



EL OBSERVATORIO DE LOS ARCHIVOS UNIVERSITARIOS ESPAÑOLES: OBSERVANDO LA INTELIGENCIA ARTIFICIAL DESDE LA INTELIGENCIA NATURAL

THE OBSERVATORY OF SPANISH UNIVERSITY ARCHIVES: OBSERVING ARTIFICIAL INTELLIGENCE FROM NATURAL INTELLIGENCE

Autor:

OBSERVATORIO DE LOS ARCHIVOS UNIVERSITARIOS ESPAÑOLES¹

Resumen:

Se resumen las experiencias, las actividades y los resultados del Observatorio de los Archivos Universitarios Españoles en el ámbito de la Inteligencia Artificial durante los años 2023 y 2024.

Abstract:+

The experiences, activities and results of the Spanish University Archives Observatory in the field of Artificial Intelligence during the years 2023 and 2024 are summarized.

Palabras clave: Inteligencia Artificial; Archivos Universitarios Españoles

Keyword: Artificial Intelligence; Spanish Universities Archives

¹ El Observatorio de los Archivos Universitarios Españoles está formado por Jesús Ávila Gómez e Ingrid Gijón Ibáñez, de la Universidad Loyola Andalucía; Jorge López Conde, de la Universidad Francisco de Vitoria; Almudena Enríquez de Salamanca Gómez y Rodrigo de Luz Carretero, de la Universidad CEU San Pablo; María José Molins Lillo, de la Universidad CEU Cardenal Herrera; Pedro Olassolo Benito, de la Universidad Pontificia Comillas de Madrid; y por Pilar Gil García y Luisa María Hernández Rabilero, de la Universidad de Castilla-La Mancha.

Per considerationem sapientiae

LOS ORÍGENES²

En las XXVIII Jornadas de Archivos Universitarios (Universidad de Cantabria, 2023) la mayoría de los miembros del Observatorio de Archivos Universitarios Españoles (en adelante OAUE), que permanecía latente desde el cierre de su última observación sobre el Linked Open Data en 2021, propuso la recreación del Observatorio con el fin de observar el proceloso mundo de la Inteligencia Artificial (en adelante IA) aplicada a los procesos de gestión de documentos. Así fue, como animados por las expectativas generadas por la IA en ámbitos afines, como la gestión administrativa, se estableció como objetivo el análisis del estado de la cuestión, la problemática, la aplicación y valoración del uso de la IA en los archivos universitarios.

LOS OBJETIVOS

Aunque en términos generales, el OAUE se dedica esencialmente a la búsqueda y valoración de la información y a la localización de proyectos desarrollados en los archivos, basados en la IA, en la práctica se alinea con los objetivos principales que, dentro de *Inter PARES Trust AI*³, se engloban bajo el lema “Confío en la Inteligencia Artificial” y que quedan resumidos en:

- Identificar tecnologías de IA que puedan asumir documentos críticos y funciones de archivo desafiantes.
- Determinar las posibilidades y riesgos que puede aportar el uso de tecnologías de IA en documentos y archivos oficiales.

² Para más información sobre los orígenes del Observatorio de los Archivos Universitarios Españoles, ver: ÁVILA, J.; CONDE, J.; ENRÍQUEZ DE SALAMANCA, A.; GIL, P.; GIJÓN, I; HERNÁNDEZ, L.M.; LUZ, R.; MOLINS, M.J.; OLASSOLO, P. Adquiriendo sabiduría: del Observatorio de la Web 2.0 al Observatorio de la Inteligencia Artificial y Archivos (2010-2024). En *Boletín ANABAD*. LXXIV (2024), NÚM. 1-2, ENERO-JUNIO. MADRID. ISSN: 2794-0519 (USB) - 2444-7293 (Internet)

³ “InterPARES Trust AI (2021-2026) is a multi-national interdisciplinary project aiming to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records by forming a sustainable, ongoing partnership producing original research, training students and other highly qualified personnel (HQP), and generating a virtuous circle between academia, archival institutions, government records professionals, and industry, a feedback loop reinforcing the knowledge and capabilities of each party” Ver <https://interparestrustai.org/trust>

- Garantizar el uso eficaz de conceptos y principios de archivo en el desarrollo de aplicaciones de IA relacionadas con los archivos.
- Ensayar y difundir las aplicaciones mencionadas en el punto anterior mediante estudios de campo y aplicaciones de muestra⁴.

LAS OBSERVACIONES: EL PRIMER AÑO

A continuación, se resumen los hallazgos obtenidos en este primer año de observación, durante el curso académico 2023/2024. Si bien el período es muy breve, la tecnología de la IA se ha desarrollado vertiginosamente en este nuestro primer año de observación y ya nos ha mostrado algo más que pistas de por dónde va a transitar por nuestros archivos en un futuro, con todas sus ventajas y todos sus riesgos.

En términos generales, desde nuestro Observatorio se han constatado avances tan ilusionantes como inquietantes de la aplicación de la IA en las técnicas archivísticas, como son la clasificación, la descripción, el análisis de datos y textos o la indización. Y profundizando un poco, los avances de la IA, específicamente de la IA generativa, se trasladan a los sistemas de gestión de archivo configurando un nuevo paradigma de transformación digital de los procesos archivísticos en una triple vertiente:

- La gestión documental orientada al dato: generación de tesauros, taxonomías de datos y entidades clasificadas.
- La optimización de procesos: evaluación documental, optimización de sistemas de calidad y verificación o generación de contenidos como respuesta a procesos.
- La generación de informes, memorias y estadísticas, en general o partiendo de las propias bases de conocimiento.

⁴ International InterPARES Trust AI Project Archives 4.0: Configuring Artificial Intelligence and Secura Data Archives (2022). Disponible en https://gazete.hacettepe.edu.tr/en/news_detail/international_inter pares_trust_ai_project_archives_40_configuring_artificial_in-25

LA IA APLICADA AL PROCESO TÉCNICO

La clasificación, la descripción, el metadato y la indización

Es en el proceso técnico de los documentos donde la aplicación de la IA ha impactado más profundamente, especialmente en la clasificación automática de documentos. Las aplicaciones de IA permiten el tratamiento inteligente de los documentos para analizar el contenido de estos, comprender su contexto y asignarlos a la categoría correcta.

A su vez, incorporan funcionalidades para la identificación y extracción normalizada de información clave de los documentos de archivo, como datos y metadatos, entidades normalizadas e información en contexto para la descripción y son capaces de identificar los datos de carácter personal para su anonimización o pseudonimización.

Se observa que la tecnología sigue y va a seguir avanzando en la automatización del proceso técnico simplificando tareas y haciéndolas más eficientes, ello va a obligar a los profesionales de los archivos a redefinir su trabajo aprovechando las nuevas posibilidades que ofrece la IA.

LA IA APLICADA AL ACCESO A LOS DOCUMENTOS Y LA INFORMACIÓN

La búsqueda documental

La IA permite la búsqueda y recuperación de la información contenida en los documentos a través de lenguaje natural, por contenido semántico así como por el análisis y la extracción de resúmenes documentales.

Las tecnologías de procesado, reconocimiento, transcripción e indexación de imágenes de documentos de texto, especialmente los textos históricos manuscritos, tiene una de sus aplicaciones más sólidas y populares en *TranSkriptorium IA*⁵, *spin-off* de la Universitat Politècnica de Valencia, basada en la indexación probabilística. La mayor aportación de *TranSkriptorium* viene precisamente de la mano de este tipo de indexación para lograr la transcripción automatizada y la indexación de los textos manuscritos con un acierto de más del 70%.

⁵ <https://www.transkriptorium.com/>

Aunque la solución puede ser incorporada a cualquier institución que gestione fondos de carácter administrativo, histórico o cultural, el ejemplo más significativo y exitoso dentro del ámbito archivístico es su incorporación al motor de búsqueda de PARES⁶: “Este nuevo motor de búsqueda de PARES presenta una experiencia innovadora de búsqueda directa sobre los objetos digitales utilizando técnicas de inteligencia artificial. Se han aplicado a cinco series documentales métodos de Reconocimiento de Texto Manuscrito (HTR, siglas en inglés de Handwritten Text Recognition) con Indexación Probabilística (PrIx, siglas en inglés de Probabilistic Indexing). La búsqueda identifica la distribución de las palabras directamente sobre las imágenes digitales.”⁷

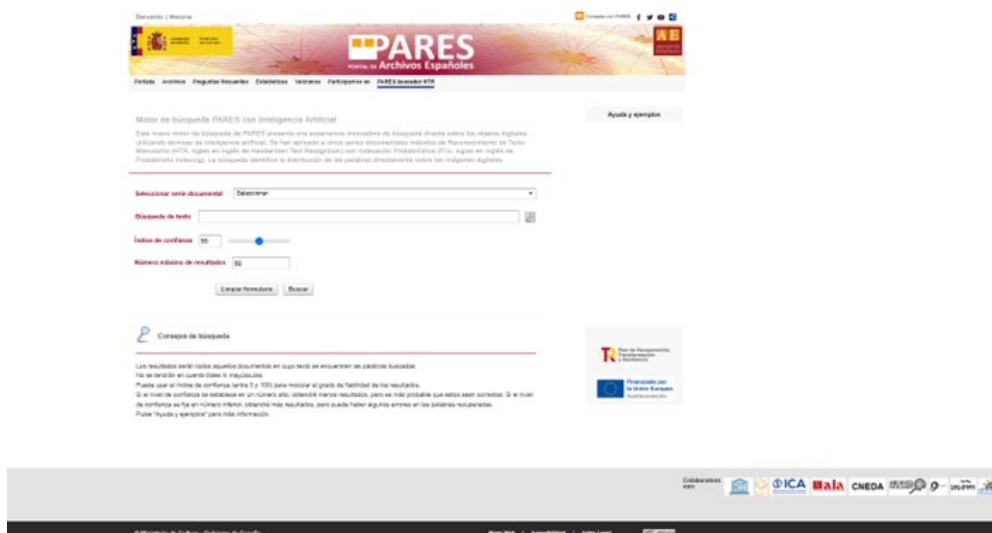


Imagen 1
Buscador HTR en PARES

Fuente: <https://pares.cultura.gob.es/pares-htr/>

La anonimización de los documentos

Una de las tareas más tediosas que pueden existir en el procedimiento de acceso a los documentos es la anonimización de los datos personales para facilitar al máximo el cumplimiento del derecho de acceso a la información, pero armonizándolo con el respeto al derecho fundamental a la protección de los datos personales.

⁶ <https://pares.cultura.gob.es/inicio.html>

⁷ <https://pares.cultura.gob.es/pares-htr/>

Se constata como la IA es una gran ayuda para el archivero en este proceso de anonimización de documentos, permitiendo detectar los datos personales y llevando a cabo su eliminación o sustitución por otros caracteres. Como ejemplos, podemos citar dos empresas que están desarrollando este tipo de productos Add4u⁸ y NYMITZ⁹

La información de referencia y documental

Otro campo donde la IA proporciona ayuda a los profesionales de los archivos es la resolución de consultas de información de referencia, en general, y del propio servicio de archivo en particular (servicios, procedimientos, horario y plazos de atención al usuario...) mediante chatbots.

Del mismo modo, los chatbots pueden resolver consultas de información sobre los fondos documentales y los servicios del archivo en general, partiendo de las propias bases de conocimiento e incorporando procesos de retroalimentación, bases que al ser elaboradas y mantenidas por los propios servicios de archivo garantizan que las respuestas están elaboradas partiendo de información de calidad.

La formación y la tutorización virtuales

En otro ámbito donde los chatbots están demostrando su utilidad es en la formación; en concreto en:

- La formación continua de los profesionales de los archivos
- La formación de usuarios de archivos, internos y externos.
- La tutorización del personal de nueva incorporación al servicio para el relevo de plantillas mediante la creación de gemelos digitales como propone la empresa ENTELGY:

⁸ <http://add4u.com>

⁹ <https://www.nymiz.com>

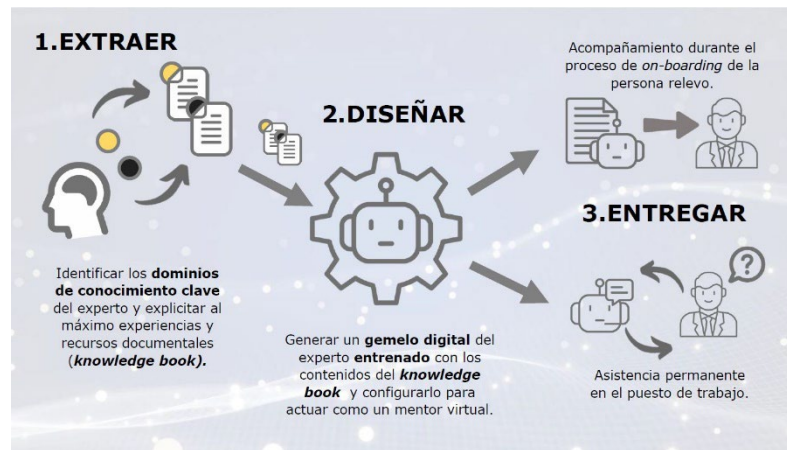


Imagen 2

“Mentores basados en la IA

Fuente: ENTELGY¹⁰

LA IA APLICADA A LA DIFUSIÓN ARCHIVÍSTICA

La generación de contenidos para las redes sociales

El uso de aplicaciones de IA en las redes sociales, plataforma esencial para la comunicación y el marketing de servicios en empresas, organizaciones e instituciones, se ha generalizado en los últimos tiempos, especialmente para la generación de contenidos: texto, gráficos, imágenes, video o música.

Herramientas como GPT-4¹¹, Copilot¹², DALL-E¹³ o Canva¹⁴ permiten crear contenidos atractivos e innovadores de manera rápida y eficiente que pueden ser utilizados para difundir productos y atraer al público objetivo en las redes sociales.

El mundo de los archivos, como ha constatado el Observatorio, no ha sido ajeno a esta tendencia, un ejemplo de ello es la campaña llevada a cabo el pasado mes de junio por varias asociaciones profesionales de archiveros y gestores documentales con motivo de la Semana de la Administración Abierta en la que se utilizaron algunas imágenes generadas por la IA.

¹⁰ Ver la presentación “Mentores basados en la IA” Disponible en <https://socinfodigital.es/2024/03/27/estrategia-nacional-de-inteligencia-artificial-avances-retos-y-casos-de-uso-reales/>

¹¹ <https://openai.com/index/gpt-4/>

¹² <https://copilot.microsoft.com/>

¹³ <https://openai.com/index/dall-e-3/>

¹⁴ <https://www.canva.com/es-es/>



Imagen 3
Campaña de la Semana del Acceso Abierto 2024
Fuente: ANABAD Federación¹⁵

No obstante, debemos ser conscientes de que la creación de contenidos por medio de la IA plantea problemas de derechos de autor y propiedad intelectual sobre la originalidad del contenido generado, por un lado y, por otro, sobre el contenido preexistente.

En principio la legislación española sobre propiedad intelectual solamente considera autor a la persona natural, en tal caso, en los contenidos generados por la IA: ¿quién poseería los derechos de estos contenidos? Es un asunto complejo que aún está por resolver.

Y, en cuanto al contenido preexistente, se plantean dos variables, en primer lugar sobre la originalidad del contenido generado por la IA, las herramientas de la IA se alimentan y entrenan con grandes volúmenes de datos y obras existentes, lo que puede llevar a la creación de contenido no completamente original, sino basado en esas obras protegidas por derechos de autor; y, en segundo lugar, con los datos y obras usados para entrenar los modelos de la IA, que a menudo incluyen contenido protegido por derechos de autor.

¹⁵ @anabad_fed

Este Observatorio recomienda citar los contenidos generados por la IA siguiendo las normas de citación, en cualquier caso, se debe incluir el nombre de la herramienta de la IA que generó el contenido, la empresa que la creó y el prompt que lo generó.

CONCLUSIONES

Como decíamos en otro lugar: “No es poco lo observado y sorprende la magnitud de la potencialidad de las herramientas y la aplicación, que ya está en marcha, de la IA en las tareas propias del profesional del archivo, especialmente desde la perspectiva -inquietante- de la confianza en la tecnología si no va de la mano de los profesionales y la ciencia.”¹⁶

El desarrollo y la aplicación de la IA en el ámbito archivístico trae consigo novedades que vuelven a demostrar la utilidad de la tecnología como herramienta que permitirá ganar tiempo en la realización de tareas rutinarias para dedicarlo a tareas intelectuales que no pueden ser resueltas por la IA, pero la IA conlleva la exigencia de la adaptación del perfil profesional y de la formación profesional, si bien esto no es nada nuevo sino la realidad que estamos viviendo desde hace años con la irrupción, el desarrollo y la evolución de las tecnologías de la información y la comunicación en nuestro día a día.

Observamos que, al igual que ya ha ocurrido con la implantación de la administración electrónica y la digitalización, los profesionales de los archivos son necesarios en la incorporación de la IA a los procesos archivísticos, pues son ellos quienes mejor los conocen y los que deben marcar el camino a seguir.

Más allá de la ortodoxia archivística y nuestros límites geográficos debemos estar atentos al desarrollo de iniciativas y proyectos internacionales como el mencionado *InterPARES Trust AI*. Y entre tanto nosotros seguiremos Observando y dando cuenta de nuestras observaciones generadas por una inteligencia nada artificial en nuestro repositorio ObservandoArchivos.

¹⁶ ÁVILA, J.; CONDE, J.; ENRÍQUEZ DE SALAMANCA, A.; GIL, P.; GIJÓN, I; HERNÁNDEZ, L.M.; LUZ, R.; MOLINS, M.J.; OLASSOLO, P. Adquiriendo sabiduría: del Observatorio de la Web 2.0 al Observatorio de la Inteligencia Artificial y Archivos (2010-2024). En *Boletín ANABAD*. LXXIV (2024), NÚM. 1-2, ENERO-JUNIO. MADRID. ISSN: 2794-0519 (USB) - 2444-7293 (Internet)

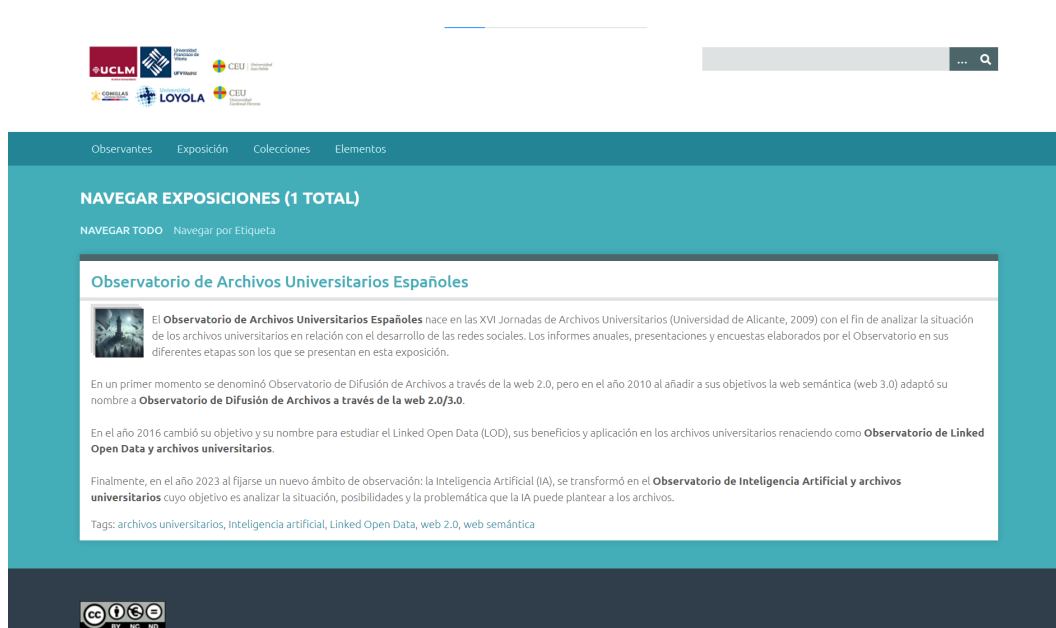


Imagen 4
Observandoarchivos.
Repositorio Open Access del Observatorio de Archivos Universitarios Españoles
Fuente: <https://observandoarchivos.omeka.net/>