

El álbum de Zoila Aurora Cáceres: inteligencia artificial y humanidades digitales en el abordaje ético y crítico del archivo

Ainaí Morales Pino, Javier Vera, Mónica Arakaki

Pontificia Universidad Católica del Perú

lmoralesp@pucp.edu.pe

jveraz@pucp.edu.pe

monica.arakaki@pucp.edu.pe

Recibido: 29/10/2024

Aceptado: 10/11/2024

Resumen

El álbum personal de la escritora, asociacionista y activista feminista Zoila Aurora Cáceres (Lima, 1872-Madrid, 1958) reúne documentos heterogéneos (en términos de contenido y materialidad) que registran su trayectoria entre 1880 y 1910. Destacan correspondencias, fotografías, postales, notas periodísticas, manuscritos y reseñas literarias que evidencian sus redes intelectuales.

Este conjunto documental es el objeto de estudio de un proyecto de investigación desde las herramientas de las humanidades digitales, la inteligencia artificial (IA) y las premisas de la ciencia abierta. El interés en este álbum como artefacto radica en su marginalidad dentro de los archivos institucionales, debido a distintos factores vinculados tanto con las notas patriarcales que imperan en tales repositorios (Derrida, 1995) como con los desafíos que plantea su heterogeneidad para un abordaje crítico convencional.

Desde una concepción del álbum como archivo en sí mismo, planteamos una aproximación feminista para un uso ético y crítico de los recursos de la inteligencia artificial en dos áreas de aplicación: por un lado, el reconocimiento automático de textos, angular para facilitar la accesibilidad, legibilidad y visibilidad de los múltiples documentos contenidos en el álbum; y, por el otro, la inteligencia artificial generativa mediante la creación de plataformas y recursos interactivos. Es nuestra intención que ambas tecnologías acerquen el álbum y la autora a una comunidad diversa de usuarios que tengan en los documentos de Cáceres —y, en etapas futuras del proyecto, de otras intelectuales— una fuente esencial para la construcción de conocimiento sobre su legado, la literatura, el feminismo y la historia intelectual.

Palabras clave: inteligencia artificial; RAG; HTR; archivos feministas; Zoila Aurora Cáceres.

Zoila Aurora Cáceres's Album: Artificial Intelligence and Digital Humanities in the Ethical and Critical Approach to the Archive

Abstract

The personal album of Zoila Aurora Cáceres (Lima, 1872-Madrid, 1958) —writer, associationist, and feminist activist— contains heterogeneous documents (in terms of content and materiality) as a record of her path between 1880 and 1910. Notable items include correspondence, photographs, postcards, journalistic notes, manuscripts, and literary reviews that highlight her intellectual networks.

This collection of documents is the focus of a research project where digital humanities tools, artificial intelligence, and principles of open science are applied. We are interested in this album as an artifact because of its marginality within institutional archives, associated with prevailing patriarchal narratives in such repositories (Derrida, 1995), as well as challenges posed by its heterogeneity for conventional critical approaches.

Conceiving the album as an archive in itself, we propose a feminist approach for an ethical and critical use of artificial intelligence resources in two application areas: on one hand, automatic text recognition which becomes crucial for enhancing accessibility, readability, and visibility of the multiple documents contained in the album; and on the other hand, generative artificial intelligence for creating interactive platforms and resources. Our intention is for both technologies to bring the album and its author closer to a diverse community of users who view Cáceres's documents —as well as those of other intellectuals in future phases of the project— as an essential source for building knowledge about her legacy, literature, feminism, and intellectual history.

Keywords: artificial intelligence; RAG; HTR; feminist archives; Zoila Aurora Cáceres.

O álbum de Zoila Aurora Cáceres: inteligência artificial e humanidades digitais na abordagem ética e crítica do arquivo

Resumo

O álbum pessoal da escritora, associacionista e ativista feminista Zoila Aurora Cáceres (Lima, 1872-Madri, 1958) reúne documentos heterogêneos (em termos de conteúdo e materialidade) que registram sua trajetória entre 1880 e 1910. Destacam-se correspondências, fotografias, cartões-postais, notas jornalísticas, manuscritos e resenhas literárias que evidenciam suas redes intelectuais.

Esse conjunto documental é o objeto de estudo de um projeto de pesquisa que utiliza ferramentas das humanidades digitais, inteligência artificial (IA) e premissas da ciência aberta. O interesse por este álbum como artefato reside em sua marginalidade dentro dos arquivos institucionais, devido a fatores vinculados tanto às narrativas patriarcais que imperam nesses repositórios (Derrida, 1995), quanto aos desafios impostos por sua heterogeneidade para uma abordagem crítica convencional.

A partir da concepção do álbum como arquivo em si mesmo, propomos uma perspectiva feminista para o uso ético e crítico dos recursos de inteligência artificial em duas áreas de aplicação: por um lado, o reconhecimento automático de textos, essencial para facilitar a acessibilidade, legibilidade e visibilidade dos múltiplos documentos contidos no álbum; e, por outro, a inteligência artificial generativa, com o desenvolvimento de plataformas e recursos interativos. Nosso objetivo é que essas tecnologias aproximem o álbum e sua autora de uma comunidade diversa de usuários, que possam encontrar nos documentos de Cáceres —e, em fases futuras do projeto, de outras intelectuais— uma fonte essencial para a construção de conhecimento sobre seu legado, a literatura, o feminismo e a história intelectual.

Palavras-chave: inteligência artificial; rag; htr; arquivos feministas; Zoila Aurora Cáceres.

Aunque la inteligencia artificial lleva varias décadas con nosotros, su mención, especialmente en el campo de las humanidades, concita una serie de debates, interrogantes y, sobre todo, resistencias. La capacidad creativa, la autenticidad y, principalmente, la ética, suelen ser los grandes ámbitos temáticos e ideológicos en los que podemos agrupar estas interpelaciones, agudizadas al considerar el talante hegemónico prevalente en la denominada inteligencia artificial generativa.¹ Al producir contenidos y respuestas sobre la base de material predominante en distintos repositorios a los que tradicionalmente han tenido acceso figuras, lenguas, imaginarios ideológicos centrales o de poder (considerando las tácitas pero existentes jerarquías culturales, sociosexuales y geopolíticas), la inteligencia artificial parece contribuir con la perpetuación de sesgos que resultan especialmente visibles y problemáticos para quienes trabajamos con documentos no hegemónicos, como lo serían, en este caso, las escrituras de mujeres latinoamericanas en el tránsito al siglo XX, sus registros, los materiales que las soportan y, en suma, sus archivos. Así, en pleno siglo XXI, parece que la llamada era digital nos lleva de vuelta a las mismas batallas no superadas en torno a la invisibilidad y minorización de las mujeres, su producción intelectual, sus documentos; al igual que en lo tocante a las diversas instancias que evidencian sus formas de participación y negociación con una modernidad de la que, pese a todo, y en medio de múltiples contradicciones, también fueron partícipes. Rita Felski, en *The Gender of Modernity* (1995), plantea que, si bien las lecturas sobre la modernidad que enfatizan la dominación masculina tienen el mérito de visibilizar estas diferencias de poder que son parte clave de proyectos de modernidad, también reafirman una reductiva visión monolítica sobre la otredad y ausencia de poder de las mujeres:

The resulting vision of male agency and female powerlessness precludes any consideration of women's distinctive roles within and active contributions to historical processes. Within the constraints of a single mythic narrative, it is inevitably an who assumes the role of collective subject of history, while woman can exist only only as Other, as the object rather than the subject of historical narrative (Felski, 1995, p. 7).²

Sin ánimos de desconocer las sistemáticas diferencias y desigualdades sociosexuales que van de la mano con la modernidad (Amorós, 1991; Valcárcel, 2001), un abordaje feminista a los archivos de cara a la ampliación crítica del discurso historiográfico debe plantear un marco que haga posible ver las formas de participación de las mujeres, sus inscripciones, intervenciones y, como ya lo señalábamos, sus pactos y complejas (hasta contradictorias) negociaciones.

Por otro lado, cuando se trata de los procesos de construcción de las historias socionacionales y de fijación de imaginarios canónicos y canonizados, no son pocos los esfuerzos por asentar datos, cronologías, conformar archivos y repositorios institucionales que fungen como base del discurso historiográfico. Sin embargo, en lo relativo a las escritoras e intelectuales mujeres, la operación parece ser diametralmente opuesta. Así, nos topamos con olvidos o silencios sistemáticos que tornan cada vuelta sobre su producción una suerte de «descubrimiento» o rescate excepcional, celebrado con euforias transitorias. Esto socava la posibilidad de articular tradiciones, continuidades y genealogías, no solo estéticas e ideológicas, femeninas y feministas, sino, también, críticas. En este sentido, los archivos y repositorios de mujeres parecen carecer, intencionalmente, de eso que Hugo Achugar (2004), en diálogo con los postulados de Hannah Arendt y René Char, ha denominado «la metáfora del testamento», es decir, esos documentos que posibilitan la filiación, la historicidad y la viabilidad del futuro:

La relación entre pasado y presente es una relación entre pasado y futuro. Narración del pasado, memoria, tradición, herencia, testamento, funcionan en cierto nivel, como sinónimos. La filiación —como el testamento— establece una tradición, una memoria, una herencia. Suele ocurrir, sin embargo, que hay herencias que se rechazan, que hay legados que despojan, que hay tradiciones

¹ Diversos trabajos críticos abordan las problemáticas que plantean las tecnologías vinculadas con la inteligencia artificial generativa. Destacamos, en tal sentido, los planteamientos de Joaquín Fernández Mateo (2024) acerca de la literatura en los tiempos del «algoriceno», sus posibilidades, pero, sobre todo, las resistencias articuladas desde la escritura (pp. 4-6).

² «La visión resultante de la agencia masculina y la impotencia femenina excluye cualquier consideración en torno a los papeles desempeñados por las mujeres dentro de los procesos históricos, al igual que sus activas contribuciones a los mismos. Una de las limitaciones que plantea esta narrativa mítica es que el hombre es quien asume el papel de sujeto colectivo de la historia, mientras que la mujer solo puede existir como Otro, como objeto y no como sujeto de la narrativa histórica» (Felski, 1995, p. 7; la traducción es nuestra).

que se cambian, en lugar de memorias, hay olvido. Entre otras cosas, por la sencilla razón de que el testamento supone la existencia de un sujeto —individual o colectivo— que lo enuncie (Achugar, 2004, p. 30).

De allí que el trabajo con las mujeres y sus archivos, llenos de vacíos, opacidades y silencios que pasan, sobre todo, por la negación de la posibilidad de ser sujetos de enunciación y desplegar una subjetividad propia (no en balde los imperativos genéricos que se imponen sobre las mujeres reducidas a un «eterno» —y externo— femenino [Torres-Pou, 1998]), exija como ejercicio ético-estético y político lo que Saidiya Hartman (2021) ha denominado las «fabulaciones críticas». Según Hartman, se trata de respuestas y resistencias a las consabidas elisiones de los archivos institucionales, o el minorizado lugar que ocupan en ellos los materiales de sujetos con identidades intersectadas por factores sociosexuales, socioraciales, ideológicos y educativos.

El caso que nos ocupa, el de la escritora, intelectual, asociacionista y activista feminista, nacida en Lima y fallecida en Madrid, Zoila Aurora Cáceres (1872-1958), es elocuente en este sentido, sobre todo en lo tocante a los «balbuces» (Achugar, 2004) de su archivo, pues, pese a no haber estado atravesada por una minorización socioracial ni haber enfrentado las dificultades para el acceso a la educación que sí tuvieron las escritoras que le precedieron (de hecho, mientras muchas de sus antecesoras fueron autodidactas, Cáceres llega cursar estudios doctorales en La Sorbona), sus materiales siguen estando disgregados en diversos repositorios en los que, con frecuencia, ingresa en calidad de apéndice o documento secundario. De hecho, su álbum personal, objeto de estudio de este trabajo, está incluido dentro del archivo dedicado a su padre, el general Andrés Avelino Cáceres,³ en las Colecciones Especiales del Sistema de Bibliotecas de la Pontificia Universidad Católica del Perú.⁴ Esto implica que, si bien gozó de un importante reconocimiento por sus contribuciones literarias e intelectuales (de lo que dan cuenta los prestigiosos sellos editoriales que la publicaron y medios impresos que la celebraron), a la fecha, Cáceres sigue siendo poco conocida más allá de ámbitos socioliterarios muy específicos. Con frecuencia, la autora es vista como una extensión bien sea de la figura paterna, o bien del escritor modernista guatemalteco Enrique Gómez Carrillo, con quien estuvo casada por menos de un año, en 1906: una etapa muy temprana dentro de su carrera como escritora, intelectual y asociacionista feminista.

Vemos entonces cómo a las dificultades que enfrentan las mujeres para ingresar al discurso historiográfico se le añaden las complejidades y azares que atraviesan sus tránsitos al archivo. Cuando logran situarse dentro de estos repositorios angulares para el discurso histórico posterior, lo hacen desde un lugar de subordinación que las concibe como extensión de una figura masculina (en este caso, la figura paterna). Esto evidencia aquello que Jacques Derrida denominó las notas «patriarchivas» de estos repositorios que, al decidir lo que se conserva y lo que no, cómo se conserva, qué se recuerda y qué se olvida, son la base de discursos de historia y memoria que soslayan la presencia y participación activa de las mujeres (Derrida, 1995, p. 12).

El álbum personal de Zoila Aurora Cáceres, objeto de estudio de este trabajo, reúne un corpus heterogéneo de documentos de diversa materialidad, intencionalidad y contenido: cartas de visita, correspondencias intelectuales políticas y/o afectivas, artículos periodísticos, reseñas literarias y culturales, menús de cenas, tarjetas postales, recibos de anuncios en prensa, fragmentos de textos literarios, ensayísticos o declaraciones de principios de las mismas asociaciones fundadas por Cáceres, como es el caso del manuscrito sobre el deber ser del Centro Social de Señoras, establecido por la autora en el año 1905. Así, el álbum da cuenta de las múltiples facetas de la vida de Cáceres entre 1880 y 1910. Planteamos esta fecha como arco temporal pues, si bien muchos documentos carecen de coordenadas geográficas, autoriales y temporales (lo que las sitúa en el horizonte de las disidencias frente a los archivos hegemónicos, en tanto capitalizan los vacíos ante la linealidad y coherencia «patriarchiva»), buena parte de los textos se sitúan entre los años de exilio familiar en Tacna y Arica, entonces bajo jurisdicción chilena tras la Guerra del Pacífico; también durante el exilio en Buenos Aires; durante su vida París; al igual que en sus diversos tránsitos por algunas ciudades italianas, alemanas y españolas. Complejizan esta cartografía de redes, relaciones y desplazamientos los lugares de origen de las múltiples correspondencias que recibe y conserva en este archivo personal, con lo que el álbum deviene un mapa denso y dinámico que pone en evidencia la presencia y agencia de la autora en diversos circuitos literarios, intelectuales y sociales.

³ Andrés Avelino Cáceres (Ayacucho, 1836- Lima, 1923) fue un militar y político peruano que desempeñó un rol clave en la guerra contra Chile (1879-1883). Fue presidente del Perú entre 1886-1890 y 1894-1895. Además, desempeño funciones diplomáticas en diversas capitales europeas.

⁴ <https://hdl.handle.net/20.500.12534/VDSQKG>

Ahora bien, hemos comentado que los textos de la autora siguen siendo escasamente conocidos, reeditados e incorporados dentro de las historias literarias e intelectuales latinoamericanas, que continúan reduciendo a Cáceres a la hija de un prohombre o una suerte de voz excepcional, *rara avis*, algunas veces por excéntrica en su singular recorrido y trayectoria intelectual y, otras, por su declarado catolicismo y su apego a los valores burgueses. Todo ello pese a haber sido publicada en las principales editoriales del momento (como Garnier en Francia, o Renacimiento en España), y pese a que sus textos contaron con prólogos de destacadas figuras dentro del campo literario e intelectual trasatlántico, como Rubén Darío, Amado Nervo, Miguel de Unamuno, entre otros (Morales Pino, 2020). Si este silencio se impone sobre sus textos, cabe entonces preguntarse qué ocurriría con un dispositivo como el álbum personal, el cual representa, por su heterogeneidad de contenidos y materialidades, múltiples desafíos vinculados tanto con las aproximaciones desde las disciplinas tradicionales, que podrían descartar su relevancia al concebirlo como un objeto femenino, íntimo y despolitizado, como con los sistemas de clasificación y, por supuesto, las políticas de conservación.

Si el carácter patriarcal de los repositorios institucionales expone, según Derrida, los límites de la democracia y la ciudadanía (1995, p. 12), especialmente cuando se trata de las mujeres, los álbumes, pensados como archivos, evidencian formas de contestación al poder y de participación de una modernidad en la que el funcionamiento desigual y divergente de la democracia fue, sin embargo, insuficiente para desincentivar la participación de las mujeres. Bien señala Ellen Gruber Garvey (2012) que la «escritura con tijeras», en los álbumes o *scrapbooks*, fue un recurso al que apelaron con frecuencia sujetos femeninos que, en el nuevo escenario de reproductibilidad técnica y acceso más difundido al papel, podían también reorganizar su contexto e historia (p. 209) desde enfoques que, no obstante, suelen ser blanco de cuestionamientos debido a su fragmentación. Reelaborando los planteamientos de Achugar (1999), podemos señalar que, así como es posible hablar de un balbuceo teórico latinoamericano, también es viable hablar de un balbuceo archivístico que evidencia las formas de articulación de historia, memoria y tránsitos a la modernidad por parte de las mujeres. En consecuencia, es posible ver en la fragmentación, ausencia de datos o continuidad de estos álbumes una contestación ético-política a las estructuras patriarcales que sostienen el archivo.

En esta línea de reflexión feminista que busca reconocer, más allá de la victimización u otredad de las mujeres, sus formas de participación e inscripción como sujetos modernos (Felski, 1995) y de modernidad (Araujo, 2010), leemos en el álbum un archivo que, al desbordar las parcelaciones tradicionales entre lo público y lo privado, y al insertar una serie de ambivalencias dentro de significantes e imaginarios que el sistema patriarcal plantea como monolíticos, constituye una instancia clave para ampliar críticamente los discursos historiográficos interpelados desde lo que se ha visto, tradicionalmente, como sus márgenes. Es decir, una reflexión histórico-cultural que sea, a su vez, una propuesta metodológica del borde al centro, considerando la tradicional minorización del álbum personal de la autora y, en diálogo con ello, de estas fuentes angulares para ponderar y comprender otras aristas de las esferas de acción e intervención de las mujeres. Es decir, consideramos al álbum, con todo y su heterogeneidad o, más bien, gracias a ella, como un documento central para una reflexión sociohistórica con impronta de género que pondere las divergencias, los silencios, los balbuceos y las contradicciones como desacatos productivos, en términos ideológicos y semánticos, y desafiantes desde un enfoque metodológico.

En las páginas que siguen, comentaremos el flujo de trabajo realizado en este proyecto, que, como mencionamos, aspira a ser una contribución metodológica y teórico-crítica, pero, también, ética en los tiempos en que se impone la hegemonía de los algoritmos. En primer lugar, ahondaremos en el proceso vinculado con el reconocimiento automático de textos, lo que es clave para facilitar la lectura y accesibilidad de los documentos contenidos en el álbum, en distintos idiomas, caligrafías y soportes materiales concebidos como semántica e ideológicamente significativos. También presentaremos los trabajos avanzados en la elaboración de plataformas interactivas que, si bien se enmarcan en la inteligencia artificial generativa, permiten una aproximación ética y crítica por parte de los usuarios. El objetivo es mostrar la posibilidad de usar la inteligencia artificial generativa como una herramienta que acerque a textos y autoras desafiantes, que permita lidiar con su compleja heterogeneidad, pero que, al mismo tiempo, no incurra en las alucinaciones, los sesgos o las falencias en las que suelen incurrir las plataformas tradicionales de inteligencia artificial generativa. Se trata entonces de capitalizar estos recursos tecnológicos hegemónicos y conciliarlos con materiales no hegemónicos, sin que esto vaya en detrimento de la agencia, la responsabilidad, el sentido crítico y ético que ha de imperar en un trabajo situado (en el campo de los estudios feministas latinoamericanos) en las humanidades en tiempos digitales.

El reconocimiento automático de textos

Desde su nacimiento como instituciones al servicio de la preservación y búsqueda del conocimiento, las bibliotecas y los archivos han llevado a cabo grandes esfuerzos para poner sus colecciones a disposición del público, labor que se ha visto apuntalada gracias a la tecnología. En este sentido, la digitalización ha contribuido a democratizar el acceso a considerables volúmenes de documentos impresos y manuscritos, y, como resultado, una amplia comunidad está en condiciones de consultar fuentes primarias y apreciar detalles de su composición visual.

A lo anterior se suman los avances logrados en la última década en materia de inteligencia artificial a través de redes neuronales y aprendizaje automático (*Machine Learning*), que hicieron posible el desarrollo del reconocimiento automático de textos y, de esta manera, abrieron las puertas a los contenidos textuales de los documentos, cuyas transcripciones habían sido anteriormente costosas de obtener.

A diferencia de una tecnología como OCR (*Optical Character Recognition*), abocada al reconocimiento de caracteres individuales, el HTR (*Handwritten Text Recognition*) busca identificar patrones en líneas completas de texto para producir las transcripciones. Lo consigue efectuando cálculos probabilísticos para extraer las características visuales de la escritura bajo análisis y generando modelos a partir de datos de entrenamiento previamente validados por intervención humana. Se trata de un caso de *aprendizaje supervisado*, en el que, en palabras de Hodel (2022), «the algorithm aligns input with a human-determined outcome, and training processes try to get the algorithm's actual output as close as possible to the desired result» (pp. 160-161).

Este es el enfoque llevado a la práctica por el proyecto READ (*Recognition and Enrichment of Archival Documents*, 2016-2019) que, con financiamiento de la Unión Europea y el liderazgo de la Universidad de Innsbruck, lanzó Transkribus como una aplicación gratuita descargable basada en Java en 2015 y como plataforma web en 2017 (Muehlberger et al., 2019). Desde entonces, según reporta la literatura académica, ha sido utilizada principalmente en la investigación en humanidades (Nockels et al., 2022).

Los modelos pueden entrenarse para reconocer diversas tipografías y caligrafías, siempre y cuando se cuente con suficiente cantidad de datos validados. En un primer momento, al vincular las imágenes digitalizadas con las correspondientes transcripciones manuales, los usuarios generan datos que, en el ámbito del aprendizaje automático, reciben el nombre de *ground truth*. Sobre la base de lo anterior, los motores de HTR se entrenan en reconocer patrones para, en un segundo momento, ser capaces de generar transcripciones cuando se les presenta material similar no visto previamente. El proceso de entrenamiento de las redes neuronales sigue la misma metodología, con el consiguiente potencial de reconocer escrituras en cualquier lengua y de cualquier período (Muehlberger et al., 2019).

Una característica resaltante de Transkribus es el fomento de una comunidad de usuarios (conformada por archivistas, investigadores en humanidades, científicos computacionales y público en general), cuya dinámica se refleja, entre otras cosas, en el hecho de compartir los modelos desarrollados. De acuerdo con su sitio web, a la fecha, 155 modelos se encuentran disponibles de forma pública para el reconocimiento de textos del siglo IX en adelante, en 49 lenguas y 32 sistemas de escritura. A manera de ejemplo, podemos mencionar el modelo entrenado con las cartas que Carlos V escribiera en 1543 a su hijo Felipe,⁵ o aquel basado en páginas del *Diario de Madrid*⁶ (1788-1825).

El álbum personal de Cáceres y los avances en su transcripción

La Pontificia Universidad Católica del Perú (PUCP) tomó la decisión, en 2013, de poner en marcha su repositorio institucional, con el fin de «recopilar y organizar los documentos de carácter académico, científico e institucional —producidos o custodiados por la PUCP— para su divulgación, en apoyo a la investigación y creación de nuevo conocimiento» (PUCP, 2024). A la fecha, esta plataforma alberga más de 50 000 documentos de diversas unidades académicas.

Como se ha mencionado, el álbum de Zoila Aurora Cáceres es conservado por el Sistema de Bibliotecas PUCP en sus Colecciones Especiales. A lo largo de sus 96 hojas (25 x 37 cm), encontramos recortes de periódico, fotografías, tarjetas postales, tarjetas de invitación, cartas, carnés, etcétera, que revelan detalles de su vida personal, académica y profesional, aproximadamente entre 1890 y 1910. Fue digitalizado e inicialmente

⁵ <https://readcoop.eu/model/carlos-v-charles-v/>

⁶ <https://readcoop.eu/model/spanish-print-xviii-xix/>

publicado en el Repositorio Institucional PUCP en 2016 y, posteriormente, migrado al Portal de Datos Abiertos PUCP en 2024. Esta última plataforma brinda acceso tanto al documento completo (un archivo PDF de 304 páginas) como a las páginas individuales del álbum (304 archivos JPG en alta resolución).⁷

Una de las primeras acciones de nuestro proyecto involucró catalogar los ítems presentes en el álbum, tarea que tuvo como resultado un *dataset* con 14 metadatos descriptivos (campos) y 277 registros. Esta identificación individual de los ítems facilitó en gran medida la carga de las imágenes digitalizadas del álbum a Transkribus.

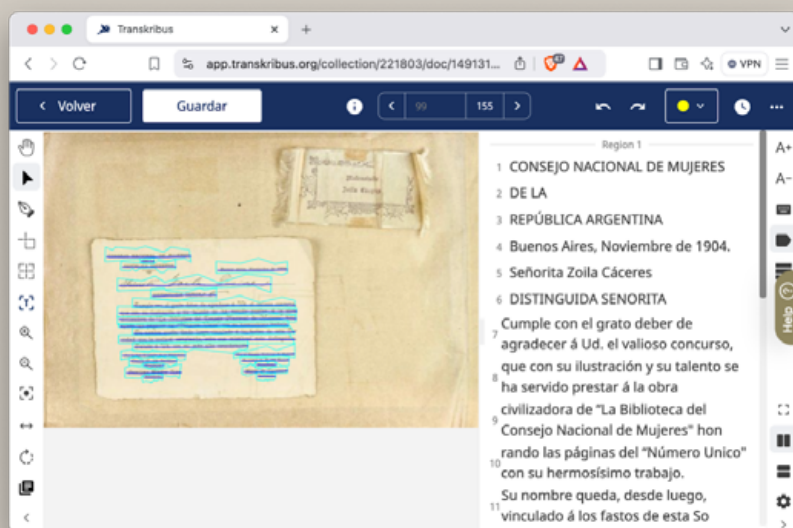


Fig. 1. Transcripción de tarjeta de agradecimiento a Cáceres

Después de la obtención de transcripciones para los primeros 50 ítems impresos, nos encontramos en el proceso de cubrir la totalidad de esta colección, caracterizada por una gran diversidad documental y multiplicidad idiomática (español, francés, alemán, inglés e italiano).

La figura 2 muestra el carné que acreditaba a Zoila Aurora Cáceres como estudiante matriculada en la Escuela de Altos Estudios Sociales de La Sorbona. Se aprecia la segmentación del texto y la correspondiente asociación entre las líneas (circunscritas a polígonos) y la secuencia de caracteres (transcripción manual). Estas transcripciones son de enorme valor para mejorar la accesibilidad y el recorrido a través del álbum y, sobre todo, la puesta en diálogo de los distintos documentos que lo integran. De esta manera, intentamos lograr superar la reducción del álbum de Cáceres a una fuente de información para extraer datos históricos o socioculturales y considerarlo, en cambio, como un documento complejo conformado por múltiples, y rizomáticas, w.

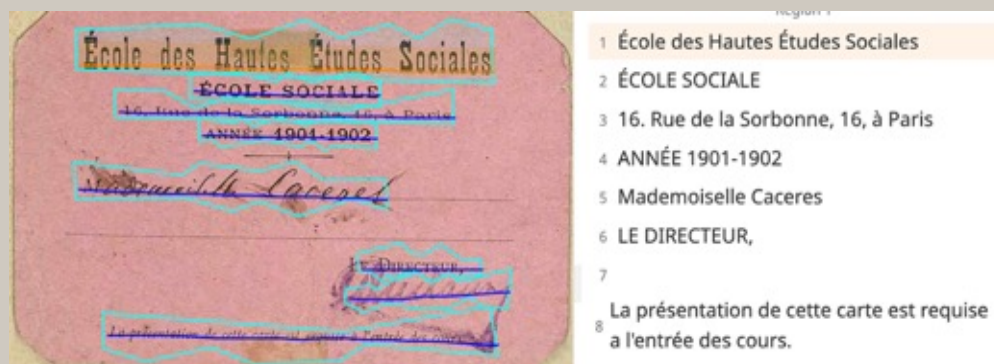


Fig. 2. Transcripción del carné de Cáceres

⁷ <https://hdl.handle.net/20.500.12534/VDSQKG>

Aplicaciones de IA generativa a partir de la obra de Zoila Cáceres

El estudio de una figura como Zoila Aurora Cáceres, su trayectoria y su vasta producción intelectual puede beneficiarse enormemente de la aplicación de IA generativa, tanto para su análisis crítico como para la difusión de su legado. Una estrategia clave sería el desarrollo de aplicaciones que utilicen modelos de IA generativa, como chats interactivos o sistemas de generación aumentada por recuperación (RAG, por sus siglas en inglés), enfocados en mitigar los sesgos y promover una representación más inclusiva y precisa de las figuras femeninas en la historia y literatura latinoamericanas. A continuación, se presentan varias posibilidades que surgen desde la creación de bases de datos centradas en la obra y los documentos de Cáceres, con el fin de construir herramientas de IA que no solo reflejen su relevancia, sino que también desafíen las limitaciones impuestas por los sistemas de IA tradicionales.

Chatbots interactivos

Un chatbot alimentado por una base de datos exhaustiva de la obra y correspondencias de Zoila Cáceres podría servir como una plataforma educativa para usuarios interesados en explorar el pensamiento feminista y la participación intelectual de mujeres en los procesos de modernización en el ámbito latinoamericano y transatlántico. Este chatbot, en lugar de reproducir los sesgos comunes de las IA entrenadas con datos predominantemente occidentales o masculinos, sería diseñado específicamente para confrontar tales sesgos al tomar como punto de partida documentos y archivos de sujetos situados en otros lugares enunciativos.

Así, el chatbot sería un modelo de IA entrenado con documentos digitalizados y anotados de Cáceres y otras escritoras de su época. El objetivo central es asegurar que las respuestas no minimicen las contribuciones de Cáceres ni las de otras mujeres, lo que permite ofrecer una experiencia interactiva en la que los usuarios puedan hacer preguntas sobre temas vinculados con el feminismo, las luchas de las mujeres por el sufragio, las asociaciones femeninas y feministas, la agencia política de las mujeres, la literatura y la teoría literaria, e incluso los procesos de modernización latinoamericana. A través del uso de datos históricamente excluidos o minimizados, el chatbot promovería una representación equitativa de las mujeres y sus documentos, y con ello crearía las condiciones de posibilidad para cuestionar los sistemas de entrenamiento de IA tradicionales, que suelen estar dominados por narrativas masculinas o eurocéntricas.

Para asegurar que el chatbot confronte activamente tales sesgos, se podría emplear un enfoque metodológico inspirado en las «fabulaciones críticas» de las que habla Saidiya Hartman, las cuales llenarían los vacíos y silencios históricos naturalizados. De este modo, lograríamos que el chatbot no solo proporcione información fáctica, sino que, también, fomente una reflexión crítica sobre las formas en que las mujeres han sido representadas o excluidas en los archivos históricos (Hartman, 2021).

Generación aumentada por recuperación para archivos

En este trabajo describimos el uso de RAG como aplicación fundamental en el estudio de la obra y trayectoria de Zoila Aurora Cáceres, lo que le permitiría a los investigadores y usuarios acceder de manera más eficiente a las múltiples facetas de la autora, de las que da cuenta su heteróclito archivo personal. RAG, o *Retrieval-Augmented Generation*, es una técnica de IA que combina la capacidad generativa de un modelo con la recuperación de información en tiempo real. A diferencia de otros modelos que generan respuestas basadas únicamente en la información que contienen internamente, RAG realiza una búsqueda activa en bases de datos para extraer información relevante al contexto de la consulta, mejorando la precisión y pertinencia de las respuestas obtenidas. En contraste con otros chatbots generativos (basados en LLM, o modelos grandes de lenguaje como el ChatGPT), los basados en la tecnología RAG combinan la capacidad de crear textos escritos que parecen haber sido realizados por una persona, con la búsqueda de información en una base de datos previamente creada.

En la figura 3 se muestra el prototipo de buscador basado en RAG, desarrollado a partir de diversos datos relacionados con la obra de Zoila Cáceres. Este prototipo, disponible en <https://buscador-zoila.streamlit.app/>, permite una interacción dinámica con la información, facilitando el acceso a datos que de otro modo podrían quedar ocultos, o ser parte de los silencios históricos que, con frecuencia, se imponen sobre las mujeres y sus obras. Al aprovechar esta tecnología, se contribuye a democratizar el conocimiento, permitiendo que

la obra y legado de Cáceres circulen ampliamente en canales digitales, promoviendo una mayor visibilidad y siendo paradigmáticos para la generación de conocimiento.

Técnicamente, el prototipo está completamente programado en Python, utilizando la biblioteca Streamlit y la API⁸ de la plataforma Vectara.⁹ Cada vez que algún usuario realice una consulta, el sistema busca automáticamente la información en la base de datos. Para ofrecer una respuesta con un estilo humano, el texto generado proviene de LLM. Así, integramos las capacidades de los LLM, las técnicas de RAG y la base de datos sobre la obra de Zoila Aurora Cáceres:

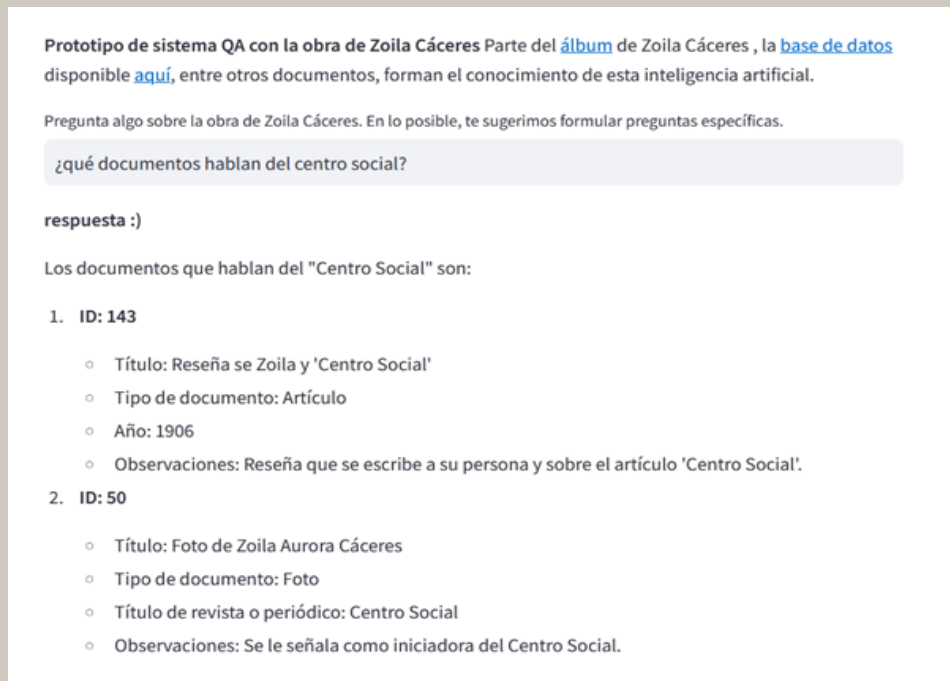


Fig. 3. Prototipo de un sistema pregunta-respuesta basado en RAG

Este prototipo de chat mediante un modelo de RAG está vinculado a la base de datos de sus escritos, correspondencias y artículos periodísticos. Cuando un usuario realiza una consulta sobre temas específicos como el feminismo, la política o la modernidad desde la perspectiva y los archivos de Cáceres, el sistema busca la información dentro del corpus de la autora y su repositorio personal y, posteriormente, genera una respuesta basada en la información recuperada. Esto permite una experiencia rica y contextualizada, en la que las respuestas no solo estén basadas en inferencias generadas por el modelo, sino también respaldadas por material original y específico del archivo de Cáceres.

Mediante este sistema, es posible contribuir a la superación de la fragmentación y disgregación que caracteriza los archivos de mujeres, en la medida en que permite recuperar, reorganizar y vincular información dispersa o desarticulada. Esto va de la mano con el aumento de su accesibilidad y el trazado de redes de diálogo entre documentos, en apariencia, heteróclitos. Esta propuesta metodológica basada en los recursos de la inteligencia artificial desafía las estructuras tradicionales de los archivos en los que predominan las voces y figuras masculinas, y los documentos que siguen formatos más convencionales (Nockels et al., 2024).

Bases de datos colaborativas para una IA más inclusiva

El desarrollo de una base de datos colaborativa, alimentada por documentos digitalizados y críticamente anotados, resulta fundamental para construir IA más inclusivas. Esta base no solo reuniría la obra de

⁸ api (Application Programming Interface o interfaz de programación de aplicaciones) es un código que permite la interacción entre la plataforma rag y el prototipo implementado mediante la librería de Python Streamlit.

⁹ <https://www.vectara.com/>

Cáceres, sino también las de otras mujeres de su tiempo que conformaban, de una u otra manera, lo que Elena Grau-Llevería (2022) ha denominado comunidades ideológicas y políticas. Para capturar las interacciones y encuentros en el tiempo y el espacio, un enfoque valioso sería emplear grafos de conocimiento (*Knowledge Graphs*), los cuales permiten modelar datos y establecer redes de relaciones diversas. Un aspecto clave de esta base de datos es su capacidad de recibir y compartir contribuciones de investigadores y usuarios, creando así un repositorio dinámico y en constante crecimiento.

Las IA generativas basadas en esta base de datos podrían estar diseñadas para producir contenido, ya sea artículos, respuestas o análisis que prioricen narrativas antihegemónicas. Estas narrativas, al integrar voces desde los márgenes (como mujeres y minorías), podrían desafiar las jerarquías tradicionales que suelen dominar los sistemas de IA entrenados con datos mayoritariamente centrados en figuras de poder. Además, la base de datos colaborativa permitiría el desarrollo de sistemas de IA adaptables que reflejen las realidades y experiencias de una variedad de grupos históricamente excluidos o minimizados en la historia, reduciendo así los sesgos presentes en las IA tradicionales (Terras, 2022).

Al permitir que los usuarios contribuyan con nuevas interpretaciones y análisis de la obra de Cáceres y sus contemporáneas, estas bases de datos se convertirían en espacios de cocreación y validación crítica, donde las IA generativas no solo reproducirían el conocimiento existente, sino que también lo enriquecerían con perspectivas diversas y transformadoras. De este modo, no solo se estarían entrenando IA generativas desde perspectivas antihegemónicas, sino que también se estaría redefiniendo la concepción de IA en general: en lugar de una IA pensada desde el centro, sería una IA concebida (y entrenada) desde los márgenes.

La IA como herramienta crítica

Suscribimos a las palabras de Terras (2022) cuando sostiene que

las transcripciones generadas por HTR permitirán búsquedas en grandes volúmenes de manuscritos (...), extendiendo el rango de las fuentes primarias al alcance de los investigadores y, a la vez, haciendo posible que la información se analice mediante minería de texto, se visualice y se estudie mediante diferentes técnicas avanzadas de Humanidades Digitales. Los aspectos transformadores de lo anterior para la investigación serán, por tanto, múltiples y proporcionarán «la base para un análisis computacional semántico, lingüístico y geoespacial avanzado de material de fuentes primarias históricas» (p. 181).

Por otro lado, de cierta manera, la IA en forma de HTR ha comenzado a brindarnos acceso a información derivada de manuscritos consultables de forma remota; a transcripciones más precisas, sobre todo en el caso de documentos con composiciones visuales complejas; y, también, a un rango más amplio de materiales institucionales y personales. Sin embargo, queda todavía pendiente una agenda que involucre preservar textos en lenguas en peligro, así como también extender el espacio para que diversas comunidades se vean representadas (Nockels et al., 2024).

La obra de Zoila Cáceres ofrece una oportunidad única para reflexionar sobre cómo la IA generativa puede servir como herramienta crítica frente a los sesgos y las estructuras de poder tradicionales. A través de aplicaciones como chatbots interactivos, sistemas RAG y bases de datos colaborativas, es posible no solo preservar el legado de Cáceres y aumentar su visibilidad, redes y relaciones, sino también promover una representación más equitativa e inclusiva en la IA. Estas herramientas, centradas en la obra de mujeres marginadas en los discursos hegemónicos, tienen el potencial de desafiar las limitaciones actuales de la IA y abrir nuevos caminos para la creación de conocimiento en el siglo XXI. Consideramos que el álbum, documento cultural multifacético, constituye un punto de partida para una reflexión teórico-crítica y metodológica, pero, también, una praxis ética y humanista que reorganice los archivos y problematice sus sesgos y jerarquías desde un enfoque crítico. Este deviene aún más significativo al tener en cuenta el trabajo de Cáceres en su lucha por derechos sociales, pero, sobre todo, políticos, para los sujetos femeninos en el Perú del temprano siglo XX (Pachas Maceda, 2024). De esta manera, los textos legibles y relacionables, al igual las plataformas interactivas, posibilitan caminos para superar el balbuceo que se impone sobre los archivos de mujeres y, de ese modo, trazar pasajes para tránsitos al futuro desde una clara conciencia ética y política sobre el pasado, las trayectorias, tradiciones y genealogías. Así, resulta viable superar las narrativas sobre la excepcionalidad que, paradójicamente, terminan justificando la disgregación dada la aparente ausencia de anclajes o «testamentos»

(Achugar, 2004, p. 30) y, al mismo tiempo, plantear rutas para los abordajes desde humanidades digitales que tengan como centro lo humano y lo humanista, contextual e ideológicamente situado, en medio de las nuevas dinámicas y hegemonías algorítmicas.

Agradecimientos

Las autoras y el autor de este artículo desean agradecer a Mariana Reyes y Dina Cornejo por su valioso apoyo en la realización de esta investigación. Sin la cuidadosa y esmerada catalogación de los documentos del álbum de Cáceres no sería posible el tratamiento computacional que el proyecto involucra.

Referencias bibliográficas

- Achugar, H. (2004). *Planetas sin boca: Escritos efímeros sobre arte, cultura y literatura*. Montevideo: Trilce.
- Amorós, C. (1991). *Hacia una crítica de la razón patriarcal*. Barcelona: Anthropos Editorial.
- Araujo, K. (2010). Configuraciones de sujeto en la modernidad latinoamericana: El caso de Perú a inicios del siglo XX. *Revista Chilena de Literatura*, (76), 5-25. <https://revistaliteratura.uchile.cl/index.php/RCL/article/view/1127>
- Derrida, J. (1995). *Mal de archivo*. México: Ediciones Siglo XXI.
- Felski, R. (1995). *The Gender of Modernity*. Boston: Harvard University Press.
- Fernández Mateo, J. (2024). Meta-literatura en el Algoriceno: explorando las fronteras de la inteligencia artificial. *La Palabra*, (48), 1-17. https://revistas.uptc.edu.co/index.php/la_palabra/article/view/17234
- Grau-Lleveria, E. (8 de noviembre de 2022). Aproximaciones feministas a la producción literaria e intelectual de las escritoras de entre siglos: hacia una comunidad ideológica y política. [Conferencia dictada en la Pontificia Universidad Católica del Perú]. Lima, Perú. <https://red.pucp.edu.pe/riel/videoteca/aproximaciones-feministas-a-la-produccion-literaria-e-intelectual-de-las-escriptoras-de-entre-siglos-hacia-la-articulacion-de-una-comunidad-ideologica-y-politica-a-cargo-de-la-doctor/>
- Gruber Garvey, E. (2012). *Writing with Scissors. American Scrapbooks from the Civil War to Harlem Renaissance*. Oxford: Oxford UP.
- Hartman, S. V. (2021). *Gallery 214: Critical Fabulations* [Audio]. *The Museum of Modern Art*. <https://www.moma.org/audio/playlist/298/4088>
- Hodel, T. (2022). Supervised and unsupervised: Approaches in machine learning for textual entities. En L. Jaillant (Ed.), *Archives, access and artificial intelligence: Working with born-digital and digitized archival collections* (pp. 157-177). Bielefeld: Bielefeld University Press.
- Morales Pino, L. A. (2020). Cuestionando el canon literario: a propósito de la reedición de «La Princesa Suma Tica», de Zoila Aurora Cáceres. En M. Blanco (Ed.), *Zoila Aurora Cáceres Moreno. La princesa Suma Tica* (pp. 8-24). Red Literaria Peruana. <https://redlitperu.com/wp-content/uploads/2022/07/zoila-aurora-caceres-moreno-la-princesa-suma-tica-28-de-junio.pdf>
- Muehlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinoecker, A., Grüning, T., Hackl, G., Haukkoara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, M., ... Zagoris, K. (2019). Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation*, 75(5), 954-976. <https://doi.org/10.1108/JD-07-2018-0114>
- Nockels, J., Gooding, P., & Terras, M. (2024). The implications of handwritten text recognition for accessing the past at scale. *Journal of Documentation*, 80(7), 148-167. <https://doi.org/10.1108/JD-09-2023-0183>
- Nockels, J., Gooding, P., Ames, S., & Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: A systematic review of Transkribus in published research. *Archival Science*, 22, 367-392. <https://doi.org/10.1007/s10502-022-09397-0>
- Pachas Maceda, S. (2024). *La historia de feminismo peruano*. Quito: Heraldos Editores.
- Segato, R. L. (2016). *La guerra contra las mujeres*. Madrid: Traficantes de Sueños.

- Terras, M. (2022). Inviting AI into the archives. En L. Jaillant (Ed.), *Archives, access and artificial intelligence: Working with born-digital and digitized archival collections* (pp. 179-204). Bielefeld: Bielefeld University Press.
- Torres-Pou, J. (1998). El e[x]terno femenino. Aspectos de la representación de la mujer en la literatura latinoamericana del siglo XIX. PPU.
- Valcárcel, A. (2001). *La memoria colectiva: Retos para el feminismo*. CEPAL. <https://www.cepal.org/es/publicaciones/5877-la-memoria-colectiva-retos-feminismo>