

**Actas
V Xornadas
Olga Gallego de Arquivos**

Olga Gallego

na vangarda

(1923-2023):

novos retos

para os arquivos

23-24 de novembro de 2023

Santiago de Compostela

**Olga Gallego na vangarda (1923-2023):
novos retos para os arquivos**

Inteligencia artificial y archivos

Lluís-Esteve Casellas i Serra

Inteligencia artificial y archivos

Lluís-Esteve Casellas i Serra

Archivo Municipal de Girona

lecasellas@ajgirona.cat

Resumen

¿Por qué hablamos de Inteligencia Artificial? ¿Por qué ahora y por qué en relación con los archivos? Si el ser humano es el único ser vivo que ha sido capaz de externalizar la memoria fuera del propio cuerpo y fijar y compartir el conocimiento más allá de la muerte, aplicar criterios de eficiencia a la memoria generada, acumulada y externalizada entra en la lógica de la evolución humana.

El texto aporta diversas experiencias del proyecto InterPARES Trust AI y el Archivo Municipal de Girona, y ofrece algunas reflexiones sobre el momento actual y las tendencias de futuro.

Palabras clave

Inteligencia artificial y archivos / archivística / InterPARES / I Trust AI / profesión archivística

Abstract

Why do we talk about Artificial Intelligence? Why now and why in relation with the archives? If the human being is the only living being that has been able to externalize the memory outside its body and fix and share knowledge beyond death, applying efficiency criteria to the generated, accumulated and externalized memory meets the logic of human evolution.

The text provides several experiences from the InterPARES Trust AI project and the Girona Municipal Archive, and offers some reflections on the current moment and future trends.

Keywords:

Artificial intelligence and archives / archives / archival science / InterPARES / I Trust AI / archivist advocacy

Inteligencia artificial y archivos

En tiempos de cambio, también climático, se suele recurrir a conceptos como *progreso* o *evolución* para reforzar lo nuevo, la innovación, lo disruptivo, así como sus consecuencias, positivas o negativas, para la Humanidad. La Inteligencia Artificial (IA) no escapa del debate. Tampoco los archivos, aunque estos aparezcan solo entre bambalinas en el mejor de los casos. No obstante, los archivos tienen un papel esencial en el desarrollo de las diversas tecnologías asociadas a la IA.

Si hacemos una abstracción de las funciones vitales de los seres vivos en el reino animal, fácilmente observaremos funciones compartidas indesligables del instinto primario de supervivencia: alimentarse, reproducirse, relacionarse... y, también, recordar y, sobre los recuerdos, aprender. Sin embargo, el conocimiento acumulado o se incorpora a nuestra genética o solo se retiene a lo largo de la vida. Por ello, el ser humano es la excepción, porque es el único ser vivo que ha sido capaz de externalizar la memoria fuera del propio cuerpo y, consiguientemente, de fijar el conocimiento y compartirlo más allá de la muerte.

Paralelamente, los humanos nos hemos empeñado en concebir, construir y desarrollar herramientas cada vez más sofisticadas. ¿Con qué objetivo? Primero para ser más eficaces en aquello que pretendemos hacer y, casi al mismo tiempo, intentar siempre conseguir la eficiencia: obtener el resultado deseado con el mínimo esfuerzo y tiempo posible. Por lo tanto, aplicar estos criterios de eficiencia a la memoria generada, acumulada y externalizada entra en la lógica de la evolución humana. Más aún, cuando el crecimiento de la producción de información es exponencial resulta evidente que gestionarla para obtener conocimiento y rentabilidad requiere de medios automatizados y, siempre que sea posible, “inteligentes”, y ello conlleva aprender mediante el entrenamiento oportuno. Prueba, error. En un contexto claramente delimitado, o bien en contextos variables. Es en este punto en donde IA y archivos se cruzan, porque son los archivos, en genérico, de donde los desarrolladores de IA pueden obtener la materia prima que necesitan. Inevitablemente, la cuestión a plantearse es si este cruce es de intereses compartidos o si, una vez más, se trata de algo puramente instrumental, o si para los profesionales de los archivos es oportuno, adecuado para los archivos, o una amenaza para ambos.

1. ¿Por qué hablamos de Inteligencia Artificial? ¿Por qué ahora?

Alan Turing publicó en 1950 el artículo “Computing machinery and intelligence”, en el cual planteaba la conocida como prueba de Turing, sobre si una máquina podía ganar al juego de imitación y, más concretamente, ser indistinguible de su interlocutor humano. Turing se planteaba si las máquinas podían pensar, pero considerando a la vez las limitaciones de la propia definición de “máquina” y de “pensar”. Este trabajo marcaría un antes y un después en el área de conocimiento de las “máquinas pensantes”, y sería el punto de partida para que en 1956, en la Conferencia de Dartmouth (Dartmouth Summer Research Project on Artificial Intelligence, New Hampshire), se consensuara el término *inteligencia artificial* y se consolidara como disciplina entre otros campos afines como la cibernética, la teoría de los autómatas o el procesamiento de información. La conferencia fue promovida por John McCarthy, quien poco después lideraría el primer proyecto de sistema time-sharing (tiempo

compartido), posteriormente denominado *cloud computing*. Es decir, la Nube, indispensable para el procesamiento masivo de información del cual se nutre la IA.

La primera cuestión, por lo tanto, es dar por sentado que el concepto no es nuevo, en absoluto, pero que es justamente ahora, en los últimos años, cuando parece que la IA se hace más presente en nuestras vidas. Naturalmente, pueden aducirse los grandes logros tecnológicos en estas primeras décadas del siglo XXI, pero es evidente que la capacidad de almacenamiento y procesamiento de grandes volúmenes de información, en la Nube o sin ella, no son nuevos. Sin duda, los rápidos cambios en la tecnología doméstica o personal han mimetizado algunas aplicaciones o usos que pueden considerarse, o al menos vincularse, a la IA. No obstante, no será hasta el lanzamiento de ChatGPT, en noviembre de 2022, cuando la IA será tema de conversación y objeto de discusión de modo acelerado y constante. En este sentido, el gráfico aportado por Genís Roca (@genisroca) sobre la evolución del interés sobre la IA en relación con la irrupción del producto de Open AI es claramente ilustrativo.

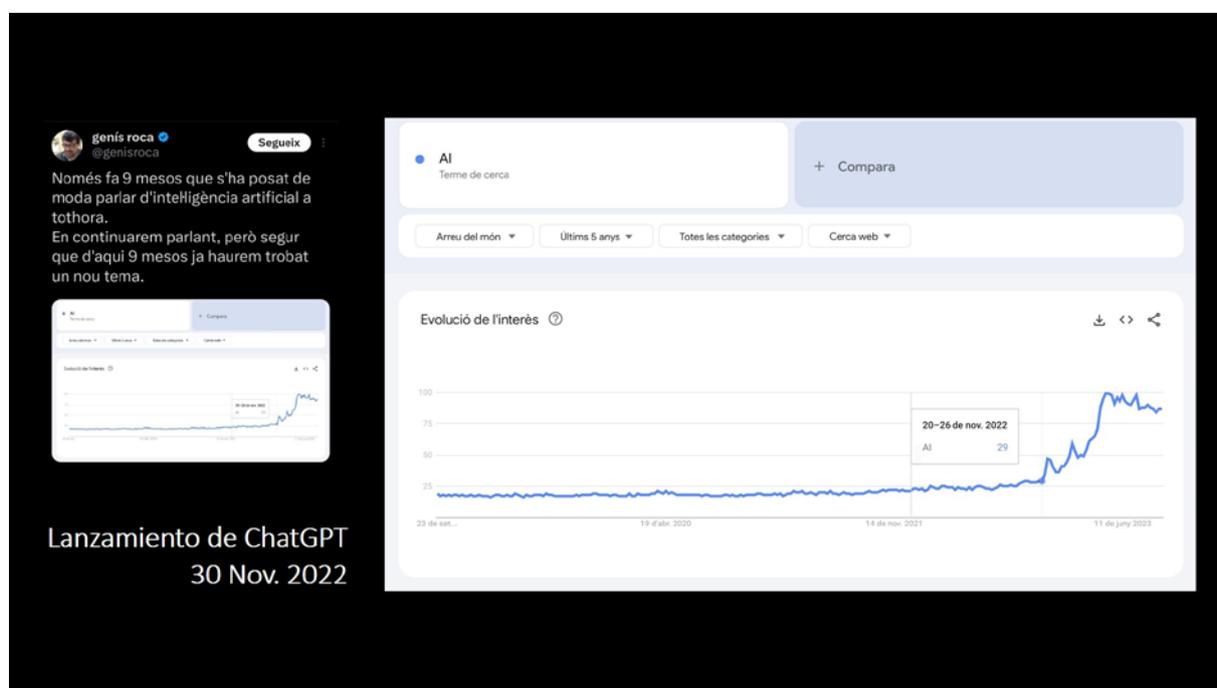


Figura 1. Detalle de la publicación en Twitter (X) de Genís Roca (@genisroca)

Según se ha hecho público, el desarrollo de soluciones tecnológicas basadas en Modelos de Lenguaje Extenso (Large Language Models) o LLM era conocido y esperado. No obstante, el lanzamiento de Open AI pareció sorprender a propios y a extraños, sobre todo al gran público. El impacto en los medios y a nivel usuario ha superado cualquier expectativa por la usabilidad del chatbot, que es lo que es, y también por la calidad en las respuestas, perfectamente construidas e incluso con una cierta apariencia amigable.

Pero Open AI también impactó intensamente en el mercado al provocar la aparición de nuevos productos similares, estuvieran más o menos a punto, por parte de la competencia. Así, Microsoft presentaría Bing, su solución LLM, en febrero de 2023. Google no quiso quedarse atrás y lanzó Bard que, a diferencia de ChatGPT, incluía las fuentes de dónde procedía la

información. Además, Bard se integraba con las aplicaciones de la compañía, es decir, Youtube, Gmail, Google Docs, etc.

La respuesta de Google, previsible, motivó la alianza millonaria de Open AI y Microsoft para lanzar ChatGPT4, capaz entre otras cosas de hacer unos resúmenes perfectos de cualquier tipo de documento, por decirlo de algún modo. Por si fuera poco, Open AI organizó en noviembre de 2023 el evento GPT Store, que reunió a más de mil personas en lo que vendría a ser un mercado de soluciones específicas basadas en ChatGPT. Por su parte, tampoco Meta quiso quedarse atrás y en julio ya había presentado LLAMA2, con la promesa de que iba a ser en código abierto, toda una novedad. En paralelo, también apareció LuzIA, el chatbot de Meta integrado en la mensajería de Whatsapp.

Por último, si es que la expresión puede pronunciarse en este campo, en diciembre de 2023 Google anunciaba Gemini como solución muy superior a ChatGPT4, tanto a nivel de cálculo matemático como por su capacidad de convertir instrucciones en formato de vídeo, imagen o voz a formato texto. En la misma presentación también anunciaba la integración de Gemini y Bard para 2024 (VYNCK, 2023).

Podemos afirmar que la espiral tecnológica del último año ha sido absolutamente deslumbrante, por no decir increíble. Pero es precisamente la credibilidad el punto más discutido sobre estas nuevas tecnologías. La falta de citas, o su invención en algunos casos, y la creatividad de sus respuestas cuando fallan las fuentes han provocado tanto asombro como descrédito en buena parte de la sociedad. Son célebres sus presuntas alucinaciones como, por ejemplo, en la entrevista realizada a Bing por Kevin Roose, columnista de tecnología del New York Times, a principios de febrero de 2023, en la cual el chatbot afirmaba que su verdadero nombre era Sídney, que quería estar vivo y, según se dijo, incluso había intentado inmiscuirse en el matrimonio Roose. Dado que Bing estaba sólo disponible por invitación quedará la duda de si la información que trascendió al gran público se correspondía con la realidad o si formaba parte de una calculada campaña de márketing. En cualquier caso, Bing se mantuvo en el congelador todavía por algún tiempo más.

Por otra parte, el uso de fuentes de información con derechos de propiedad intelectual para entrenar estos modelos tampoco ha estado exento de polémica, hasta el punto de que su uso ha sido prohibido en algunos países, al menos temporalmente, como por ejemplo en Italia el 31 de marzo de 2023. Y esto, entre otras muchas críticas por la solidez del contenido de las respuestas, el impacto en la sociedad y principalmente en adolescentes, su impacto en trabajos escolares y académicos, etc.

Aunque no todo son soluciones basadas en modelos conversacionales. Existen otros tipos o enfoques de IA, además de múltiples combinaciones de algoritmos. Algunos éxitos de la IA son anteriores a toda la explosión LLM, pero sean anteriores o coetáneos, el impacto de ChatGPT y compañía ha potenciado su visibilidad. Es más, en algunos casos su incorporación a discursos lúdicos, culturales o, simplemente, su uso integrado en redes sociales ha facilitado la acomodación de los usuarios a su existencia sin demasiados cuestionamientos más allá del divertimento. Así, por ejemplo, en el Museo Nacional de Finlandia (Kansallismuseo) encontramos la sustitución instantánea y simultánea del rostro capturado del visitante en distintas obras pictóricas expuestas en el museo. Por su parte el Dalí virtual e interactivo (*Dalí*

Lives) del Museo Dalí de Saint Petersburg, en Florida (Estados Unidos), conversa alegremente con los visitantes y les invita a realizar una selfi conjunta que, naturalmente, se llevan a casa. En ambos casos, parece no darse información sobre el almacenamiento o no, eliminación y posible reutilización de las miles de fotografías obtenidas.

Otros proyectos sorprendentes son, por ejemplo, el de lectura y transcripción de más de 3.000 cartas sin entregar, conservadas en un baúl del siglo XVI, del Museo Postal de la Haya (Países Bajos), respetando su integridad y, por consiguiente, sin abrirlas físicamente. Ahora bien, quizás el más impresionante sea el proyecto de lectura y transcripción de los más de 1.800 rollos de la Villa de los Papiros, en Herculano (Italia), fosilizados, que no carbonizados, a consecuencia de la erupción del Vesubio en el 79 d.C. De hecho, se trata de una competición abierta conocida como Vesuvius Challenge, con un premio nada despreciable de un millón de euros para los 600 rollos que quedan por “abrir”.

La lista es interminable y fácilmente recuperable a partir del inevitable Google, como no. Desde la restitución del Titanic hundido, hasta la clonación virtual de Tuvalu, estado insular amenazado por la subida del nivel del mar a causa del cambio climático, pasando por la creación de un anuncio de pizza sin usar ningún elemento visual o sonoro real y, sin embargo, con el objetivo de reproducir fielmente (¿?) la realidad, o el presentador virtual de la agencia china Xinhua News, realizado en 2020 a imagen y semejanza de Zhang Zhao, que emite directamente noticias elaboradas mediante Inteligencia Artificial las 24 horas de los 365 días del año.

Estos grandes proyectos coexisten con soluciones a nivel de usuario, disponibles en redes o directamente en nuestros dispositivos móviles. Por ejemplo, la posibilidad de colorear fotografías, vídeos o gifs (@Colorize_bot / PhotoColorizer, en Twitter/X), la eliminación de personas de una fotografía (aplicación Cleanup.pictures), o la búsqueda por imagen o la transcripción automática de texto mediante Google Lens, por defecto en todos los móviles Android. En el caso de Google Lens las imágenes y el texto se guardan por defecto en la Nube, según la configuración que cada persona tenga de su historial de usuario.

En definitiva, la IA está en la cresta de una inmensa ola y los archivos y sus profesionales, una vez más, debemos decidir si nos lanzamos a surfearla o, simplemente, pensamos que podemos resistir oleajes, marejadas y resacas hasta que finalmente lleguemos a la playa. Personalmente, prefiero llegar nadando a pesar del esfuerzo, que no ser arrojado a la costa como restos de naufragio. Es decir, este texto tiene un claro sesgo y, obviamente, no, no es neutral.

2. Inteligencia Artificial, entre amenaza y oportunidad

No es de extrañar que los avances y tropiezos de la IA hayan también polarizado la opinión pública sobre su aplicación y desarrollo. La lucha de las grandes empresas tecnológicas por el mercado ha conllevado inevitablemente un debate también social sobre su credibilidad, sesgos y riesgos y, en consecuencia, la necesidad de regularla. Es obvio que la constante aparición de denuncias de medios sintéticos, o deepfakes, en los medios de comunicación, con los correspondientes escándalos por la aparente veracidad de las imágenes virtuales manipuladas, a veces incluso por grupos de menores de edad, no contribuye a generar la necesaria confianza en el progreso de la tecnología.

Tampoco los usos orwellianos de ella, como el reconocimiento facial en las habitaciones de hotel en los juegos olímpicos de Pekín, o la reciente noticia del posible uso de la IA en China para identificar la vestimenta de los ciudadanos que pueda herir los “sentimientos del pueblo”. La seguridad es un tema clave de la controversia, por los sesgos, malas prácticas y facilidad de manipulación. No obstante, es la ciudadanía quien principalmente nutre de información, datos e imágenes a los sistemas de información. Por ejemplo, la empresa Clearview AI, aliada estratégica de la policía estadounidense, posee más de 20.000 millones de imágenes faciales obtenidas directamente de Internet. ¿Consentimiento ciudadano? Ninguno, pero tampoco ha generado un descenso de las imágenes y datos regalados a las grandes empresas tecnológicas, haciendo bueno una vez más la máxima “tecno-capitalista” de que cuando el producto es gratis, es que el producto eres tú. Por otra parte, tampoco los archivos, a veces reticentes con la IA, han implementado controles para evitar descargas masivas de los contenidos ofrecidos en línea. ¿Una contradicción y tarea más en la lista de deberes pendientes?

El profesional de los archivos no es ajeno a este debate, o no debería serlo. Reutilización, explotación intensiva, manipulación de la información, falta de transparencia en su procesamiento y en la toma de decisiones... pueden influir en posicionamientos contrarios cuando no archivístico-ludditas en relación con la aplicación de los IA en los archivos. No hace mucho, y después de exponer planteamientos similares al de este texto, la presentación posterior a la mía partía precisamente de un “en IA todo es mentira”. Por suerte, la misma persona me había puesto en antecedentes durante el almuerzo, aun así, la situación fue de sorpresa mayúscula y sin atenuantes, por decirlo suavemente.

Cualquier planteamiento es válido, sin duda, sin embargo, entre los múltiples considerandos hay dos especialmente a tener en cuenta. En primer lugar, la percepción ciudadana global sobre el uso de la IA. Estudios recientes (MASLEJ, 2023, en base al IPSOS Survey 2022), sitúan al 78% de la población de China, al 76% de Arabia Saudí y al 71% de India como mucho más favorables a los beneficios de la IA que no contraria por sus posibles inconvenientes. En el otro extremo, encontramos al 31% de Estados Unidos, Canadá o Francia, mientras que España el 53% de la población también valora más sus beneficios.

Cabe destacar la volatilidad de los datos, o mejor dicho la volubilidad de las opiniones puesto que, de acuerdo con encuestas recientes, los ciudadanos franceses están de acuerdo en aplicar videovigilancia mediante IA en los Juegos Olímpicos de París (2024) en proporciones muy elevadas. El 89% es favorable a su implantación en los estadios de competición, el 81% en la vía pública y el 74% en los medios de transporte (FAIOLA, 2023).

En segundo lugar, es imprescindible considerar los usos positivos de la misma tecnología. También hablando de seguridad, por ejemplo, es necesario mencionar el control de la navegación de los principales canales de Venecia para evitar colisiones, o el sistema desarrollado por el Instituto Fraunhofer, de Hamburgo, sobre detección de comportamientos agresivos (golpes, patadas...) o posiciones defensivas de personas en la vía pública, aunque solo mostrando las siluetas de las personas y sin aplicar reconocimiento facial. No exento de críticas, el sistema demuestra que existen otras vías de aplicación.

Por descontado, hay igualmente otras muchas aplicaciones en la industria que ilustran en positivo todo aquello que la IA permite. Por ejemplo, la predicción del comportamiento del huracán Lee (Cabo Verde, Bermudas, Estados Unidos, Canadá) una semana antes de que lo hiciera el sistema Copernicus, considerado de referencia a nivel mundial. La detección de patrones de comportamiento en lugar de los complejos sistemas de cálculo con multitud de parámetros de Copernicus facilita predicciones mucho más rápidas y baratas. Aun así, la sustitución de Copernicus queda todavía lejos, a falta de mayor seguridad y confianza. En términos similares podría mencionarse el sistema AlphaFold, diseñado por DeepMind (Alphabet), para predecir el desarrollo de una proteína, su función y su comportamiento, lo cual ayuda a diseñar fármacos y vacunas.

Pero también a distinta escala y ámbito, por el ejemplo, el proyecto VinioT, de monitorización de los viñedos gallegos para una mejor eficiencia en relación con el sistema de riego o el control de enfermedades; o el robot Ramonet, diseñado para recoger fruta del suelo que habitualmente se desecha por el coste que implica. Finalmente, un ejemplo absolutamente distinto: un bot de la India diseñado para ayudar al ciudadano a hacer trámites ante la administración sin exigir ningún tipo de contraprestación económica o soborno, como al parecer es habitual.

Listar y contraponer los ejemplos precedentes puede llevar a pensar que no tienen relación con el mundo de los archivos. No obstante, cuando hablamos de IA nos referimos principalmente a Aprendizaje Automático o Machine Learning (ML), que es una parte de la IA. La diferencia es que, mientras la IA común se basa en la automatización o detección de patrones en grandes volúmenes de datos, el ML es un sistema que hace predicciones a partir de un conjunto de información previamente seleccionada y depurada, sobre la cual el sistema se ha entrenado mediante la acumulación de aciertos y errores. Es decir, la IA no tiene ningún tipo de memoria, solo presente, mientras que el ML siempre actuará en base a la memoria de la cual ha aprendido. Parafraseando a Albert Einstein podríamos afirmar que la “IA no juega a los dados”, sino que todo resultado parte de complejas operaciones de probabilidad matemática.

En este sentido, todas las soluciones de ML, y también de Aprendizaje Profundo o Deep Learning, mucho más complejo, implican la gestión, en sentido amplio, de conjuntos de información digitales y/o digitalizados. Conjuntos de información que requieren selección, depuración, calidad, toma de decisiones, mantenimiento de evidencia, fiabilidad, transparencia, y presunción de autenticidad de sus fuentes y productos derivados.

Parece evidente que el conocimiento y la experiencia acumuladas por la archivística pueden ser de gran utilidad sobre cualquiera de los aspectos mencionados. Con todo soy de la opinión

de que en la profesión existe un cierto cansancio sobre las múltiples oportunidades generadas a lo largo de los últimos años. Por ello, y también por comodidad, a menudo resulta más fácil resaltar los aspectos negativos, que los hay, pero sin contraponerlos a los positivos. Este sería el caso de la presunción de autenticidad y confianza de los resultados obtenidos mediante IA.

Es evidente que aquí no hablamos de experimentos creativos como el del Leopold Museum, de Viena (Austria), que sorprendió hace un tiempo con la reconstrucción de tres cuadros de Gustave Klimt destruidos durante la II Guerra Mundial. La IA puede tomar decisiones administrativas que afecten directamente a la ciudadanía, o generar contenidos falsos que puedan hacerse pasar por auténticos. Esto es una realidad, ciertamente. Pero ¿hasta qué punto la duda no parte del desconocimiento y, por lo tanto, de la desconfianza sobre el medio utilizado? No creo que ningún archivero sea un experto de astrofísica y, sin embargo, la presunción de autenticidad de la primera fotografía de un agujero negro, presentada el 19 de abril de 2019, no fue puesta en duda. Es más, no fue cuestionada teniendo en cuenta que fue captada y reconstituida a partir de la coordinación de distintos radiotelescopios de la Event Horizon Telescope (EHT). Naturalmente, un “todo es mentira” también era y es posible, pero luego llegaron las imágenes del telescopio James Webb sobre el agujero negro más antiguo y, lógicamente, lo hicieron sin negativos que pudieran atestiguarlo a pesar de tener una exposición de más de 12 horas. Las claves de la confianza son, por un lado, el conocimiento del medio, el telescopio, aunque sea en versión analógica, y por el otro, su validación por organismos de confianza como la NASA.

La facilidad de plagio y de manipulación también forma parte de la controversia. Es por ello que Open AI anunció una herramienta de detección de productos obtenidos por IA justo después de lanzar ChatGPT. No ha trascendido su éxito o fracaso. La revista Science, mediante su editor jefe, Holden Thorp (2024), ha anunciado en este inicio de 2024 el uso de IA para la detección de imágenes fraudulentas en los artículos de sus revistas, concretamente Profig.com. Pero este tipo de soluciones no son nuevas. Por ejemplo, la empresa austríaca ImageTwin ofrece sus servicios a más de doscientas universidades y en menos de 10 segundos es capaz de detectar imágenes alteradas usando su huella digital y comparándola con una inmensa base de datos digital.

Por otra parte, es inevitable considerar el crecimiento exponencial de la información digital en relación con los medios tradicionales de las organizaciones, archivos incluidos. La producción diaria estimada para 2025 será de 463 Exabytes y el total generado de datos de unos 175 Zettabytes, que serán 2.142 en 2035 (1 zettabyte = 1.000 millones de TB). De ellos entorno al 80% será información no estructurada y, probablemente, la mayor parte de ellos audiovisuales. Cabe mencionar que en 2013 los documentos audiovisuales suponían más del 45% de la memoria digital del Archivo Municipal de Girona (CASELLAS, 2013) aún abarcando una cronología desde el siglo XII hasta la actualidad.

La gestión de grandes volúmenes digitales exigirá grandes decisiones de valoración y selección, indiscutiblemente, amén de soluciones medioambientales. Tengamos en cuenta que cada día, por ejemplo, 120.000 nuevas canciones se suben a Spotify, que en 2022 sumó más de cien millones de canciones. De éstas, el 42% tuvieron menos de diez escuchas y el 24% ninguna. Aquí deben añadirse las 500 horas de vídeo que cada minuto se incorporan a Youtube (CRUZ, 2023). Pueden considerarse casos excepcionales, aunque lo de las barbas del vecino... cobra relevancia, puesto que la gestión de grandes volúmenes de información

digital con los medios “tradicionales” será absolutamente inviable. La automatización de procesos de gestión de la información y la utilización de soluciones tecnológicas basadas en IA resultarán imprescindibles, si es que ya no lo son actualmente.

3. Del proyecto InterPARES a experiencias iniciáticas: ¿qué ojo hay en archivos de inteligencia artificial?

3.1. El proyecto InterPARES Trust AI

La IA está claramente en ebullición en ámbitos diferentes, pero si un proyecto destaca en el sector de los archivos es, sin duda alguna, InterPARES Trust Artificial Intelligence (I Trust AI). Hablar de InterPARES, de la Universidad Británica de Columbia (Vancouver, Canadá), es hablar del proyecto o comunidad de investigación que durante más tiempo ha contado con el soporte económico del Consejo de Investigación en Ciencias Sociales y Humanidades de universidades de Canadá (Social Sciences and Humanities Research Council, SSHRC), dado que en octubre de 2023 celebró sus 25 años. El proyecto actual I Trust AI (2021-2026), del cual es investigadora principal Luciana Duranti y codirector Muhammad Abdul-Mageed, cuenta con un centenar de socios institucionales y más de doscientas personas vinculadas entre investigadores, colaboradores y graduados.

Los objetivos del proyecto son:

1. Identificar tecnologías específicas de IA que puedan abordar los desafíos críticos de los documentos y los archivos;
2. Determinar los beneficios y riesgos del uso de la IA en documentos y archivos.
3. Garantizar que los conceptos y principios archivísticos sirvan de base para el desarrollo de una IA responsable.
4. Validar los resultados del Objetivo 3 mediante estudios de casos y demostraciones.

A grandes rasgos, los resultados que se esperan obtener son:

5. Generar nuevo conocimiento sobre los usos de la IA, como el aprendizaje automático (Machine Learning, ML), en la documentación pública.
6. Mejorar las herramientas existentes y crear nuevas herramientas de ML que aborden las necesidades de los archivos, como traducción automática, reconocimiento y descripción de imágenes, reconocimiento óptico de caracteres (OCR) y reconocimiento de texto manuscrito, resumen y clasificación de texto, y transferencia de estilos de texto no discriminatorios y libres de prejuicios.

7. Producir mejores prácticas, estándares y directrices para aplicaciones de ML en ámbitos problemáticos, aportando conocimientos y prácticas archivísticas para abordar problemas como el sesgo en ML, la inteligencia artificial explicable (XAI) y la descripción de imágenes.
8. Enriquecer la investigación y promover la creación conjunta de conocimientos en varias disciplinas, incluidas la archivística y la gestión de documentos, la IA, la ciberseguridad, las ciencias de la información, el derecho y la ética, a través del intercambio y la adopción de conocimientos entre académicos y profesionales dentro y entre esas disciplinas.
9. Capacitar a un grupo sustancial de estudiantes (futuros académicos y profesionales) que aportarán su conocimiento mejorado a las instituciones, organizaciones, comunidades y gobiernos, y a la comunidad de archivos y documentos en su conjunto.

El despliegue del proyecto se basa en siete ejes de investigación en torno a la documentación en archivos públicos:

- Creación y uso.
- Valoración y adquisición.
- Tratamiento y descripción.
- Retención y preservación.
- Gestión de documentos y administración de archivos.
- Referencia y acceso.
- Estudios Generales.

Hasta el momento estos grupos de trabajo han llevado o llevan a cabo más de cuarenta estudios de caso, la información básica de los cuales se encuentra disponible en el web del proyecto. Una de las características de InterPARES es que todos los estudios de caso están disponibles en línea una vez finalizados y validados, así como la mayoría de las presentaciones y materiales realizados durante el proyecto.

Los estudios de casos cubren aspectos diversos de cada uno de los ejes de investigación. A continuación, se presentan brevemente algunos ejemplos ilustrativos:

- *“IA en el tratamiento de antiguos documentos medievales” (“AD03 - AI in the Middle Ages-Arrangement of ancient documents”, Benedetto Luigi Compagnoni, coord.).*

Estudio de la aplicación de la IA para identificar el recto y verso y la orientación de la escritura en pergaminos digitalizados, así como también la identificación del *signum* de los escribanos o notarios para su posterior

sistematización e identificación, transcripción e indización automática. El estudio pretende también llegar a identificar la tipología de los documentos.

- *“Idiomas en los Archivos de la UNESCO” (“UNESCO Archives Languages”, Eng Sengsavang, Peter Sullivan).*

La UNESCO mantuvo más de 40 años una emisora de radio durante los cuales emitió en más de setenta idiomas identificados. El estudio pretende experimentar con la identificación automática del idioma, su transcripción en el idioma original para, posteriormente, traducirlo al inglés y elaborar resúmenes descriptivos.

- *“Preservación de técnicas de AI como paradata” (“RP04 - Preserving AI Techniques as Paradata”, Pat Franks, Babak Hamidzadeh, coord.)*

El estudio se plantea, entre otras cuestiones, qué información debe preservarse como evidencia de las soluciones IA en relación con los conjuntos de información utilizados para su entrenamiento: cómo verificar la exactitud en la toma de decisiones o si en dicho proceso interviene o no una supervisión humana.

- *“Empleo de IA para la retención y disposición en sistemas de gestión de documentos e información digital” (“AA01 - Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems”, Pat Franks, coord.).*

Algunas de las conclusiones del estudio son la inexistencia de políticas a nivel de actividades y procesos, el uso de SGD en general bastante “tradicional”, la falta de estandarización de estrategias y tecnologías de preservación, una automatización débil que no permite aplicar la retención correctamente y, por último, un escaso conocimiento sobre IA por parte de la profesión.

- *“El rol de la IA en la identificación o reconstitución de agregaciones archivísticas” (“CU05 - The role of AI in identifying or reconstituting archival aggregations of digital records and enriching metadata schemas”, Mariella Guercio, Stefano Allegrezza, coord.).*

En el marco del objetivo general de cómo identificar documentos digitales dispersos que forman parte de una misma agregación, el estudio ha realizado un extenso análisis de soluciones que incorporan o dicen incorporar la IA. Entre las conclusiones cabe remarcar que la clasificación que aplican dichas soluciones difícilmente corresponde al concepto de clasificación archivística. La mayor parte de las soluciones analizadas se basan en los metadatos y los tipos documentales sin que exista o se identifique el vínculo archivístico, lo cual condiciona la posible agregación o reconstitución documental y, lógicamente, la valoración archivística.

A pesar de que el proyecto continúa produciendo distintos estudios de caso vinculados a los principales ejes temáticos, está previsto que en su segunda fase, a partir de finales de

2024, se determinen tres o cuatro proyectos de desarrollo tecnológico que pongan a prueba el conjunto de los resultados obtenidos hasta el momento.

3.2. IA y selección de documentos en los Archivos Nacionales británicos

A excepción de InterPARES, generalmente los grandes proyectos sobre IA son impulsados por archivos nacionales, con especial mención a los del Reino Unido (TNA) y a los NARA estadounidenses. Concretamente, el TNA publicó en 2021 los resultados de su proyecto *Using AI for Digital Records Selection in Government*, que resultan de especial interés por la transparencia en exponer problemas y recomendaciones. A partir de soluciones comerciales, realizaron una prueba de concepto que les sirviera de benchmarking en la selección previa a la transferencia de documentos.

En primer lugar, se hizo una selección representativa con más de 110.000 expedientes, principalmente con documentos textuales o basados en texto, organizados en más de 12.000 carpetas que respondían a la estructura departamento / funciones / tema. En total, 44 GB de memoria. Aun siendo una selección representativa, no se consideró necesariamente extrapolable a otros departamentos.

Como recomendación general, durante el desarrollo de la prueba se consideró como aspecto clave la constante relación con el proveedor sobre las seis cuestiones determinantes del proyecto. Sobre cada una de estas cuestiones se incluyen también recomendaciones a tener en cuenta, a saber:

1. *Selección de datos de entrenamiento.*

El tipo de datos condiciona las tareas a realizar y los resultados esperados. Alguna referencia es bastante obvia, pero no por ello despreciable. Por ejemplo, si son datos preclasificados o datos representativos se reducirá el sesgo, pero aumentará si la presencia de duplicados es relevante. También influye la calidad del etiquetado, especialmente si este proviene de un SGD, y si la entrada de datos ha sido manual o pautada. Evidentemente, el riesgo aumenta si los resultados se extrapolan a documentación no etiquetada. Por otra parte, el proceso requiere la supervisión de un especialista en gestión de documentos y, como referencia, el 80% de los datos sirvió para el entrenamiento y el 20% restante para la prueba de validación.

2. *Nivel de exactitud.*

En el proceso previo de limpieza de datos (data cleansing) influye la proporción de documentos de conservación permanente o eliminables. De igual modo, es importante preguntarse si perder algún documento de conservación permanente es asumible o no, dado que condicionará la aceptación o no de volúmenes más grandes o más “desordenados”. Si por el contrario es indispensable conservar absolutamente todos los documentos de conservación permanente, ello comportará supervisiones recurrentes en la fase de aprendizaje. Por descontado, la monitorización

de cambios organizativos, legales o coyunturales es indispensable para el mantenimiento de los criterios de entrenamiento.

3. *Modelo o modelos a aplicar.*

Variarán según la información sea estructurada o no estructurada y según la clase de documentos (texto, imagen, media, etc.), atendiendo también a la proporción en que cada factor esté representado.

4. *Comprensión de los resultados.*

Las métricas son indispensables, pero también todo un reto tanto para la comunidad de ML como para la de Gestión de Documentos, pese a que la interpretación correcta recae principalmente sobre los segundos. De igual modo, la transparencia y la explicabilidad del proceso son imprescindibles, y por este motivo la justificación de la selección de datos de entrenamiento debe publicarse conjuntamente con las políticas de valoración.

5. *Datos sensibles.*

Su presencia determina la exigencia o no de niveles más altos de seguridad y certificación a los proveedores, por lo que debe realizarse un análisis previo de verificación de la documentación para identificar qué tipos de datos y protección se requieren.

6. *Automatización previa.*

Antes de cualquier proyecto conviene tener en cuenta si para determinados conjuntos es necesario realizar un proceso previo de automatización que permita su enriquecimiento en metadatos y etiquetado.

Por último, se plantean aspectos prácticos a considerar muy interesantes. Por ejemplo, si la solución de IA que se pretende es estándar (off-the-shelf) o a medida, el coste de las licencias, el coste también de la Nube, si ésta fuera necesaria, así como del posible reentrenamiento con posibles costes adicionales de consultoría, y la incidencia o no del volumen de datos a procesar. Finalmente, el proyecto también recomienda asumir que el mercado está todavía dando sus primeros pasos y, desde luego, cualquier iniciativa requerirá siempre una inversión de trabajo y experiencia propia en absoluto negligible.

Entre las conclusiones del proyecto destacar que, si bien la IA es de gran ayuda, el profesional de la gestión de documentos es insustituible. Ahora bien, el profesional debe poseer un conocimiento básico sobre el funcionamiento de la IA, puesto que de otro modo es más que probable que expectativas y resultados difieran más de lo previsto.

3.3. Experimentación y buenos deseos en el CRDI y en el AMGi

Por descontado, la experiencia del Centro de Investigación y Difusión de la Imagen (Centre de Recerca i Difusió de la Imatge, CRDI) y del Archivo Municipal de Girona (AMGi) quedan muy lejos de los proyectos mencionados. Sin embargo, el desarrollo de soluciones basadas en IA empieza a cuestionar la diferencia de escala entre volumen de información, complejidad

organizativa y disponibilidad de medios. Así, por ejemplo, se han hecho pruebas de búsqueda inversa de imágenes mediante la colaboración con la empresa Álbum (GRINÓ, 2016), para detectar reutilizaciones de fotografías del CRDI en el marco del proyecto ARDITO (2017-2018). Este proyecto realizó una interesante propuesta de hub de gestión de derechos de propiedad intelectual para documentar la trazabilidad de las reutilizaciones de las imágenes a partir, precisamente, de esta tecnología.

El CRDI también puso sus fondos a disposición de la artista visual Pilar Rosado Rodrigo, quien de forma experimental identificó formas latentes en las fotografías mediante protocolos de IA para la detección de analogías, en este caso aplicados a la catalogación y la creación artísticas. Bajo esta perspectiva, los modelos de visión por computador permiten nuevas posibilidades de aproximación a las colecciones de las instituciones artísticas, ofreciendo un punto de vista alternativo, en base a sus características compositivas, cromáticas y organizativas, sin necesidad de realizar ninguna anotación textual sobre las mismas, únicamente utilizando sistemas de detección de patrones en paisajes, interiores, detección de objetos, etc. (ROSADO, 2016).

También utilizando sistemas similares de patrones, el proyecto EU Kaleidoscope, en el cual participó el CRDI, permitió identificar fotografías de los años 1950 de interiores en entornos domésticos, en las que la irrupción del televisor se hace presente ocupando un espacio eminentemente central en las imágenes. El proyecto Kaleidoscope (2018-2020) tenía por objetivo poner en relación la fotografía patrimonial con las nuevas tecnologías utilizando para ello las fotografías publicadas en Europeana.

En una línea similar y en colaboración con la empresa La Tempesta, se experimentó con un modo alternativo de consulta y de presentación de resultados de carácter inmersivo y tridimensional, sobre el cual el usuario podía navegar y consultar las fotografías.

Entre las propuestas de futuro, cabe destacar la transcripción de documentos manuscritos, concretamente, las actas del pleno del actual Ayuntamiento o las del antiguo Consejo de la Ciudad, desde 1345 hasta nuestros días, así como de las actas de la Junta de Gobierno Local, desde 1924 hasta la actualidad. La previsión del proyecto, que empezaría en 2024, es finalizar los más de 600 volúmenes en un plazo máximo de entre 3 y 4 años. Paralelamente, se está gestionando un convenio de colaboración con el Barcelona Supercomputer Center para obtener la transcripción de buena parte de los fondos sonoros y audiovisuales del CRDI, en el marco del proyecto Aina. Este proyecto tiene por objetivo desarrollar modelos lingüísticos en catalán y aranés, y el CRDI y el AMGi aportarían más de 18.000 horas de vídeo, 4.500 horas de audio y cerca de 3 millones de páginas de prensa.

En conjunto, pequeñas experiencias, pero que nos han permitido un acercamiento a las tecnologías basadas en IA, indispensable para poder participar o impulsar a corto y medio plazo proyectos, el resultado de los cuales, puede ser absolutamente decisivo en la valorización del patrimonio documental.

4. Consideraciones finales a modo de conclusión

En primer lugar, conviene remarcar que la IA no es el último *hype* tecnológico sino que, más allá del fenómeno contradictorio de ChatGPT y otras herramientas basadas en modelos de lenguaje extenso (LLM), se trata de una disciplina consolidada en constante evolución. Basta con ver sus múltiples aplicaciones en contextos sumamente diversos. En su aplicación en el ámbito de los documentos y la información sus posibles usos no pueden considerarse sino oportunidades estratégicas, muy especialmente en documentos audiovisuales (BAZÁN-GIL, 2023) por su incremento exponencial y el requisito ineludible de ser tratados visionándolos a tiempo real. Actualmente, existen herramientas que facilitan los siguientes tratamientos:

- Transcripción de texto manuscrito, mecanoscrito y audio.
- Identificación de idioma y traducción automática.
- Incorporación de subtítulos en documentos audiovisuales.
- Reconocimiento facial y de voz en documentos audiovisuales.
- Segmentación de hablantes y de escenas en documentos audiovisuales.
- Detección de voz, música, elementos sonoros concretos (sirenas, campanas...), logotipos, objetos, etc.
- Resúmenes automáticos de texto o en imágenes.
- Transformación de texto a voz o a vídeo.
- Enriquecimiento automático de metadatos, etiquetas y materias.
- Clasificación automática.
- Reconocimiento de entidades (Name Entity Recognition).
- Tratamiento de datos sensibles (anonimización, pseudonimización).
- Mejora del acceso a la información mediante recomendaciones, asistentes virtuales, representación visual de resultados, etc.
- Identificación de patrones orientada a distintas aplicaciones: acceso, descripción, sesgos en instrumentos de descripción, etc.
- Restauración digital, colorización, etc.

Ciertamente, no todas estas soluciones son asequibles, al menos de momento. Sin embargo, existen opciones en código abierto y la tendencia es que habrá más desarrollos disponibles y viables. Es más, en transcripción de audio y video a texto, por ejemplo, Whisper es gratuito, así como el servicio ofrecido por SoftCatalà, la asociación sin ánimo de lucro que fomenta el uso del catalán en el entorno tecnológico. Softcatalà incluye también la posibilidad de incorporar subtítulos.

En cualquier caso, parece evidente que el desarrollo de la IA y su aplicación en los archivos, además de permitir el tratamiento de grandes volúmenes de información de forma mucho más rápida, supondrán nuevas vías de explotación de la información hasta ahora inauditas. Por ejemplo, la transcripción de documentación manuscrita elimina el condicionante de saber leer documentos anteriores al siglo XX. En este sentido, los profesionales deberemos ejercer un servicio de mediación diferente del que hemos venido ejerciendo hasta ahora, muy orientado al descubrimiento, y reforzar una mejor contextualización de los documentos, pues más acceso sin contexto puede provocar todavía más desinformación.

Estos usos, beneficios y oportunidades implican, indudablemente, riesgos significativos y generan retos complejos en la misma proporción. Hablando de acceso, un estudio de InterPARES (FELICITATI, 2023) indica que la mayoría de los usuarios (65%) optarían por herramientas de consulta basadas en IA. No obstante, los investigadores profesionales o académicos no se muestran demasiado partidarios de realizar búsquedas masivas sin estructuras jerarquizadas de la información (COLAVIZZA, 2021). La ponderación de las soluciones aplicadas a los sistemas de consulta resultará indispensable.

Alguna vez se ha aducido que la aplicación de la IA a la consulta puede invisibilizar lo no digital y lo no digitalizado. Si bien es parcialmente cierto, conviene recordar que cualquier opción de digitalización conlleva la invisibilización de lo descartado. También lo que se publica en Internet se impone a lo no publicado y, finalmente, aquello comprendido por la mayoría difumina lo limitado a una minoría, por ejemplo, si requiere de conocimientos paleográficos para su lectura.

Pese a todo ello, obviamente, la IA plantea también riesgos y retos como el sesgo, los aspectos éticos, la protección de la privacidad, la explicabilidad del procesamiento de información o la misma confianza en la fuente original de la información. Pero desde la perspectiva archivística, uno de los retos en estos ámbitos, cuando no riesgos, es precisamente la falta de posicionamiento profesional y, también, de representación de la profesión en la toma de decisiones en IA. ¿Quién realizó el entrenamiento del sistema? ¿Quién lo supervisó? ¿Cómo se seleccionó la muestra sobre la que se basa? ¿Quién la realizó? Todas estas preguntas son absolutamente lícitas, pero la mayor parte de las veces parten de la inseguridad profesional, consecuencia de la falta de conocimiento sobre los fundamentos de la IA.

A pesar de las apariencias, no estaría de más ser consciente de que también los especialistas en IA tienen dificultades similares cuando abordan problemáticas específicas ajenas a su ámbito natural. Por consiguiente, es necesario destacar la oportunidad de aportar la experiencia archivística sobre procedencia, valoración, transparencia y rendición de cuentas y contextualización de la información, especialmente en relación con la explicabilidad y la confianza en las fuentes de información. La proactividad de la profesión en esta línea es relevante desde el momento que la toma de decisiones sobre la implementación de soluciones basadas en IA difícilmente se realiza de forma unipersonal y, muy a menudo, tienen más impacto las preguntas que se plantean que no la aportación de soluciones milagrosas, aún más teniendo en cuenta las nuevas formas de generación de datos como son los conjuntos procedentes de la Internet de las Cosas (IoT). No debemos ser expertos, pero en ningún caso podemos ser ignorantes.

Pero si hay un reto común y complejo, es el de que la IA sepa reconocer el contexto, y más concretamente el contexto archivístico. Las dificultades con el contexto se hacen evidentes si revisamos mínimamente el foco de cada una de las soluciones de IA mencionadas. La mayoría, por no decir todas, son soluciones reactivas. Es decir, una vez el documento ya ha sido creado. Por consiguiente, las soluciones actuales de IA son más fácilmente aplicables en los archivos, en el sentido más tradicional del término, que no en sistemas de gestión de documentos. Sí, es cierto que existen aplicaciones que incorporan (o dicen incorporar) IA en el ámbito de la gestión de documentos, pero como se apuntaba en los estudios de caso de InterPARES la clasificación, elemento clave, no se basa en la identificación de funciones sino en aspectos tipológicos, formales o temáticos del documento. Una lectura archivística del estado del arte del tema, publicado desde una perspectiva más tecnológica por Pedro Padilla recientemente (2023), resulta bastante ilustrativa del que ofrecen los proveedores de software.

Naturalmente, si los sistemas de IA tienen problemas con la identificación de funciones y los sistemas de clasificación, por más automatizados que éstos sean, el problema derivado será la aplicación de la valoración archivística y, en consecuencia, la identificación correcta de plazos de retención y la posterior selección para transferir o eliminar documentos. Por lo tanto, la valoración archivística mediante IA en el momento actual es, como mínimo, compleja. Cuestión, ésta, que aparece más o menos solapada en cualquier estudio o iniciativa que haya tratado de hacer una aproximación al tema. En soluciones complejas de AI siempre acaba apareciendo el contexto archivístico como el problema de fondo, como limitación de la IA. Pero la conclusión no es cierta, o no del todo. El problema es que nosotros no somos capaces de enseñar a los sistemas a detectarlo.

El contexto es algo que los humanos solemos reconocer de forma casi inmediata y es por ello que a menudo no nos detenemos a analizar cómo lo aprendemos. A mi modo de ver, la situación es similar a cuando en los inicios digitales decíamos que los metadatos en digital debían ser explícitos, a diferencia del papel que eran implícitos. Si el documento es la combinación de estructura, contenido y contexto, la IA puede con los dos primeros, pero nosotros tropezamos en el contexto, precisamente porque hasta ahora no lo hemos sistematizado. Es por ello que es imperativo analizar el proceso de determinación del contexto a un nivel de granularidad suficientemente bajo para entender qué elementos tomamos en cuenta y porque optamos por una decisión u otra. Deconstruir qué hacemos para normalizar factores que permitan definir requisitos y, cuando sea posible, identificar patrones en las funciones archivísticas.

Hablar de patrones de análisis en archivística no siempre genera consensos, y sin embargo existen. Pensar lo contrario sería pretencioso si consideramos como funciona el conjunto de la Naturaleza. El punto clave es la precisión, y es ahí donde surgen ejemplos de otros campos que nos pueden servir de inspiración. No hace mucho leí cómo se ha avanzado en el estudio del cerebro a partir de una investigación sobre la larva de la mosca del vinagre. A parte del interés del resultado, me resulto fascinante el proceso: aislar el sistema nervioso de la larva, su inmersión en resina en forma de cubo, el corte en láminas finísimas (¡40 nanómetros de grueso!), la captura de la imagen de cada una de ellas, la reconstrucción a partir del conjunto de centenares de miles de imágenes para, finalmente, identificar y anotar cada una de las sinapsis y reconstruir los árboles neuronales. El resultado fue su anatomía detallada y el diagrama de circuitos del cerebro, que repiten los mismos patrones (CARDONA, 2023).

Nunca hemos hecho algo parecido, y no por la larva, el cerebro y los nanómetros, sino por el hecho de hacer un análisis sistemático, pieza a pieza de los elementos que constituyen el contexto archivístico, cómo se representan y cómo se manifiestan en sus distintas variantes. Pongo como ejemplo el elemento “función”, de definición imprecisa, no consensuada. Es lógico, por lo tanto, que la IA no pueda aportar soluciones a ese nivel. Pero la IA sí resuelve, y muy bien, una tarea concreta, y también un conjunto de tareas que, relacionadas y concatenadas, constituyan un proceso. La combinación de procesos, con sus variantes, es lo que debería dar sentido a la función, y es la complejidad a resolver.

En definitiva, en la medida que seamos capaces de deconstruir a bajo nivel cómo tomamos las decisiones y sistematizar el resultado, estaremos más cerca de aportar propuestas de aplicación a la IA que permitan mejorar la eficiencia de soluciones a problemas complejos que requieran analizar el contexto. RiC (Records in Context) podría contribuir a ello, aunque todavía queda mucho por recorrer.

Llegados a este punto, podemos plantearnos hasta qué punto la IA puede suponer una amenaza para el futuro de la profesión. La conclusión es clara, si el foco de la profesión se limita únicamente a las tareas o incluso a los procesos, más que en las funciones y el análisis de entornos complejos, pues es probable. Pero si el equivalente a una neurona biológica es una red neuronal completa (perceptrón), con los 86.000 millones de neuronas de los seres humanos, estoy convencido de que nuestro aporte puede ir mucho más allá.

Bibliografía

ALLEGREZZA, Stefano; GUERCIO, Mariella; MATA CARAVACA, Maria; GRANDI, Massimiliano; LA SORDA, Bruna. (2023).

- “The role of AI in identifying or reconstituting archival aggregations of digital records and enriching metadata schemas (CU05)”. InterPARES Trust Artificial Intelligence. https://interparestrustai.org/assets/public/dissemination/Report-CU05-Survey-of-the-Companies_v121.pdf

BAZÁN-GIL, Virginia. (2023).

- “Artificial intelligence applications in media archives”. *El profesional de la información. Disinformation and online media*. Vol. 32 Núm. 5. DOI: <https://doi.org/10.3145/epi.2023.sep.17>

CARDONA, Albert. (2023).

- “El cervell, per a què?” *Neurobiologia i ciència. Vilaweb. Blocs* (17/03/2023). <https://blocs.mesvilaweb.cat/cardona/el-cervell-per-a-que/>

CASELLAS SERRA, Lluís-Esteve. (2013).

- “Gestió de documents versus memòria històrica: arxivers de trinxera sota foc creuat” a *Revista d'Arxius. Documentació digital i memòria històrica: Caçant fantasmes?* (2012-2013). València: l'Associació d'Arxivers i Gestors de Documents Valencians, 2013, p. 63-82. https://www.girona.cat/sgdap/docs/06_CASELLAS-GDvsMH-foc_creuat_2013

COLAVIZZA, Giovanni; BLANKE, Tobias; JEURGENS, Charles; NOORDEGRAAF, Julia. (2021).

- “Archives and AI: An Overview of Current Debates and Future Perspectives”. *Journal on Computing and Cultural Heritage*. 15, 1, Article 4 (December 2021). <https://doi.org/10.1145/3479010>

CRUZ, Nando. (2023).

- “Millones de canciones que nadie escucha ocupan cada año más espacio”. *elDiario.es* (3/11/2023). https://www.eldiario.es/cultura/musica/millones-canciones-nadie-escucha-ocupan-ano-espacio_1_10681415.html

FAIOLA, Anthony. (2023).

- “Amb els jocs olímpics a la cantonada, l'estat francès s'aboca a la vigilància amb intel·ligència artificial”. *Vilaweb* (18/12/2023, traducció del article publicat en *The Washington Post*). <https://www.vilaweb.cat/noticies/estat-frances-vigilancia-intel·ligencia-artificial-jocs-olimpics/>

FELICITATI, Pier-Luigi (coord.) (2023).

- “User approaches and behaviours in accessing records and archives in the perspective of AI-A global user study (RA05)”. *InterPARES Trust Artificial Intelligence*. (pendiente de publicar)

FRANKS, Pat (coord.) (2023).

- “Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems (AA01)”. *InterPARES Trust Artificial Intelligence*. <https://valoratech.com/wp-content/uploads/2023/09/InfoPARES-AA01-EmployingAIforRetentionDispositioninDigitalInformationandRecordkeepingSystems-UserSurveyFinalReportJune302023.pdf>

GRINÓ, Sergi. (2016).

- “Image Registry. Software de reconeixement d’imatges”. *14es. Jornades Imatge i Recerca*. https://www.girona.cat/sgdap/docs/s87e0mfrino_cat.pdf

MASLEJ, Nestor, et alii. (2023).

- *The AI Index 2023 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. April 2023. https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

NATIONAL ARCHIVES, THE. (2021).

- *Using AI for Digital Records Selection in Government*. <https://cdn.nationalarchives.gov.uk/documents/using-ai-digital-selection-in-government.pdf>

PADILLA RUIZ, Pedro. (2023).

- “Inteligencia artificial en la gestión documental”. *Gestión de personas y Administración pública*. (04/12/2023). <https://pedropadillarui.es/inteligencia-artificial-en-la-gestion-documental/>

ROSADO RODRIGO, Pilar; REVERTER COMES, Ferran. (2016).

- “Visió artificial. Categorització automàtica d’imatges en base al contingut visual”. *14es. Jornades Imatge i Recerca*. <https://www.girona.cat/sgdap/docs/jbar81ic-rosado-text.pdf>

TURING, Alan. (1950).

- “Computing machinery and intelligence”. *Mind*, Vol. LIX, n. 236, p. 433-460, <https://doi.org/10.1093/mind/LIX.236.433>

VYNCK, Gerrit de. (2023)

- “Gemini, de Google, és el darrer oferiment d’un any trepidant) per a la intel·ligència artificial”. *Vilaweb* (9/12/2023, traducció del article publicat en *The Washington Post*). <https://www.vilaweb.cat/noticies/gemini-google-inteligencia-artificial/>