METODOLOGÍA ASISTIDA POR IA PARA LA CONFORMACIÓN DE UNA BASE DE DATOS DE UN REPOSITORIO DOCUMENTAL: LAS PUBLICACIONES DE LA CASA DE CHILE EN MÉXICO, 1974-1993

AI-ASSISTED METHODOLOGY FOR CREATING A DATABASE OF A DOCUMENT REPOSITORY: SERIAL PUBLICATIONS OF LA CASA DE CHILE IN MEXICO, 1974-1993

Claudia Fedora Rojas-Mira
https://orcid.org/0000-0002-1673-7504
Universidad Tecnológica Metropolitana, Chile.
E-mail: c.rojasm@utem.cl

José Antonio Román-Brugnoli
https://orcid.org/0000-0001-8151-9821
Universidad Tecnológica Metropolitana, Chile.
E-mail: j.romanb@utem.cl

Pablo Seguel-Gutiérrez
https://orcid.org/0000-0001-6837-287X
Universidad Alberto Hurtado, Chile.
E-mail: pseguel@uahurtado.cl

Emilio J. Castro-Navarro
https://orcid.org/0000-0003-4931-6180
Universidad Tecnológica Metropolitana, Chile.
E-mail: ecastron@utem.cl

DOI: https://doi.org/10.36132/g0hxam50

Recibido: 4 diciembre 2025 / Aceptado: 11 octubre 2026 / Publicado: 26 octubre 2026

Resumen: Este artículo presenta criterios metodológicos para la conformación de un repositorio con apoyo de inteligencia artificial generativa (IA) del proyecto *Exilio, solidaridad y resistencia política en La Casa de Chile en México: análisis de sus boletines informativos 1973-1993*. Se describen procesos de digitalización, cribado humano, normalización y clasificación de un acervo, así como las herramientas de IA para el reconocimiento óptico de caracteres, la generación de metadatos y la organización temática. La reflexión enfatiza la IA como complemento del investigador, con base en criterios historiográficos y archivísticos, destacando avances, limitaciones y aprendizajes transferibles a repositorios sobre memoria y derechos humanos.

Palabras clave: repositorio documental, inteligencia artificial, archivística, exilio, solidaridad

Abstract: His article presents methodological criteria for the creation of a document repository supported by generative artificial intelligence (AI) for the project *Exile, Solidarity, and Political Resistance at La Casa de Chile in Mexico: Analysis of its Newsletters, 1973-1993*. It describes processes of digitization, human review, standardization, and classification of heterogeneous collection, as well as the AI tools for optical character recognition, metadata generation, and thematic organization. The reflection emphasizes AI as a complement to the researcher, based on historiographical and archival criteria, highlighting advances, limitations, and lessons transferable to the creation of other repositories on memory and human rights.

Keywords: documentary repository, artificial intelligence, archiving, exile, solidarity

* Este artículo es resultado del proyecto “Exilio, solidaridad y resistencia política en La Casa de Chile en México: análisis de sus boletines informativos 1973–1993”, financiado por el Fondo de Fomento a la I+D+i o Creación de la Vicerrectoría de Investigación y Postgrados de la Universidad Tecnológica Metropolitana (LPR23-26).

INTRODUCCIÓN Y JUSTIFICACIÓN

El exilio chileno posterior al golpe de Estado de 1973 constituye un fenómeno significativo de la historia reciente del país y del espacio Iberoatlántico[1]. La violencia política, la persecución y el cierre del sistema democrático expulsaron a decenas de miles de personas, que conformaron comunidades exiliares en más de ochenta países[2]. México se convirtió en un destino prioritario por su tradición de asilo y por la activa solidaridad desplegada por autoridades, intelectuales y organizaciones sociales[3]. En este marco, la fundación de La Casa de Chile en México en 1974 configuró una plataforma política y cultural para cohesionar a la comunidad exiliada, articular redes de solidaridad internacional y sostener la denuncia de las violaciones de derechos humanos cometidas por la dictadura civil-militar encabezada por Augusto Pinochet[4].

La Casa de Chile en México desarrolló, durante casi dos décadas, una intensa producción editorial que incluyó los *Boletines Informativos*, los *Cuadernos Casa de Chile*, la *Presencia Casa de Chile en México* e *Informativo Casa de Chile*. Si bien la investigación desde la historia y las ciencias sociales ha abordado a la institución como caso de estudio[5], estas no han considerado la producción editorial de la institución como objeto de indagación.

Esta limitante se debe a lo reciente de la recuperación de las publicaciones de la Casa de Chile en México, que fueron donadas al Museo de la Memoria y los Derechos Humanos (MMDH) de Chile en 2023 y archivadas como series documentales dentro de dos fondos según la procedencia de la donación: "UAM Xochimilco. Casa de Chile en México" y "Fernando Schultz". Este criterio, centrado en el origen de la donación y no en la institución productora, fragmenta el conocimiento del conjunto documental y dificulta su análisis como archivo orgánico.

El proyecto *Exilio, solidaridad y resistencia política en La Casa de Chile en México: análisis de sus boletines informativos 1973-1993* busca subsanar esa dispersión. Su primer objetivo ha sido reunir las publicaciones seriadas de la Casa de Chile y organizar una base de datos orientada a la construcción de un repositorio que responda a criterios historiográficos[6]. El repositorio apunta a garantizar acceso público al patrimonio documental del exilio chileno en México, en línea con experiencias de preservación de archivos de derechos humanos, y a facilitar la cooperación internacional en torno a una memoria

[1] Silvina Jensen, "«¿Quién es el dueño de la historia?». Agenda para una historia conectada de la represión y los exilios en el espacio iberoatlántico en la segunda mitad del siglo XX", en *Cartografías de la represión en América Latina: historias comparadas y conectadas de la violencia estatal (1960–1990)*, ed. Gabriela Águila, Santiago Garaño y Pablo Scatizza (Buenos Aires: Imago Mundi, 2025), 95–114.

[2] Claudia Rojas Mira, *Las Moradas del Exilio: La Casa de Chile en México (1973-1993)* (Santiago de Chile: Ediciones Universidad Tecnológica Metropolitana, 2023).

[3] Raquel Rebolledo-Rebolledo y Gabriel Reyes-Arriagada, "El exilio chileno en México. Organización y denuncia como resistencia al desarraigo", *CUHSO* 33, nº 2 (2023): 134–149, https://doi.org/10.7770/cuhso-v33n2-art647.

[4] Nancy Nicholls, "Defensa de DDHH en Chile en el contexto transnacional del movimiento de defensa de los derechos humanos, 1973–1990", *Estudos Ibero-Americanos* 45, nº 1 (2019): 43–56, https://doi.org/10.15448/1980-864X.2019.1.31796; Pablo Seguel, *Soldados de la represión. Anticomunismo, contrasubversión y seguridad nacional en las fuerzas armadas chilenas, 1970–1975* (Santiago de Chile: Ediciones Universidad Alberto Hurtado, 2022).

[5] Rojas Mira, *Las Moradas del Exilio*.

[6] Consejo Internacional de Archivos, *Principios básicos sobre el papel de archiveros y gestores de documentos en la defensa de los derechos humanos* (septiembre de 2016), https://www.ica.org/resource/basic-principles-on-the-role-of-archivists-and-records-managers-in-support-of-human-rights/.

transnacional cuyas huellas están repartidas en distintos países[7].

En Chile, iniciativas recientes articulan inteligencia artificial, archivos y memoria de la dictadura. Destaca *NuestramemorIA*, proyecto que combina reconocimiento óptico de caracteres, extracción de entidades, grafos de conocimiento y validación humana para relacionar documentos dispersos y preservar su trazabilidad[8]. Asimismo, el estudio del caso Quiñones aplicó modelos de lenguaje y GraphRAG a 13283 páginas judiciales sobre una desaparición forzada de 1976[9]. Aunque comparte el propósito de organizar y consultar grandes corpus mediante IA, nuestra investigación amplía ese enfoque hacia las publicaciones político-culturales del exilio chileno en México, vinculadas con denuncia, solidaridad, resistencia política y construcción transnacional de la memoria histórica.

El carácter inédito del archivo, el volumen y la heterogeneidad de sus materiales, plantean desafíos que los métodos tradicionales de procesamiento documental difícilmente podrían resolver en tiempos humanos razonables. La integración de herramientas de inteligencia artificial (IA) generativa permite acelerar y abaratar costos en tareas de investigación que serían más lentas y costosas bajo métodos manuales. También la heterogeneidad de las fuentes (que combina boletines, cuadernos y registros gráficos) añade una complejidad que demanda soluciones tecnológicas capaces de superar los límites de las prácticas archivísticas convencionales. Es en este punto donde la IA adquiere un rol central.

Estudios recientes muestran que modelos grandes de lenguaje (LLMs, por sus siglas en inglés) pueden ayudar en la clasificación y análisis de grandes textos. Trabajos recientes comprobaron que combinar IA generativa con revisión humana aumenta la precisión y la coherencia en anotaciones complejas[10], mientras otras investigaciones encontraron que estos modelos detectan patrones semánticos sin necesidad de entrenamiento previo[11]. En archivística, la inteligencia artificial potencia el procesamiento documental, aunque suscita desafíos metodológicos y éticos[12]. Estos hallazgos apoyan el uso de IA generativa en esta investigación, que complementa al trabajo experto sin reemplazarlo.

Diversas técnicas y herramientas de IA permiten optimizar procesos como el reconocimiento óptico de caracteres (OCR), la normalización de textos, la extracción automática de metadatos y la clasificación semántica de documentos[13]. Todo ello contribuye a

[7] Kim Christiaens, Idesbald Goddeeris y Magaly Rodríguez, eds., *European Solidarity with Chile, 1970s–1980s* (Frankfurt am Main: Peter Lang, 2014).

[8] NuestramemorIA, "El Proyecto", *NuestramemorIA*, consultado el 7 de octubre de 2026, https://nuestramemoria.ing.uc.cl/el-proyecto/.

[9] Hugo Rojas Corral, Antonia Fonck, Darwin Sanhueza, Domingo Mery y Jocelyn Dunstan, "Artificial Intelligence, Judicial Archives, and Transitional Justice: Analysis of the Quiñones Case in Chile", *Tapuya: Latin American Science, Technology and Society* 9, nº 1 (2026), https://doi.org/10.1080/25729861.2026.2698211.

[10] Yuting Guo, Anthony Ovadje, Mohammed Ali Al-Garadi y Abeed Sarker, "Evaluating Large Language Models for Health-Related Text Classification Tasks with Public Social Media Data", *Journal of the American Medical Informatics Association* 31, nº 10 (2024): 2181–2189, https://doi.org/10.1093/jamia/ocae210.

[11] Tharindu Kumarage, Amrita Bhattacharjee y Joshua Garland, "Harnessing Artificial Intelligence to Combat Online Hate: Exploring the Challenges and Opportunities of Large Language Models in Hate Speech Detection", prepublicación en arXiv (2024), https://doi.org/10.48550/arXiv.2403.08035.

[12] Giovanni Colavizza, Tobias Blanke, Charles Jeurgens y Julia Noordegraaf, "Archives and AI: An Overview of Current Debates and Future Perspectives", *Journal on Computing and Cultural Heritage* 15, nº 1 (2022): artículo 4, 1–15, https://doi.org/10.1145/3479010.

[13] Melanie Mitchell, *Inteligencia artificial: guía para seres pensantes* (Madrid: Capitán Swing, 2024),

acelerar la organización del acervo y a enriquecer su análisis, sin sustituir los criterios historiográficos ni la necesaria verificación humana experta del proceso[14].

Este artículo tiene como objetivo presentar la metodología desarrollada en la experiencia de conformar la base de datos del repositorio documental del proyecto *Exilio, solidaridad y resistencia política en La Casa de Chile en México.* Su aporte se articula en tres niveles. En primer lugar, propone un modelo metodológico replicable, que combina estándares archivísticos con herramientas de IA para la conformación de una base de datos, ofreciendo aprendizajes transferibles a otros repositorios de memoria y derechos humanos en historia reciente[15]. En segundo lugar, plantea un enfoque historiográfico que organiza la base de datos del archivo en función de su relevancia política y cultural y no únicamente bajo parámetros técnicos de descripción. Finalmente, se trata de un corpus inédito cuya sistematización abre nuevas posibilidades de investigación sobre las prácticas de solidaridad y resistencia política en el exilio chileno, contribuyendo a ampliar un campo historiográfico en consolidación[16].

1. MARCO CONCEPTUAL Y METODOLÓGICO

La literatura sobre exilio y memoria ha destacado la importancia de los archivos como soportes de prácticas colectivas de resistencia y de construcción identitaria[17]. En este marco, la noción de repositorio documental se interpreta en un sentido amplio y dinámico. No se trata simplemente de un contenedor de documentos digitalizados, sino de un sistema que integra criterios de clasificación, preservación y acceso, con miras a su análisis histórico, documental y de memoria social.

Siguiendo la definición de la Fundación Española para la Ciencia y la Tecnología (FECYT), se entiende por repositorio a

> "(…) un conjunto de servicios que ofrece una institución a los miembros de su comunidad para la gestión y distribución de materiales digitales creados por la institución y sus miembros, organizados de tal modo que se garantice la correcta administración de los mismos, in-

59–84.

[14] Christian Henriot, *The AI-Augmented Research Process: A Historian's Perspective*, prepublicación en arXiv (2025), https://doi.org/10.48550/arXiv.2508.01779; Sarah Oberbichler y Cindarella Petz, *Working Paper: Implementing Generative AI in the Historical Studies* (Leibniz Institute of European History, 2025), https://doi.org/10.5281/zenodo.14924737.

[15] Oleksandr Bondarenko, Svitlana Orlyk y Mark Pyzyk, "Digital Technology in Historical Research: Contemporary Scholarly Current", *Skhidnoievropeiskyi Istorychnyi Visnyk [East European Historical Bulletin]* 33 (2024): 227–238, https://doi.org/10.24919/2519-058X.33.317459; Gerben Zaagsma, "Facing the History Machine: Toward Histories of Digital History", *History of Humanities* 9, nº 2 (2024): 451–491, https://doi.org/10.1086/731827.

[16] Rojas Mira, *Las Moradas del Exilio*.

[17] Patricia Baeza Duffy, "Construcción de memorias que compiten sobre el pasado chileno reciente por adultos que vivieron ese período en Chile o en el exilio", *Discurso & Sociedad* 11, nº 3 (2017): 433–457, https://dialnet.unirioja.es/servlet/articulo?codigo=6140840; Loreto Rebolledo, *Memorias del desarraigo: testimonios del exilio y retorno de hombres y mujeres de Chile* (Santiago de Chile: Catalonia, 2005); Rebolledo-Rebolledo y Reyes-Arriagada, "El exilio chileno en México"; Luis Roniger, "El giro transnacional en estudios latinoamericanos. Una hoja de ruta", *Revista Mexicana de Ciencias Políticas y Sociales* 70, nº 254 (2025): 17–46, https://doi.org/10.22201/fcpys.2448492xe.2025.254.91325; Luis Roniger, "Exile, Transnational Life, Return, and Diasporas: The Southern Cone Experience", *Middle Atlantic Review of Latin American Studies* 7, nº 1 (2023): 1–29, https://doi.org/10.23870/marlas.431.

cluyendo la preservación a largo plazo, el acceso y su difusión"[18].

En ese sentido, los repositorios pueden constituirse en institucionales, temáticos y de datos.

En el proyecto *Exilio, solidaridad y resistencia*, la construcción de la base de datos implicó crear un repositorio temático de las publicaciones seriadas de La Casa de Chile en México. A través de este proceso, la documentación antes dispersa se transformó en un corpus estructurado mediante campos normalizados como código; año; título; autor; resumen; palabras clave y tema, lo que no solo garantiza trazabilidad y coherencia, sino que habilita múltiples niveles de acceso y lectura.

A diferencia de un archivo digital sin un marco de clasificación historiográfica, el repositorio temático integra metodologías propias de la archivística y de las ciencias sociales, permitiendo que el acervo no quede reducido a su valor patrimonial o de conservación, sino que se proyecte como un insumo para la investigación, la docencia y la memoria pública. En este proceso, la incorporación de IA resulta decisiva, tanto por la relevancia que cobra en su articulación con internet y aplicaciones de diverso tipo para la generación y gestión de un conocimiento científico compartido, como para el abordaje en investigación de grandes volúmenes de documentación histórica heterogénea.

Un repositorio bien conformado busca dotar de coherencia a acervos dispersos, establecer reglas de trazabilidad y garantizar la sostenibilidad de los datos. En este sentido, se articula con los principios archivísticos de resguardo, normalización y accesibilidad, pero incorpora también un enfoque historiográfico que lo convierte en un dispositivo de memoria activa. Su énfasis es su organización significativa de cara a la investigación histórica y a la difusión pública[19].

De este modo, la base de datos de este proyecto no solo resguarda y preserva, sino que constituye una infraestructura de conocimiento organizada en torno a un tema específico, capaz de facilitar búsquedas avanzadas, comparaciones diacrónicas y nuevas narrativas históricas, problematizando el fenómeno del exilio, la solidaridad y la resistencia, al tiempo que democratiza el acceso a materiales antes fragmentados e invisibilizados por su volumen y contenido. El repositorio temático resultante transforma un conjunto documental hasta ahora inexplorado en un corpus coherente, accesible y analizable, cuya relevancia excede lo técnico y se inscribe en los debates contemporáneos sobre exilio, memoria, derechos humanos y las posibilidades que ofrecen las humanidades digitales[20].

En los últimos años, diversas experiencias han demostrado el potencial de la IA en archivos, bibliotecas y humanidades digitales. Por una parte, se ha documentado su aplicación en procesos de OCR, extracción de metadatos y clasificación automática de documentos[21]. Por otra, se ha explorado su capacidad para analizar narrativas y discursos históricos, ofreciendo nuevas perspectivas sobre la relación entre algoritmos y memoria

[18] José Manuel Barrueco Cruz, Pilar Rico Castro y Laura Valeria Bonora Eve, *Guía para la evaluación de repositorios institucionales de investigación* (Madrid: FECYT/REBIUN, 2021), 37, http://hdl.handle.net/20.500.11967/809.

[19] Museo de la Memoria y los Derechos Humanos, *Políticas de digitalización de colecciones* (Santiago de Chile: Área Colecciones, mayo de 2012).

[20] Leonor Riesco Tagle, ed., *Humanidades digitales: miradas, experiencias y proyecciones desde América Latina* (Santiago de Chile: Ediciones Universidad San Sebastián, 2025).

[21] Salvatore Spina, "Artificial Intelligence in Archival and Historical Scholarship Workflow: HTR and ChatGPT", *Umanistica Digitale*, nº 16 (2023): 125–140, https://doi.org/10.6092/issn.2532-8816/17205; Martin Horna Romero y Roxana Perca Chagua, "Quinquenio de la aplicación de la inteligencia artificial en la archivística: una revisión sistemática en revistas académicas", *BiD* 54, nº 2 (2025), https://doi.org/10.1344/bid2025.54.02.

social[22]. En el campo de la historiografía, se ha destacado que estas tecnologías no sustituyen la labor interpretativa del historiador, sino que amplían las posibilidades de lectura y cruce de datos en acervos complejos[23].

Kumarage, Bhattacharjee y Garland indican que los modelos de lenguaje agrupan documentos según su significado, incluso en áreas no exploradas[24]. Esto no solo acelera el procesamiento, sino que, también, amplía las interpretaciones al captar relaciones complejas, lo cual es clave para organizar un corpus heterogéneo. Por otra parte, Guo, Ovadje, Al-Garadi y Sarker resaltan la importancia del diseño híbrido con doble revisión humana, presente en esta metodología, para asegurar la calidad y la trazabilidad[25]. Al mismo tiempo, estas innovaciones plantean desafíos científicos y dilemas éticos que no pueden soslayarse: sesgos algorítmicos, preservación de datos sensibles y la necesidad de transparencia en los procesos de estructuración de los comandos (*prompt*) hasta el resultado final del análisis[26].

Este enfoque metodológico reconoce que la IA actúa como un complemento a los criterios archivísticos e historiográficos. Aporta rapidez y escalabilidad frente a acervos voluminosos y heterogéneos, pero siempre requiere la mediación del juicio humano experto. La conformación del repositorio, entonces, no es un mero procedimiento técnico, sino que una práctica de construcción de memoria que conjuga preservación, accesibilidad y sentido histórico. En el caso del proyecto *Exilio, solidaridad y resistencia* se trata de dotar a la documentación de la Casa de Chile en México de una estructura coherente y abierta, que permita tanto a investigadores como a comunidades acceder a una memoria hasta ahora dispersa y escasamente conocida, potenciando su dimensión transnacional y colaborativa.

2. DIAGNÓSTICO DE LA SITUACIÓN ARCHIVÍSTICA

2.1. La situación de los archivos de La Casa de Chile en México

Al momento de inicio de este proyecto, las publicaciones seriadas de La Casa de Chile en México (1974-1993) se encontraban distribuidas en dos fondos en el MMDH atendiendo al origen de la donación de la siguiente manera: "UAM Xochimilco. Casa de Chile en México" (donde encontramos la colección *Noticias de Chile. Boletín Casa de Chile en México, Cuadernos Casa de Chile en México*) y "Fernando Schultz" (donde encontramos *Presencia de Casa de Chile en México* y la colección *Cuadernos Casa de Chile*). Estos documentos no se encontraban digitalizados, por lo que la primera acción del proyecto fue establecer una colaboración con el MMDH para digitalizar cada documento de cada una de estas publicaciones seriadas.

En esta primera fase de identificación de publicaciones seriadas de la Casa de Chile en México en el inventario del MMDH para

[22] Geovanny Francisco Ruiz Muñoz, Soledad Yessenia Ramírez Mantilla, Luisa Joselyn Ortiz Villón y Juanita Irene Guevara Burgos, "Algoritmos de la memoria: aplicaciones de la inteligencia artificial en la reinterpretación histórica y el análisis de narrativas sociales", *Reincisol* 4, nº 7 (2025): 3127–3155, https://doi.org/10.59282/reincisol.V4(7)3127-3155.

[23] Henriot, *The AI-Augmented Research Process*; Laura Manrique-Gómez y Jaime Humberto Borja Gómez, "Ciencia de datos para la historia: datificar las fuentes para una historia (predictiva)", *Historia y Grafía*, nº 64 (2025): 97–145, https://doi.org/10.48102/hyg.vi64.541.

[24] Kumarage, Bhattacharjee y Garland, "Harnessing Artificial Intelligence to Combat Online Hate".

[25] Guo et al., "Evaluating Large Language Models".

[26] David B. Resnik y Mohammad Hosseini, "The Ethics of Using Artificial Intelligence in Scientific Research: New Guidance Needed for a New Tool", *AI and Ethics* 5, nº 2 (2025): 1499–1521, https://doi.org/10.1007/s43681-024-00493-8; Marnie Hughes-Warrington, "Ethics for Artificial Historians", *History and Theory* 64, nº 2 (2025): 159–177, https://doi.org/10.1111/hith.12377.

su reunión y digitalización, se identificó una brecha en el ámbito archivístico para la creación de un repositorio temático como el propuesto: los criterios de archivo del MMDH no eran coincidentes con referentes internacionales de criterios de clasificación archivística, como los del Consejo Internacional de Archivos (ICA)[27]. Particularmente, los criterios de clasificación y estructuración del MMDH habían archivado las publicaciones seriadas de la Casa de Chile en México según la fuente de la donación. Esto genera problemas archivísticos, de eventuales duplicidades de un documento en fondos diferentes; dificultando la consulta y el análisis del archivo según criterios de interés historiográficos, sensibles al contexto, la temporalidad y la fuente productora, entre otros. De esta forma, un primer desafío del proyecto *Exilio, solidaridad y resistencia* fue generar una base de datos de un repositorio documental unificado que reuniera y sistematizara su acervo editorial, siguiendo normativas internacionales de descripción archivística como las promovidas por el ICA[28].

El MMDH comprende dos cuerpos de política de archivo y digitalización: de 2012 y 2021. Las políticas de 2012 adaptan criterios internacionales y constituyen un marco general orientado a garantizar la preservación patrimonial y la coherencia con estándares internacionales, mientras que la de 2021 adopta un enfoque más operativo, centrado en la estandarización de procedimientos y la optimización de recursos.

Los criterios de almacenamiento y preservación son un punto de continuidad. Tanto en 2012 como en 2021 se estipula la generación de másteres en alta resolución, que se resguardan en servidores internos, separados de las copias de difusión. La diferencia radica en que, mientras las políticas de 2012 disponen un almacenamiento en servidores NAS (*Network Attached Storage*, dispositivo de almacenamiento centralizado al que se accede a través de la red) bajo control permanente de calidad; el instructivo de 2021 establece un sistema de organización de carpetas vinculadas a la ID asignada al documento, lo que asegura trazabilidad y facilita la administración de grandes volúmenes.

En conjunto, estas normativas evidencian un tránsito desde un marco amplio, patrimonialista y normativo (2012) hacia un enfoque operativo y técnico (2021), que estandariza procedimientos de digitalización en función de la trazabilidad, la eficiencia y el control de formatos. Esta evolución permite comprender cómo el MMDH ha ajustado sus políticas, mantiene fidelidad al principio rector de preservar la integridad de los documentos y asegurar acceso público.

Por su parte, de acuerdo con los criterios de clasificación archivística de ICA, la descripción documental se fundamenta en el respeto al principio de procedencia, que obliga a mantener la estructura y el contexto original de los documentos, evitando su mezcla o alteración, y en la aplicación de descripción multinivel, que organiza la información de general a particular, desde el fondo hasta el documento simple, consignando en cada nivel solo datos pertinentes. La descripción debe contextualizar origen, funciones e institución o persona productora, ya que el valor de la documentación depende de su marco de creación. Asimismo, requiere la normalización de elementos descriptivos, siguiendo la Norma Internacional General de Descripción Archivística, conocida por su acrónimo ISAD(G)[29] por su denominación en inglés, y que considera la especificidad de los archi-

[27] Consejo Internacional de Archivos, *ISAD(G): Norma Internacional General de Descripción Archivística*, 2.ª ed. (Madrid: Ministerio de Educación, Cultura y Deporte, 2000).

[28] Consejo Internacional de Archivos, "La Declaración Universal sobre los Archivos", *Consejo Internacional de Archivos*, https://www.ica.org/es/resource/declaracion-universal-de-los-archivos-uda/.

[29] Consejo Internacional de Archivos, *ISAD(G): Norma Internacional General de Descripción Archivística*, https://www.ica.org/es/resource/isadg-general-international-standard-archival-description-second-edition/.

vos de derechos humanos[30]. Estas establecen áreas de identificación, contexto, contenido, condiciones de acceso, documentación asociada y control de la descripción. Todo ello debe expresarse con claridad, precisión y objetividad, sin juicios de valor, garantizando consistencia y coherencia en uso de formatos y terminología. En definitiva, la finalidad central es asegurar accesibilidad y recuperación de información, facilitando el acceso a usuarios tanto en entornos físicos como digitales.

Mientras que el ICA remarca contextualización y descripción multinivel como mecanismos indispensables para comprender el valor historiográfico de la documentación, las políticas del MMDH privilegian criterios técnicos de preservación y acceso digital. Ello significa que la digitalización, aunque asegura la conservación y difusión de los documentos, no necesariamente garantiza la reconstrucción del marco de creación y densidad histórica de los fondos. En consecuencia, sin un trabajo adicional de contextualización historiográfica, los objetos digitales tienden a reducirse a unidades técnicas de almacenamiento, perdiendo las relaciones internas y externas que explican su sentido histórico.

Se propuso crear un *dataset* con criterios historiográficos que cumple con estándares técnicos de digitalización y dimensiones propias de la descripción archivística internacional: procedencia, contexto de producción, función de la institución y la relación entre series y subseries. Esto implicó adaptar normas de ISAD(G) al entorno digital, traduciendo categorías tradicionales de identificación, contexto y contenido en campos estructurados y normalizados que permiten la búsqueda, el análisis y la reutilización de los datos. En otras palabras, la digitalización se entendió no solo como un procedimiento técnico, sino también como un proceso historiográfico de reconstrucción y organización de la memoria documental.

De este modo, la brecha se transformó en un espacio de innovación metodológica: entre las normas internacionales de descripción archivística, que ponen el acento en la procedencia y el contexto, y las políticas nacionales de digitalización, orientadas a la preservación técnica, surge la necesidad de *datasets* historiográficos que integren ambas dimensiones. Solo así es posible garantizar que documentos digitalizados no pierdan su densidad histórica y los repositorios en línea cumplan una doble función: la de conservar técnicamente archivos y ponerlos en valor como fuentes vivas para la investigación y la memoria histórica.

3. LA METODOLOGÍA DESARROLLADA PARA LA CONFORMACIÓN DEL REPOSITORIO DOCUMENTAL ASISTIDA POR IA

En la actualidad, la construcción de repositorios documentales digitales constituye un desafío relevante para las ciencias sociales y humanas, tanto por la centralidad que han adquirido los datos en los procesos de investigación como por la necesidad de preservar fuentes que permitan comprender fenómenos históricos, culturales y sociales de muy antigua data. La creciente disponibilidad de grandes volúmenes de documentos, producto de procesos masivos de digitalización, la producción acelerada de información nativa digital y la apertura institucional de archivos antes restringidos, ha ampliado significativamente el universo de materiales que los investigadores deben gestionar. A ello se suma la dispersión geográfica de los acervos, la fragilidad de los soportes tradicionales y la heterogeneidad de los formatos, lo que complejiza su preservación, descripción y accesibilidad.

En este escenario, la conformación de repositorios coherentes, sostenibles y técnicamente estandarizados exige metodologías robustas capaces de operar en múltiples escalas, garantizando la preservación a largo

[30] Consejo Internacional de Archivos, "Aplicación de la Norma ISAD(G) para el tratamiento de Archivos y Derechos Humanos", *Consejo Internacional de Archivos*, https://www.ica.org/es/resource/aplicacion-de-la-norma-isadg-para-el-tratamiento-de-archivos-y-derechos-humanos/.

plazo, la posibilidad de reutilización y de acceso abierto para diversas comunidades de investigación y actores sociales. En este marco, el proyecto *Exilio, solidaridad y resistencia* buscó conformar un repositorio que permitiera organizar, resguardar y difundir un corpus documental disperso y escasamente trabajado por la historiografía.

En el caso del proyecto *Exilio, solidaridad y resistencia en la Casa de Chile en México*, el repositorio no solo cumple una función de preservación documental: también configura un espacio de investigación y memoria que transforma un corpus previamente disperso en un recurso organizado historiográficamente, accesible y sostenible. Esta arquitectura metodológica permite integrar materiales de procedencias diversas, establecer relaciones interpretativas entre ellos y plantear nuevas preguntas de investigación. Al mismo tiempo, ofrece un modelo replicable para otros acervos vinculados al exilio y a las redes de solidaridad transnacional, contribuyendo al desarrollo de las humanidades digitales y fortaleciendo la democratización del acceso a la memoria histórica.

En este sentido, el repositorio opera como una infraestructura activa de investigación y memoria, capaz de facilitar procesos de análisis comparativo, identificar patrones en grandes volúmenes de datos y sostener narrativas históricas más complejas y plurales. Al integrar la automatización con la supervisión experta, el repositorio se convierte en un espacio que no solo resguarda documentos, también habilita nuevas formas de exploración, reutilización y participación, reforzando su valor como plataforma abierta para investigadores, comunidades y proyectos de memoria.

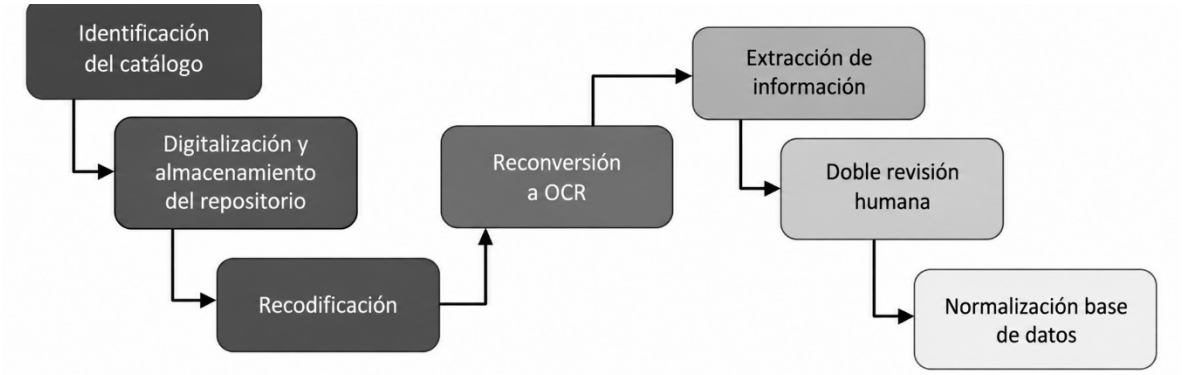
El *Proyecto Exilio, Solidaridad y Resistencia* desarrolló una metodología que combinó un enfoque de descripción archivística, las políticas de digitalización del MMDH y un criterio historiográfico para conformar un dataset temático que permitió un abordaje descriptivo específico de la información existente en el repositorio documental del proyecto.

Este artículo presenta, en clave metodológica, el proceso de conformación de la base de datos del repositorio, estructurando siete fases: (1) reconocimiento de los materiales dentro de la estructura de catalogación del MMDH; (2) digitalización y almacenamiento en el repositorio del proyecto; (3) recodificación; (4) reconocimiento de OCR; (5) extracción de información; (6) doble revisión humana; y (7) normalización. Cada fase se explica en detalle, situándose en la tradición archivística y en el marco de experiencias comparadas en archivos, bibliotecas y humanidades digitales.

3.1. Fases de la metodología de construcción del repositorio documental

La metodología para conformar la base de datos del repositorio comprende siete fases sucesivas que aseguran preservación, calidad y consistencia (Gráfico 1).

Gráfico 1. Diagrama de flujo de las fases metodológicas del proyecto



Fuente: Elaboración propia

1. Reconocimiento de los materiales dentro de la estructura de catalogación del MMDH

La fase de identificación de los documentos del catálogo tiene como propósito demarcar los documentos originales y establecer mecanismos de trazabilidad que aseguren su identificación a lo largo de todo el proceso. En la práctica archivística, la catalogación constituye el primer eslabón para transformar un conjunto de materiales dispersos en una colección coherente y corresponde a un proceso llevado a cabo por la institución que resguarda la documentación.

Para efectos de este proyecto se buscaron las publicaciones seriadas producidos por la Casa de Chile en México, disponibles en dos fondos del MMDH que abarcaron las siguientes publicaciones (tabla 1)

Tabla 1. Procedencia de los documentos en el CEDOC MMDH

Fondo	Colección	Ítems	Descripción
Fernando Schultz (ID: 1596)	Colección Publicaciones Seriadas (ID00001596000002)	63-110	Corresponde a la publicación seriada Informativo Casa de Chile en México
		216-233	Corresponde a la Publicación seriada Presencia de Casa de Chile en México
	Documentos Casa de Chile (ID00001596000040)		Corresponde a la publicación monográfica numerada Cuadernos Casa de Chile
UAM Xochimilco (ID 1595)	Publicaciones Seriadas (ID 00001595000003)	Carpetas 1 – 14 en las Cajas B7 E 28 C6 y 7	Carpetas que contienen la publicación seriada Noticias de Chile. Boletín de Casa de Chile en México
		Caja B7: E29-C7.	Carpetas que contienen la publicación seriada Cuadernos Casa de Chile en México

Fuente: Elaboración propia.

2. Digitalización y almacenamiento en el repositorio de proyecto

La segunda fase consiste en la transformación de los documentos físicos en imágenes digitales de calidad, aptas para ser procesadas posteriormente mediante OCR. El objetivo es garantizar que cada detalle relevante del documento quede registrado en la imagen digital.

La digitalización exige decisiones técnicas que condicionan la preservación y legibilidad documental. Por ello, requiere personal con conocimiento temático y sensibilidad para tratar materiales delicados, así como criterios de categorización definidos explícitamente por el equipo investigador para orientar su procesamiento.

En la fase de digitalización y almacenamiento, se resguardan los originales con criterios de trazabilidad. Esta fase cumple dos funciones fundamentales: garantizar la preservación de largo plazo mediante imágenes maestras de máxima calidad, y, democratizar el acceso a través de copias derivadas más livianas que pueden circular en línea o en bases de datos abiertas.

En esta fase se convirtieron los soportes físicos en imágenes de alta resolución y formatos maestros aptos para los procesos posteriores. La base resultante mantuvo la estructuración en que los documentos se encontraban en el inventario del MMDH y una copia fue entregada como aporte al Museo.

Para efectos de este proyecto, se decidió usar el formato PDF y aplicar el estándar archivístico del MMDH de tener para cada documento dos copias de diferente calidad: una de alta resolución de 600 dpi (High Quality o HQ), y otra de baja resolución en 300 dpi (Low Quality o LQ). El umbral de 300 dpi constituye el mínimo aceptable para asegurar un reconocimiento óptico confiable en textos impresos convencionales. Sin embargo, en el

caso de tipografías finas, documentos deteriorados o con elementos gráficos complejos, se recomienda utilizar una resolución de 600 dpi.

Antes y durante el escaneo se corrigieron problemas de alineación, manchas y brillos, se controló la iluminación y se emplearon cunas en "V" para proteger encuadernaciones. Cada sesión se documentó técnicamente, registrando equipo, operador, fecha, resolución y modo de captura.

El estándar en archivística recomienda la digitalización en dos formatos diferentes, por regla general en un formato de conservación de almacenamiento de toda la información del documento, como los formatos TIFF o PNG; y en otro formato de mayor comunicabilidad, pero como mayor pérdida de información, como por ejemplo el formato PDF. Considerando los objetivos del proyecto, orientados a la investigación historiográfica y en ciencias sociales, y de conformación de una base de datos para un repositorio documental, se optó por el formato PDF, que facilita la legibilidad y el análisis de documentos asistido por herramientas computacionales.

Finalmente, se aplicaron medidas de integridad digital como el uso de suma de comprobaciones o verificaciones, y una estrategia de respaldo 3-2-1 (tres copias en dos formatos distintos y una en una ubicación externa). De esta manera, incluso antes de la digitalización, el acervo queda protegido frente a pérdidas accidentales, corrupción de datos o fallas técnicas.

3. Recodificación

La tercera fase consistió en la recodificación del corpus de documentos. En esta fase cada documento recibió una ID o código único, al tiempo que el conjunto de los documentos fue almacenado en el repositorio de proyecto con una estructura descriptiva con fines historiográficos. Esto dio lugar a una nueva base con una nueva estructura.

Para la realización de esta fase del proyecto, implementamos una estructura jerárquica de carpetas organizada en cuatro niveles: /01_originales/, /02_scans/, /03_OCR/ y /04_entregables/. Cada nivel correspondió a un estado de procesamiento, lo que permitió mantener siempre una versión resguardada del original y separar claramente las copias intermedias de trabajo de las versiones finales de consulta. Esta separación aseguró que cualquier error técnico o fallo en las etapas posteriores pudiera revertirse sin pérdida de información.

El siguiente paso fue crear una "identidad archivística" para cada documento. Para eso aplicamos la convención de crear para cada documento un código o ID mediante una nomenclatura controlada. Cada documento recibió un nombre estructurado bajo la fórmula: "Título abreviado del documento _{número}_{año}_{calidad}. Para ello la nomenclatura utilizada fue la siguiente (tabla 2).

Tabla 2. Recodificación de documentos.

Identificación documento	Codificación
Informativo Casa de Chile en México	IChMx_{número}_{año}_{calidad}
Presencia de Casa de Chile en México	PChMx_{número}_{año}_{calidad}
Noticias de Chile. Boletín de Casa de Chile en México	NotCh_{número}_{año}_{calidad}
Cuadernos Casa de Chile en México	CCChMx_{número}_{año}_{calidad}

Fuente: Elaboración propia.

Así, por ejemplo, el archivo "CChMx_N19_1978_LQ.pdf" indica que se trata del Cuaderno de la Casa de Chile en México, número 19, publicado en 1978 y digitalizado en baja calidad. Este sistema de codificación permitió identificar de manera inmediata la procedencia y las características técnicas de cada archivo, reduciendo el riesgo de duplicado, permitiendo su recuperación, aportando trazabilidad y facilitando su ubicación en el repositorio y su consulta durante el proceso de investigación.

También, se elaboró una ficha mínima por ítem, que incluyó información esencial: código, título, año, fuente, ubicación y derechos. Este registro se complementó con notas sobre el estado físico del documento y observaciones pertinentes (dobleces, dete-

rioro, anotaciones manuscritas y otras). Así, cada documento adquirió una identidad de archivo que lo acompaña durante todo el flujo de trabajo.

4. Reconocimiento óptico de caracteres (OCR)

En la cuarta fase el objetivo fue que los materiales digitalizados fuesen susceptibles de un ulterior análisis textual computacionalmente asistido, para lo cual se estableció un reconocimiento óptico de caracteres u OCR. El OCR constituye la fase que transforma imágenes estáticas en documentos dinámicos, buscables y analizables. Su importancia radica en que habilita búsquedas por palabra clave, minería de texto y extracción automatizada de información, funciones imposibles en un simple PDF de imagen. Esta operación superpone una capa de texto al facsímil, garantizando la búsqueda, selección y análisis de contenidos sin alterar la integridad del documento.

En este caso, se realizó con la herramienta de Adobe Acrobat Pro[31], configurado en idioma español, con detección automática de orientación y corrección de "sospechosos". Estas opciones permiten que el software identifique palabras mal reconocidas y sugiera revisiones manuales.

El proceso puede describirse como un pipeline algorítmico: primero el preprocesado (*deskew* o enderezamiento de los caracteres, reducción de ruido y binarización adaptativa), luego el análisis de maquetación, seguido del reconocimiento de texto mediante modelos entrenados por idioma, un posprocesado lingüístico para corregir errores y, finalmente, la generación de un PDF-OCR con texto anclado por coordenadas.

Cada documento que antes era una imagen estática se convierte en un recurso interactivo y explorable, apto para análisis tanto cualitativos como cuantitativos.

5. Extracción de información

La cuarta fase consiste en la extracción de información estructurada a partir de los documentos procesados mediante OCR. Los materiales fueron organizados en unidades analíticas y se normalizaron campos descriptivos como título, autor, resumen y palabras clave.

En este proceso, la IA cumple un rol de apoyo, especialmente en la identificación de secciones internas de cada documento y en la sugerencia de metadatos.

Cada documento se concibe como un conjunto de unidades analíticas diferenciadas (editoriales; artículos; discursos y documentos anexos) y no como un bloque homogéneo. Siguiendo el mismo principio de codificación mediante una nomenclatura controlada. A cada unidad se le asigna una fila en una tabla con campos fijos: código, año, título, autor(es), resumen breve, tres palabras clave y un tema inferido.

La IA contribuye especialmente en tareas como la detección de encabezados, la normalización de autores y la generación automática de resúmenes y palabras clave. Sin embargo, estas sugerencias siempre requieren supervisión humana, pues los algoritmos tienden a cometer errores en tipografías irregulares o en textos con maquetaciones complejas

Esto se alinean con Guo, Ovadje, Al-Garadi y Sarker[32], quienes compararon modelos supervisados tradicionales con modelos generativos, mostrando que la combinación de extracción automática y control humano previene errores y mejora la calidad del corpus.

Los archivos digitalizados y recodificados fueron cargados en un proyecto en Chat GPT modelo 5.0 Thinking y en Gemini Pro Thinking. A través del empleo de *prompts* entrenados para cada tipo de documento analizados en este proyecto, para cada tipo

[31] Adobe, "Recognize Text in Scanned Documents with Acrobat", *Adobe Acrobat Help*, 5 de noviembre de 2025, https://helpx.adobe.com/uk/acrobat/desktop/create-documents/scan-documents-to-pdfs/recognize-text.html.

[32] Guo et al., "Evaluating Large Language Models".

de documento fue posible distinguir las secciones que lo componían, y extraer la información descriptiva de cada una de ellas para los siguientes campos: Institución, fondo, colección, ítem, identificador (ID), código, año, título, autor, resumen, tres palabras claves, tema inferido, descriptor, fecha de descripción. Lo que permitió confeccionar una base de datos en que para cada documento se especificaron los siguientes campos (tabla 3).

Tabla 3. Campos de la base de datos del repositorio documental

Campo	Explicación	Ejemplo
Institución	Corresponde a la institución donde se encuentran los documentos	Museo de la Memoria y los Derechos Humanos (MMDH)
Fondo	Corresponde a la ubicación en el archivo	UAM Xochimilco
Colección	Corresponde a la colección dentro del fondo en la que se ubica el archivo	Publicaciones Seriadas
Ítem	Corresponde a la identificación desagregada de donde se encuentra el documento	Caja B7: E29-C7.
ID	Corresponde al identificador asignado por la institución al objeto documental	00001596-000040-000001
Código	Corresponde a la recodificación efectuada por el Proyecto	CCChMx_N1_A1976_LQ
Año	Corresponde al año de producción del objeto documental	1976
Título	Corresponde al título de cada texto identificado dentro del objeto documental	Presentación
Autor (es)	Corresponde al productor del texto indicado en el objeto documental	Casa de Chile en México
Resumen	Corresponde al breve resumen del texto	Introduce la serie *Cuadernos* y explica por qué inaugura la colección con tres ensayos económicos que desmontan la propaganda del “milagro” económico de la Junta.
Tres palabras claves	Corresponden a palabras claves identificadoras del texto dentro del documento	Cuadernos; introducción; economía
Tema inferido	Corresponde al tema inferido del texto a partir del título	Lanzamiento de la serie y contexto crítico sobre la economía chilena.
Descriptor	Corresponde al investigador que efectuó el proceso de descripción	Pablo Seguel
Fecha	Corresponde a la fecha en la que se efectuó el proceso de descripción	01/07/2025
Modelo IA	Señala el modelo de Inteligencia Artificial utilizado para el proceso de extracción de la información	GPT 4.5

Fuente: Elaboración propia.

6. Doble revisión humana

La sexta fase, de doble revisión humana es indispensable para garantizar la calidad del repositorio. Se aplica un sistema de doble control: una primera revisión de extracción de la información, donde se marcan dudas o inconsistencias. Una segunda revisión de validación, que contrasta los resultados con el PDF-OCR, la imagen maestra y el sumario original.

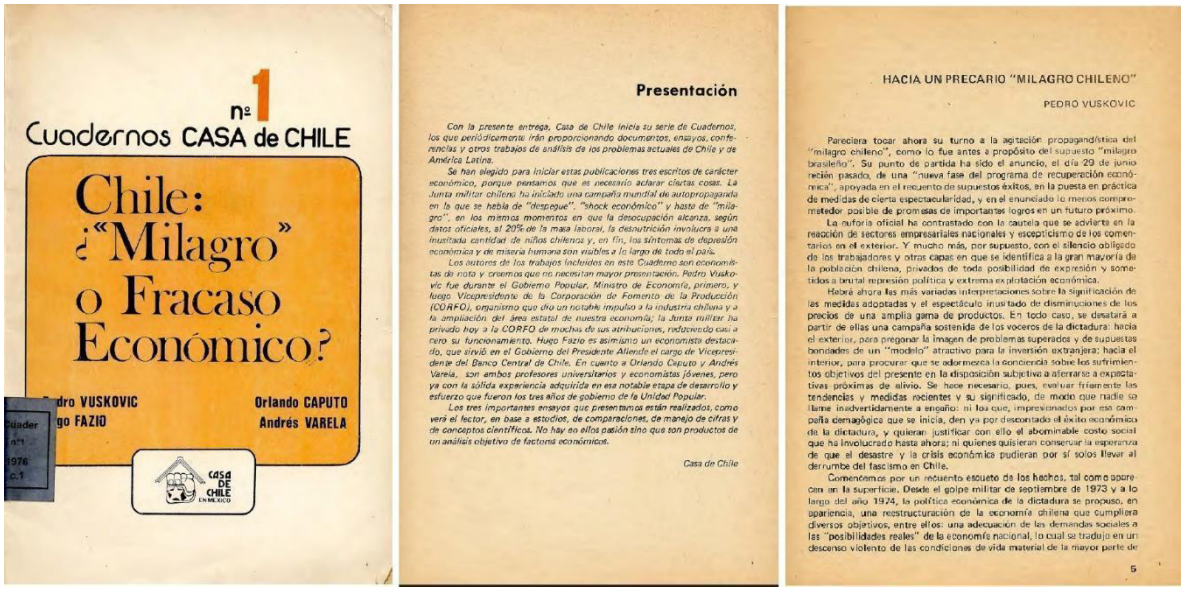
En el caso de este proyecto lo hicimos en dos etapas progresivas, cada una encargada a una persona diferente. En la primera etapa un investigador recodificó los documentos digitalizados; y en la segunda etapa, otro investigador chequeó la corrección de la recodificación y efectuó el proceso de re-

visión. Finalmente, el proceso de extracción de información, realizado por uno de los investigadores asistido por una herramienta de IA fue revisado y validado por el equipo de investigación en reuniones de trabajo específicas.

Este sistema aseguró que el repositorio no dependiera únicamente de la "confianza ciega" en la IA, sino que se construyera de manera colaborativa entre tecnología y personas. Todas las correcciones se registraron en una bitácora de cambios, lo que aportó transparencia y trazabilidad al proceso.

Dentro de los aprendizajes obtenidos, obtuvimos que la cualidad de los documentos es significativa cuando se trata del empleo de herramientas de IA para la extracción de información. Por ejemplo, cuando un documento textual tiene homogeneidad tipográfica y una jerarquización constante en su diagramación, el proceso de entrenamiento del chat box y de elaboración del *prompt* es más sencillo (imagen 1).

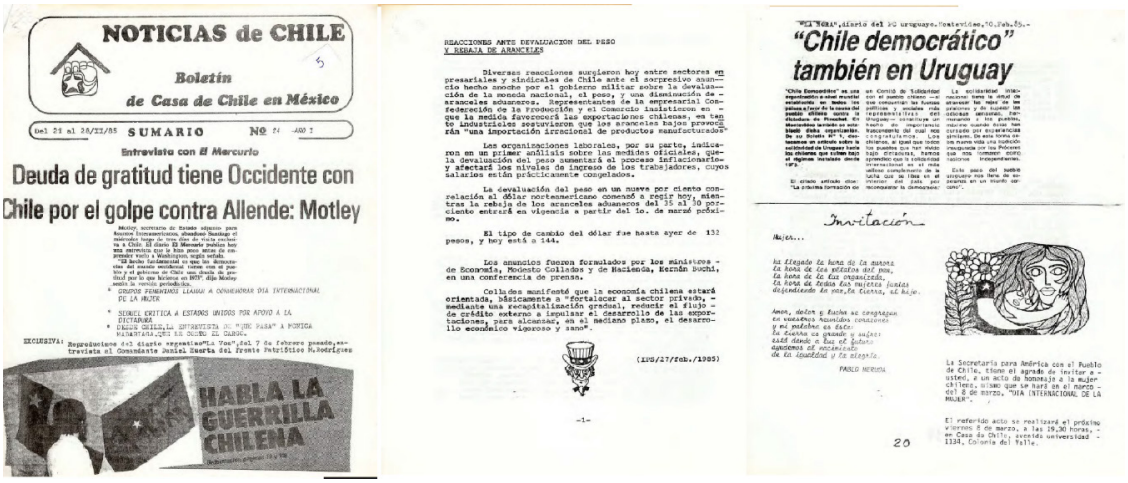
Imagen 1. Cuadernos Casa de Chile.



Fuente: MMDH.

En cambio, cuando los documentos presentan una mayor heterogeneidad en su diagramación y su tipografía, la supervisión humana se torna imprescindible para garantizar una correcta extracción de la información textual y los procesos de entrenamiento del chat y de formulación del *prompt* son más lentos y complejos (Imagen 2).

Imagen 2. Noticias de Chile. Boletín de Casa de Chile en México



Fuente: MMDH.

7. Normalización

En la fase de normalización se unificaron los criterios de metadatos, se depuraron duplicidades y se estandarizaron vocabularios, dando como resultado una base de datos final consistente y lista para su análisis historiográfico y su difusión académica.

Cuando se extrae la información de los textos dentro de cada documento con la asistencia de una IA, muchas veces no son identificados campos de información relevantes que son parte de la unidad documental. Por ejemplo, si tenemos una editorial de una revista de un año determinado, en la extracción de la información con la IA empleada, esa editorial aparece sin un año de publicación, a pesar de que le corresponde el año de la unidad documental a la que pertenece.

Por eso, esta última fase, denominada de normalización, tiene por objeto corregir eventuales pérdidas de información relevantes del texto, que sólo se pueden obtener analizándolo en su contexto documental, por ejemplo, año, autoría, género textual, entre otras.

En el caso de nuestro proyecto, se establecieron los siguientes criterios de normalización: recuperación del año de cada texto a partir del código asignado a su unidad documental; redacción de los títulos de cada texto con la denominación original en su unidad documental con la primera letra de la primera palabra en mayúsculas y todo lo demás en minúsculas (salvo sustantivos propios); identificación de autores y consignación de autores múltiples separados por punto y coma; estandarización de denominaciones heterogéneas para las mismas autorías; selección de palabras clave en minúsculas; identificación de temas a través de una taxonomía estandarizada; y finalmente, detección de duplicados y su eliminación.

El resultado es una base de datos final limpia, coherente y estandarizada, lista para ser exportada a Excel, CSV o sistemas de gestión de bases de datos.

Resultados descriptivos de la aplicación de la metodología propuesta

Con base en la metodología se analizaron 205 publicaciones, con un total de procesamiento de 6965 páginas y 3554 entradas de textos independientes identificadas (Tabla 4).

Tabla 4. Síntesis Publicaciones Analizadas

Nombre	Código	Páginas	Ejemplares	Textos
Cuadernos Casa de Chile en México	CChMx_N{número}_{Año}_{calidad}	1.310	40	392
Informativo Casa de Chile en México	ICChMx_{número}_{año}_{Calidad}	1.532	45	866
Noticias Casa de Chile en México	NotCh_N_{número}_{año}_{calidad}	3.317	103	2.126
Presencia Casa de Chile en México	PChMx_{número}_{año}_{calidad}	806	17	170
	TOTALES	**6.965**	**205**	**3.554**

Fuente: Elaboración propia.

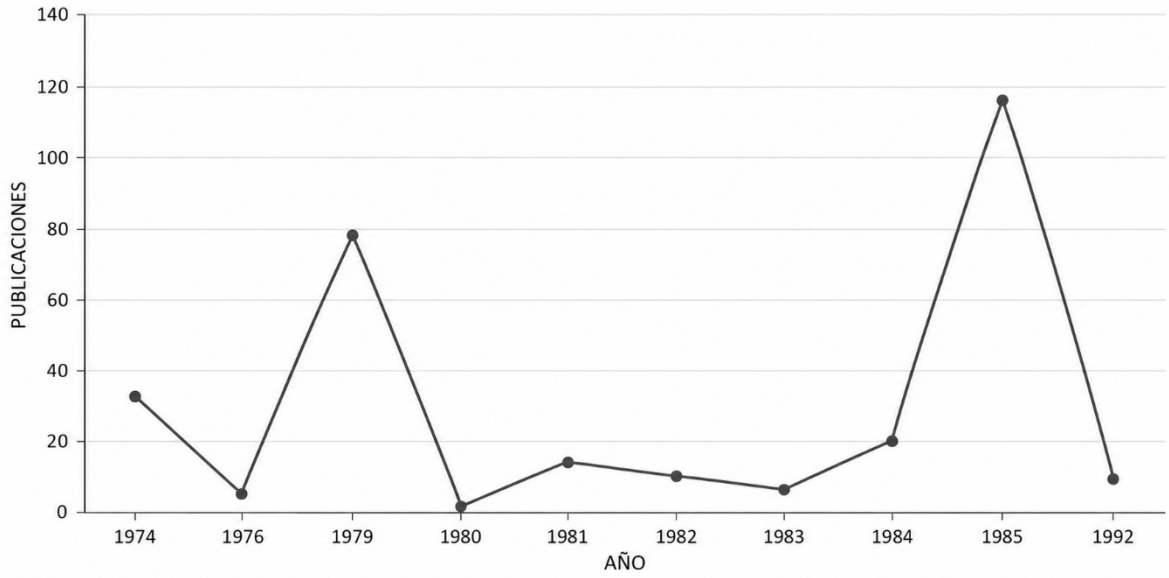
A continuación, presentaremos un breve análisis descriptivo preliminar periodizado de cada publicación, como muestra de los análisis contextualizados que una base de archivo temático como la construida permite realizar.

a) Cuadernos Casa de Chile en México

La publicación constituye una serie que se extiende entre 1976 y 1992, con un total de 40 números. Sin embargo, el conjunto presenta ocho ediciones faltantes (correspondientes a los números 20, 21, 22, 39, 40, 41, 44 y 45). Los 32 ejemplares existentes contienen 1.310 páginas disponibles para el análisis.

Al analizar el conjunto documental, se obtiene un corpus compuesto por 392 textos independientes, concentrados principalmente en los primeros años de su publicación (Gráfico 2).

Gráfico 2. Producción anual de textos en Cuadernos Casa de Chile en México.



Fuente: Elaboración propia

El análisis de la producción editorial muestra cuatro momentos claramente diferenciados. El pico fundacional de 1976-1977 concentra casi la mitad de los textos (184 de 390), lo que marca la instalación de la serie con una densidad excepcional de documentos de denuncia, análisis económicos y resoluciones políticas. En este periodo inicial, los cuadernos funcionan como una plataforma urgente de articulación del exilio chileno, orientada a visibilizar la represión y ofrecer interpretaciones inmediatas del rumbo político y económico del régimen militar.

Posteriormente, se observa un reacomodo en 1978-1979, con un descenso abrupto en 1978 (23 registros), seguido de un repunte en 1979 (56 registros). Esta variación coincide con un cambio en el tipo de contenidos: aparecen textos que analizan la consolidación del régimen, la coyuntura del plebiscito de 1978, la continuidad operativa de la Dirección de Inteligencia Nacional (DINA) pese a su formal disolución y los efectos acumulados de la crisis económica.

A comienzos de los años ochenta (1980-1983) se configura un tercer bloque, caracterizado por un volumen menor, pero por una especialización temática más clara. Los cuadernos de este periodo abordan materias específicas (economía; exilio y retorno; lucha democrática y las protestas de 1983), lo que muestra que la línea editorial se vuelve más concentrada y analítica. La serie se estabiliza como un espacio de reflexión especializada sobre el conflicto político chileno.

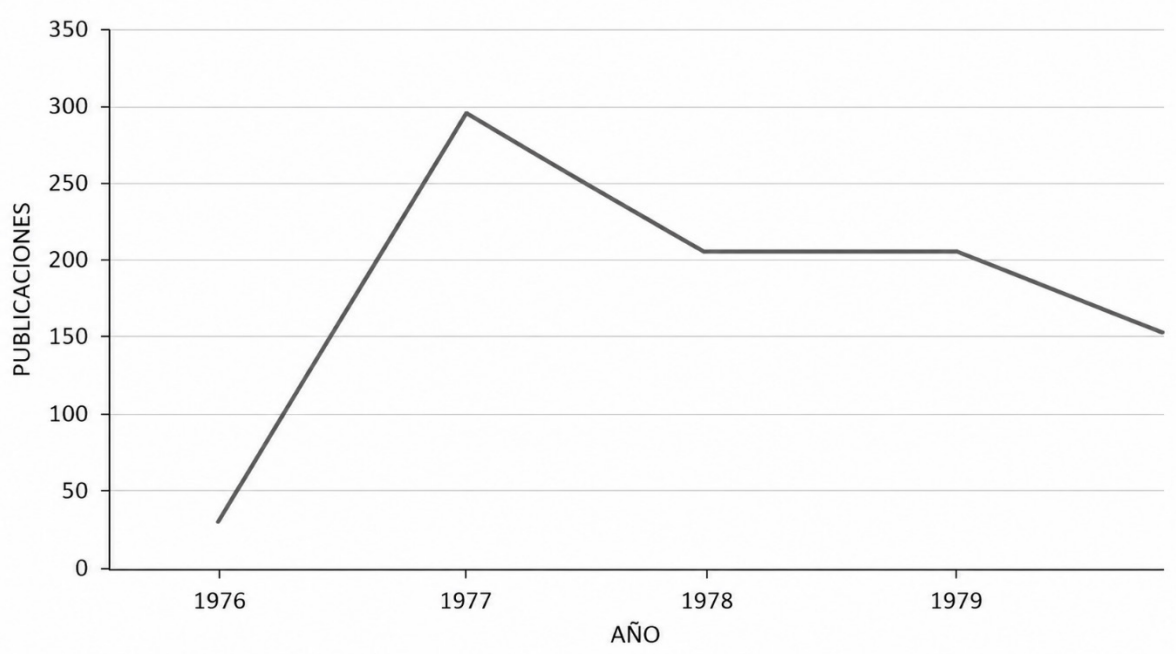
Finalmente, entre 1988 y 1992 se produce una reaparición en clave de transición a la democracia, marcada primero por el año 1988 de baja producción, centrado casi exclusivamente en temas de deuda externa y el plebiscito. En cambio, 1991-1992 muestran un retorno significativo, con 48 artículos, en los que predominan los ejes de memoria, restitución de derechos, testimonio y experiencias del exilio en el nuevo contexto democrático. Este cierre revela una relectura del periodo dictatorial desde la transición y actualiza la función de los Cuadernos Casa de Chile como espacio de memoria y balance político.

b) Informativo Casa de Chile en México

Publicaciones seriadas producidas desde 1976 hasta 1980, que abarcan desde el número 4 hasta el 57 con un total de 45 volúmenes, con varias ediciones con dos números en su interior. El conjunto presenta cinco ediciones faltantes (1, 2, 3, 7 y 52), lo que en definitiva suma de 1532 páginas.

La base de datos contiene 866 registros, cada uno correspondiente a un artículo, nota, editorial o sección identificada dentro de los *Cuadernos Casa de Chile en México*. El corpus abarca documentos publicados entre 1974 y comienzos de los años 80, con una estructura homogénea que permite observar patrones de producción editorial, énfasis temáticos y ritmos de publicación asociados a los ciclos políticos del exilio chileno y la dinámica represiva de la dictadura (Gráfico 3)

Gráfico 3. Producción anual de textos en Informativo Casa de Chile en México.



Fuente: Elaboración propia.

Entre 1976 y 1980 la producción de los *Informativos de la Casa de Chile en México* muestra un ciclo claramente escalonado, que combina expansión, estabilización y densificación temática. El año 1976 marca el inicio del proyecto, con solo 35 registros. Los contenidos iniciales se concentran en el análisis económico y político de la Junta Militar y en textos programáticos que buscan definir la identidad editorial de la colección.

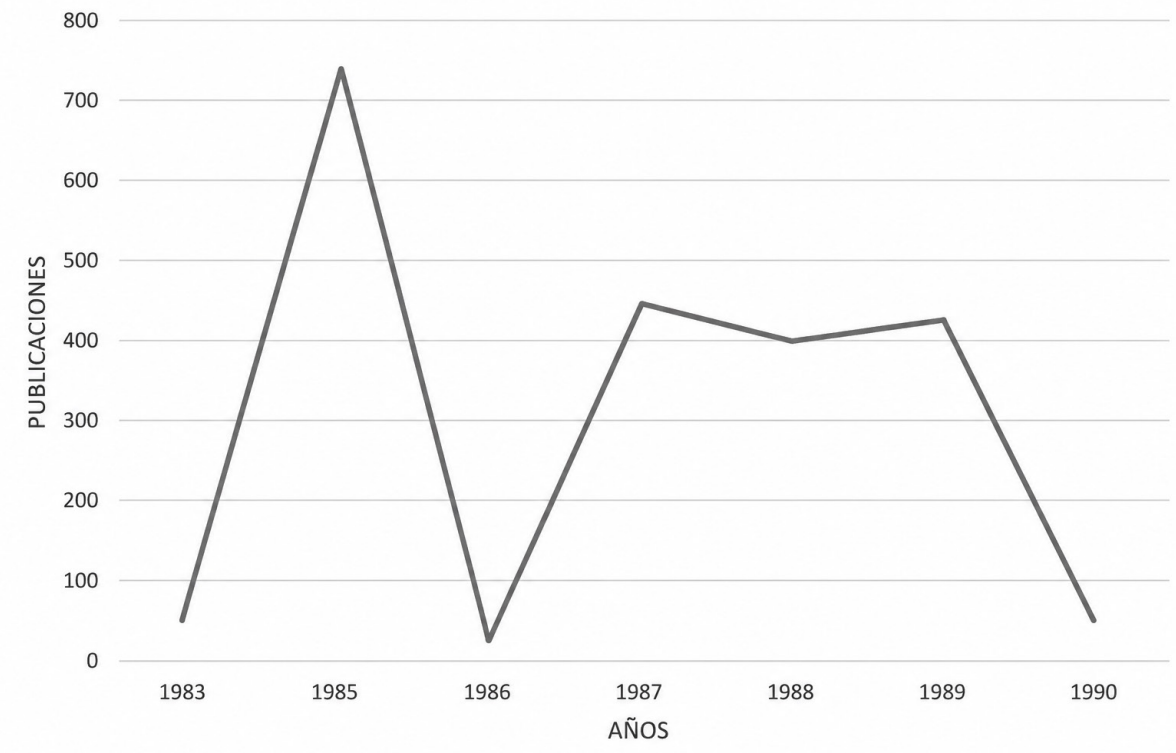
En 1977 se produce un salto cuantitativo y cualitativo decisivo. Con 289 registros, este año concentra más de un tercio de toda la base. La publicación logra una regularidad sostenida, incorpora más secciones internas y amplía la diversidad de temas: economía política, coyuntura interna chilena, relaciones internacionales, cultura y debates del exilio. Se consolidan, además, líneas de denuncia de la represión y reflexiones sobre la solidaridad internacional, aunque los derechos humanos aún no aparecen como categoría dominante.

Los años 1978 y 1979 mantienen un volumen alto, con 205 y 204 registros, respectivamente. Más que crecer, la colección se profundiza: aparecen secciones especializadas, aumentan los análisis estructurales del modelo económico y se integra un repertorio amplio de materiales breves (cables; reseñas y fichas) que densifican la información. En 1979 destacan especialmente los temas vinculados a derechos humanos, solidaridad, política y memoria, señalando una transición hacia un lenguaje más articulado de denuncia y lucha democrática, estrechamente conectado con la agenda internacional.

c) Noticias de Chile. Boletín de Casa de Chile en México

Publicaciones de noticias de Chile en el extranjero, de carácter seriado, que registra un total de 104 ejemplares con un total de 3317 páginas y 2126 registros textuales individualizados. Esta base es la con mayor cantidad de ejemplares faltantes con 142, lo

Gráfico 4. Producción anual de textos en Noticias. Boletín de la Casa de Chile en México.



Fuente: Elaboración propia.

que dificulta establecer un análisis descriptivo robusto sobre la colección (Gráfico 4).

Entre 1983 y 1985 los boletines se expanden en cobertura y frecuencia. En 1983, con 48 registros, predominan noticias sobre movilización social y violencia estatal, articuladas con una lectura internacional del conflicto, aunque todavía con un volumen acotado y una estructura editorial menos compleja.

El año 1985 marca un punto de inflexión cuantitativo y cualitativo. Con 739 registros, se configura el momento de mayor densidad de toda la serie. El boletín consolida una regularidad sostenida, multiplica sus secciones internas y amplía el repertorio temático: movilización social, represión, economía, violencia política no estatal, relaciones internacionales y solidaridad. La presencia concentrada de editoriales indica que el proyecto ya no solo informa, sino que interpreta el conflicto y busca orientar políticamente a sus lectoras y lectores.

Entre 1986 y 1990 la producción entra en una fase de reacomodo y densificación selectiva. En 1986 el número de registros desciende, pero aumenta proporcionalmente el peso de los derechos humanos y de la disputa institucional. Entre 1987 y 1988 se recuperan volúmenes altos y se estabiliza una articulación temática entre violencia estatal, violencia política no estatal, campañas internacionales y derechos humanos como eje estructurante.

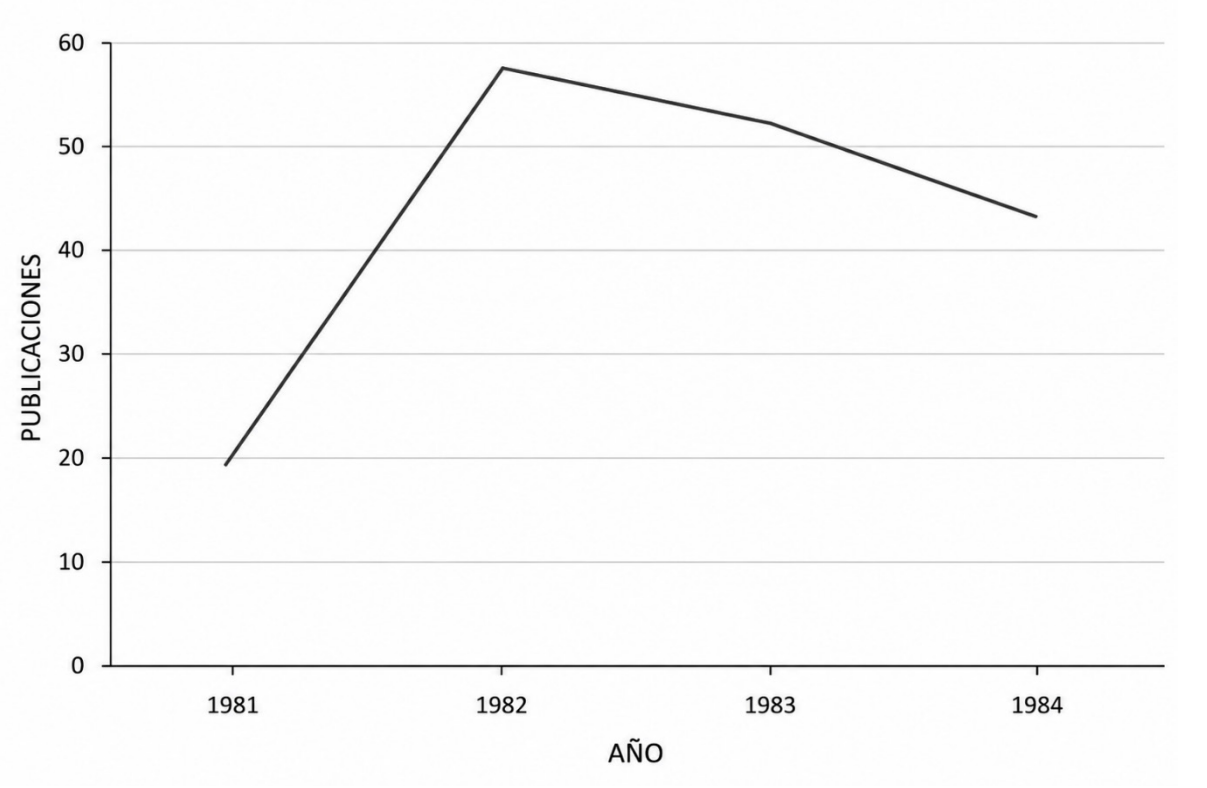
Finalmente, entre 1989 y 1990 la serie acompaña el giro transicional: sin abandonar la cobertura del conflicto, desplaza el énfasis hacia juicios, comisiones, rol de la Iglesia y negociaciones en el plano internacional, registrando la transformación de la represión en problema de legitimidad democrática.

Presencia Casa de Chile en México

Publicación seriada producida entre 1981 y 1984, que abarca un total de 17 ejemplares, con tres faltantes (1, 9 y 10) con un total de 806 páginas. En total se procesaron 170 entradas de diversas variedades temáticas (Gráfico 5).

Entre 1981 y 1984, *Presencia. Casa de Chile en México* muestra una evolución marcada en volumen, regularidad y enfoque editorial. El año 1981 constituye una fase inicial,

Gráfico 5. Producción anual de textos en Presencia. Casa de Chile en México.



Fuente: Elaboración propia.

con solo 19 registros y una estructura todavía experimental. Predominan los formatos híbridos (notas breves, cables y comentarios de coyuntura) junto con un fuerte interés por la política exterior estadounidense y por la legitimación internacional del régimen militar chileno. El periódico opera en modo de instalación, sin una periodicidad clara y con un tratamiento descriptivo más que analítico.

En 1982 la producción crece de manera sustancial, alcanzando 57 registros. Este salto cuantitativo revela un proceso de consolidación interna: mayor sistematicidad, ampliación temática (crisis económica; represión; seguridad interior y conflictos regionales) y reportajes más elaborados. La pauta se alinea con el contexto histórico de la crisis económica de 1982 y el resurgimiento de la protesta social en Chile, lo que impulsa al medio a fortalecer su función informativa y política.

El año 1983 mantiene un nivel alto de producción (52 registros), evidenciando estabilidad y madurez del dispositivo editorial. Los contenidos se orientan hacia la movilización social, las respuestas represivas del régimen y las lecturas regionales de la crisis. Se incrementa la profundidad analítica y la cobertura sistemática de la protesta, configurando un medio altamente politizado.

En 1984, aunque la cifra baja a 42 registros, se observa más un reacomodo que un declive. Disminuyen las notas breves y aumentan las entrevistas y análisis internacionales, volviendo el tono más interpretativo. En conjunto, el periodo refleja un crecimiento acelerado, una consolidación entre 1982-1983 y una transición desde un medio informativo hacia un órgano político-analítico del exilio chileno.

CONCLUSIONES

La experiencia desarrollada en el Proyecto *Exilio, solidaridad y resistencia en la Casa de Chile en México* permite observar cómo la integración entre archivística, criterios historiográficos y herramientas de IA abre un campo metodológico que hasta ahora no contaba con protocolos consolidados en la historiografía chilena ni en los estudios del exilio. La conformación del repositorio

documental no solo reorganiza un acervo disperso, sino que ensaya un modo de trabajo híbrido capaz de enfrentar corpus editoriales extensos, heterogéneos y poco normalizados, combinando automatización con criterios expertos de lectura, clasificación y verificación.

Uno de los aportes más significativos emerge del modo en que la IA permitió intervenir etapas tradicionalmente invisibles del proceso archivístico, donde el investigador suele enfrentar tareas repetitivas y de alta carga temporal: identificación de encabezados, separación de unidades textuales, sugerencias de metadatos preliminares y generación de resúmenes exploratorios. En estos segmentos, la automatización no sustituyó el juicio historiográfico, pero sí permitió ampliar el volumen de información procesada, hacer más eficiente la revisión manual y habilitar comparaciones transversales imposibles mediante métodos tradicionales. La metodología mostró que la IA es útil no por su capacidad de "acertar", sino por su capacidad de producir hipótesis descriptivas que luego fueron depuradas por la doble revisión humana.

El valor metodológico también se ve en la articulación entre normas de descripción del ICA, criterios técnicos del MMDH y decisiones historiográficas del proyecto. Al estructurar el corpus según códigos únicos, trazabilidad completa y campos normalizados, se reconstruyeron relaciones internas entre documentos que en el archivo formal aparecían fragmentadas por el criterio del donante. Esto generó un tipo de repositorio que no replica el orden institucional del MMDH, sino que complementa con una lectura del archivo centrada en la unidad de la experiencia histórica: la Casa de Chile como comunidad política y editorial del exilio. La metodología muestra, así, como un repositorio puede convertirse en una forma de interpretación y no solo de conservación.

El proceso de OCR y extracción evidenció además que las herramientas de IA funcionan de manera desigual según la estabilidad tipográfica del documento. Esto introdujo un aprendizaje metodológico crucial: la IA no opera como un dispositivo neutral, sino como una tecnología sensible a la materialidad del archivo. La necesidad de ajustar *prompts*, corregir errores sistemáticos y diseñar excepciones para publicaciones con diagramación irregular llevó a desarrollar un enfoque minucioso para el uso de modelos generativos, capaz de reconocer el flujo de trabajo necesario y cuándo requieren intervenciones adicionales. Ese punto intermedio, ni tecnofílico ni tecnofóbico, constituye uno de los aportes del proyecto.

Asimismo, la metodología de siete fases permitió transformar un conjunto de documentos heterogéneos en un corpus estructurado y explotable. El repositorio resultante no solo preserva y organiza, sino que habilita nuevas posibilidades de lectura: análisis diacrónicos de tendencias editoriales, reconstrucción de genealogías temáticas, identificación de inflexiones políticas, estudio sistemático de discursos de solidaridad o vigilancia algorítmica de patrones de lenguaje. La base de datos no reemplaza la lectura histórica tradicional, pero amplifica posibilidades y conduce hacia un terreno metodológico donde la escala y densidad documental ya no constituyen barreras.

En conjunto, se muestra que la IA puede ser integrada de forma crítica en estudios historiográficos sin perder profundidad analítica ni sensibilidad interpretativa. La clave no se ubica en delegar tareas en los algoritmos, sino en reformar los procesos de trabajo, articulando la potencia de la automatización con la densidad del razonamiento histórico. Este cruce metodológico produce repositorios que no sólo resguardan memoria, sino que crean condiciones para nuevas preguntas de investigación, fortalecen la transparencia y trazabilidad del trabajo histórico. Ofreciendo modelos replicables para otros acervos vinculados al exilio, la represión y las redes transnacionales de solidaridad.

El repositorio resultante es un instrumento historiográfico en sí mismo: una infraestructura que permite observar, inter-

pretar y comparar, ampliando el horizonte metodológico para la investigación del exilio chileno y, más ampliamente, para el estudio de fenómenos históricos cuya escala documental supera las herramientas tradicionales de la disciplina.

FUENTES

Museo de la Memoria y los Derechos Humanos. *Inventario Fondo UAM Xochimilco. Casa de Chile en México*. ID 1595, sin fecha.

Museo de la Memoria y los Derechos Humanos. *Inventario Fondo: Schultz, Fernando*. ID 1596, sin fecha.

Museo de la Memoria y los Derechos Humanos. *Políticas de digitalización de colecciones*. Santiago de Chile: Área Colecciones, mayo de 2012.

REFERENCIAS

Adobe. "Recognize Text in Scanned Documents with Acrobat". *Adobe Acrobat Help*, 5 de noviembre de 2025. https://helpx.adobe.com/uk/acrobat/desktop/create-documents/scan-documents-to-pdfs/recognize-text.html.

Baeza Duffy, Patricia. "Construcción de memorias que compiten sobre el pasado chileno reciente por adultos que vivieron ese período en Chile o en el exilio". *Discurso & Sociedad* 11, nº 3 (2017): 433–457. https://dialnet.unirioja.es/servlet/articulo?codigo=6140840.

Barrueco Cruz, José Manuel, Pilar Rico Castro y Laura Valeria Bonora Eve. *Guía para la evaluación de repositorios institucionales de investigación*. Madrid: FECYT/REBIUN, 2021. http://hdl.handle.net/20.500.11967/809.

Bondarenko, Oleksandr, Svitlana Orlyk y Mark Pyzyk. "Digital Technology in Historical Research: Contemporary Scholarly Current". *Skhidnoievropeiskyi Istorychnyi Visnyk [East European Historical Bulletin]* 33 (2024): 227–238. https://doi.org/10.24919/2519-058X.33.317459.

Carretero, Mario, y Elisa Gartner. "Artificial Intelligence and Historical Thinking: A Dialogic Exploration of ChatGPT / Inteligencia Artificial y pensamiento histórico: una exploración dialógica del ChatGPT". *Studies in Psychology: Estudios de Psicología* 45, nº 1 (2024): 80–102. https://doi.org/10.1177/02109395241241379.

Christiaens, Kim, Idesbald Goddeeris y Magaly Rodríguez, eds. *European Solidarity with Chile, 1970s–1980s*. Frankfurt am Main: Peter Lang, 2014.

Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens y Julia Noordegraaf. "Archives and AI: An Overview of Current Debates and Future Perspectives". *Journal on Computing and Cultural Heritage* 15, nº 1 (2022): artículo 4, 1–15. https://doi.org/10.1145/3479010.

Consejo Internacional de Archivos. *ISAD(G): Norma Internacional General de Descripción Archivística*. 2.ª ed. Madrid: Ministerio de Educación, Cultura y Deporte, 2000.

Consejo Internacional de Archivos. *Principios básicos sobre el papel de archiveros y gestores de documentos en la defensa de los derechos humanos*. Septiembre de 2016. https://www.ica.org/resource/basic-principles-on-the-role-of-archivists-and-records-managers-in-support-of-human-rights/.

Figueroa Chichande, Xiomara Stefania, Naya Mishelle Martillo Gallardo, Taily Brigitte Martínez Márquez y Maria Clara Torres Rodriguez. "La ética en la inteligencia artificial, desafíos y oportunidades para la sociedad moderna". *Sage Sphere International Journal* 1, nº 1 (2024): 1–24. https://sagespherejournal.com/index.php/SSIJ/article/view/1.

Guo, Yuting, Anthony Ovadje, Mohammed Ali Al-Garadi y Abeed Sarker. "Evaluating Large Language Models for Health-Related Text Classification Tasks with Public Social Media Data". *Journal of the American Medical Informatics Association* 31, nº 10 (2024): 2181–2189. https://doi.org/10.1093/jamia/ocae210.

Henriot, Christian. *The AI-Augmented Research Process: A Historian's Perspective*. Prepublicación, arXiv, 2025. https://doi.org/10.48550/arXiv.2508.01779.

Horna Romero, Martin, y Roxana Perca Chagua. "Quinquenio de la aplicación de la inteligencia artificial en la archivística: una revisión sistemática en revistas académicas". *BiD* 54, nº 2 (2025). https://doi.org/10.1344/bid2025.54.02.

Hughes-Warrington, Marnie. "Ethics for Artificial Historians". *History and Theory* 64, nº 2 (2025): 159–177. https://doi.org/10.1111/hith.12377.

Jensen, Silvina. "«¿Quién es el dueño de la historia?». Agenda para una historia conectada de la represión y los exilios en el espacio iberoatlántico en la segunda mitad del siglo XX". En *Cartografías de la represión en América Latina: historias comparadas y conectadas de la violencia estatal (1960–1990)*, editado por Gabriela Águila, Santiago Garaño y Pablo Scatizza, 95–114. Buenos Aires: Imago Mundi, 2025.

Kumarage, Tharindu, Amrita Bhattacharjee y Joshua Garland. "Harnessing Artificial Intelligence to Combat Online Hate: Exploring the Challenges and Opportunities of Large Language Models in Hate Speech Detection". Prepublicación, arXiv, 2024. https://doi.org/10.48550/arXiv.2403.08035.

Manrique-Gómez, Laura, y Jaime Humberto Borja Gómez. "Ciencia de datos para la historia: datificar las fuentes para una historia (predictiva)". *Historia y Grafía*, nº 64 (2025): 97–145. https://doi.org/10.48102/hyg.vi64.541.

Mitchell, Melanie. *Inteligencia artificial: guía para seres pensantes*. Madrid: Capitán Swing, 2024.

Nicholls, Nancy. "Defensa de DDHH en Chile en el contexto transnacional del movimiento de defensa de los derechos humanos, 1973–1990". *Estudos Ibero-Americanos* 45, nº 1 (2019): 43–56. https://doi.org/10.15448/1980-864X.2019.1.31796.

NuestramemorIA. "El Proyecto". Consultado el 7 de octubre de 2026. https://nuestramemoria.ing.uc.cl/el-proyecto/.

Oberbichler, Sarah, y Cindarella Petz. *Working Paper: Implementing Generative AI in the Historical Studies*. Leibniz Institute of European History, 2025. Zenodo. https://doi.org/10.5281/zenodo.14924737.

Perry, Mariana. "Transferencia política en el exilio chileno en los Países Bajos, 1973–1989. El caso del Instituto para el Nuevo Chile". *Historia* 50, nº 1 (2017): 175–207. https://doi.org/10.4067/S0717-71942017000100007.

Rebolledo, Loreto. *Memorias del desarraigo: testimonios del exilio y retorno de hombres y mujeres de Chile*. Santiago de Chile: Catalonia, 2005.

Rebolledo-Rebolledo, Raquel, y Gabriel Reyes-Arriagada. "El exilio chileno en México. Organización y denuncia como resistencia al desarraigo". *CUHSO* 33, nº 2 (2023): 134–149. https://doi.org/10.7770/cuhso-v33n2-art647.

Resnik, David B., y Mohammad Hosseini. "The Ethics of Using Artificial Intelligence in Scientific Research: New Guidance Needed for a New Tool". *AI and Ethics* 5, nº 2 (2025): 1499–1521. https://doi.org/10.1007/s43681-024-00493-8.

Riesco Tagle, Leonor, ed. *Humanidades digitales: miradas, experiencias y proyecciones desde América Latina*. Santiago de Chile: Ediciones Universidad San Sebastián, 2025.

Rojas Corral, Hugo, Antonia Fonck, Darwin Sanhueza, Domingo Mery y Jocelyn Dunstan. "Artificial Intelligence, Judicial Archives, and Transitional Justice: Analysis of the Quiñones Case in Chile". *Tapuya: Latin American Science, Technology and Society* 9, nº 1 (2026). https://doi.org/10.1080/25729861.2026.2698211.

Rojas Mira, Claudia. *Las Moradas del Exilio: La Casa de Chile en México (1973–1993)*. Santiago de Chile: Ediciones Universidad Tecnológica Metropolitana, 2023.

Roniger, Luis. "Exile, Transnational Life, Return, and Diasporas: The Southern Cone Experience". *Middle Atlantic Review of Latin American Studies* 7, nº 1 (2023): 1–29. https://doi.org/10.23870/marlas.431.

Roniger, Luis. "El giro transnacional en estudios latinoamericanos. Una hoja de ruta". *Revista Mexicana de Ciencias Políticas y Sociales* 70, nº 254 (2025): 17–46. https://doi.org/10.22201/fcpys.2448492xe.2025.254.91325.

Ruiz Muñoz, Geovanny Francisco, Soledad Yessenia Ramírez Mantilla, Luisa Joselyn Ortiz Villón y Juanita Irene Guevara Burgos. "Algoritmos de la memoria: aplicaciones de la inteligencia artificial en la reinterpretación histórica y el análisis de narrativas sociales". *Reincisol* 4, nº 7 (2025): 3127–3155. https://doi.org/10.59282/reincisol.V4(7)3127-3155.

Seguel, Pablo. *Soldados de la represión. Anticomunismo, contrasubversión y seguridad nacional en las fuerzas armadas chilenas, 1970–1975*. Santiago de Chile: Ediciones Universidad Alberto Hurtado, 2022.

Spina, Salvatore. "Artificial Intelligence in Archival and Historical Scholarship Workflow: HTR and ChatGPT". *Umanistica Digitale*, nº 16 (2023): 125–140. https://doi.org/10.6092/issn.2532-8816/17205.

Zaagsma, Gerben. "Facing the History Machine: Toward Histories of Digital History". *History of Humanities* 9, nº 2 (2024): 451–491. https://doi.org/10.1086/731827.