portales para la Wikipedia en español. *Revista Tecnología, Ciencia Y Educación*, (23), 155–176. <https://doi.org/10.51302/tce.2022.790>

Zanotti, A. (2022b). Wikipedia y la gobernanza de pares: Proyectos, comunidades, ecosistema. *Hipertextos*, 10(18), 056. <https://doi.org/10.24215/23143924e056>

Zukerfeld, M. (2020). Bits, plataformas y autómatas. Las tendencias del trabajo en el capitalismo informacional. *Revista Latinoamericana de Antropología del Trabajo*, 4(7). <https://ojs.ceil-conicet.gov.ar/index.php/lat/article/view/623>

Zukerfeld, M; Rabosto, A; Fredes, M. y De Marco, C. (2023). Encuesta de usos de ChatGPT en Argentina: Resultados preliminares sobre frecuencia de uso, productividad en el trabajo y sustitución de tareas. *Hipertextos*, 11 (20), e075. <https://doi.org/10.24215/23143924e075>

Fuentes de financiamiento

Se agradece al Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina por el financiamiento al proyecto *Bienes comunes informacionales: producción, uso y apropiación de conocimiento en plataformas abiertas colaborativas*, PIBAA CONICET, Res. 2022-1930-APN.

Conflicto de interés

El autor declara que no existen conflictos de interés relacionados con la presente investigación y sus resultados.

Este preprint fue presentado bajo las siguientes condiciones:

- Los autores declaran que son conscientes de que son los únicos responsables del contenido del preprint y que el depósito en SciELO Preprints no significa ningún compromiso por parte de SciELO, excepto su preservación y difusión.
- Los autores declaran que se obtuvieron los términos necesarios del consentimiento libre e informado de los participantes o pacientes en la investigación y se describen en el manuscrito, cuando corresponde.
- Los autores declaran que la preparación del manuscrito siguió las normas éticas de comunicación científica.
- Los autores declaran que los datos, las aplicaciones y otros contenidos subyacentes al manuscrito están referenciados.
- El manuscrito depositado está en formato PDF.
- Los autores declaran que la investigación que dio origen al manuscrito siguió buenas prácticas éticas y que las aprobaciones necesarias de los comités de ética de investigación, cuando corresponda, se describen en el manuscrito.
- Los autores declaran que una vez que un manuscrito es postado en el servidor SciELO Preprints, sólo puede ser retirado mediante solicitud a la Secretaría Editorial deSciELO Preprints, que publicará un aviso de retracción en su lugar.
- Los autores aceptan que el manuscrito aprobado esté disponible bajo licencia [Creative Commons CC-BY](#).
- El autor que presenta el manuscrito declara que las contribuciones de todos los autores y la declaración de conflicto de intereses se incluyen explícitamente y en secciones específicas del manuscrito.
- Los autores declaran que el manuscrito no fue depositado y/o previamente puesto a disposición en otro servidor de preprints o publicado en una revista.
- Si el manuscrito está siendo evaluado o siendo preparando para su publicación pero aún no ha sido publicado por una revista, los autores declaran que han recibido autorización de la revista para hacer este depósito.
- El autor que envía el manuscrito declara que todos los autores del mismo están de acuerdo con el envío a SciELO Preprints.