

ARQUEOLOGÍA DIGITAL: EN BUSCA DEL ARCHIVO PERDIDO

MARIANA TORO NADER • [@MARIANATORON](#)

Periodista y politóloga especialista en relaciones internacionales de la Pontificia Universidad Javeriana (Bogotá) y máster en Creación Literaria de la Universitat Pompeu Fabra (Barcelona). Actualmente es la editora de la revista *Ethic* en sus versiones web e impresa y colabora con medios como *National Geographic*, *Jot Down Cultural Magazine* y *Revista Mercurio*. Ha sido editora y redactora digital en CNN en Español, coordinadora de comunicaciones en el Centro Regional para el Fomento del Libro y la Lectura en América Latina y el Caribe (Cerlalc-Unesco) y redactora internacional, de investigación y cultura en las revistas colombianas *Semana*, *Arcadia* y *Soho*.

1. Introducción: ¿quién guarda la memoria de internet?

Hace unos días estaba buscando un meme —durante más tiempo del que me gustaría confesar—, sabía exactamente qué decía, recordaba la imagen específica, pero, a pesar de mis esfuerzos, no pude dar con él; ni Google ni los *hashtags* en Instagram y en X (Twitter) sirvieron de nada. Citando otro meme: ahora solo vive en mi memoria. Esa búsqueda frustrada resume cabalmente lo que puede suceder con internet. La cantidad de información, imágenes, artículos, videos, tuits, *reels*, *posts*, es tan colosal que a veces, sencillamente, un contenido es devorado por las fauces de la red. Quizás ahí sigue, pero ¿adónde van a parar los archivos perdidos?

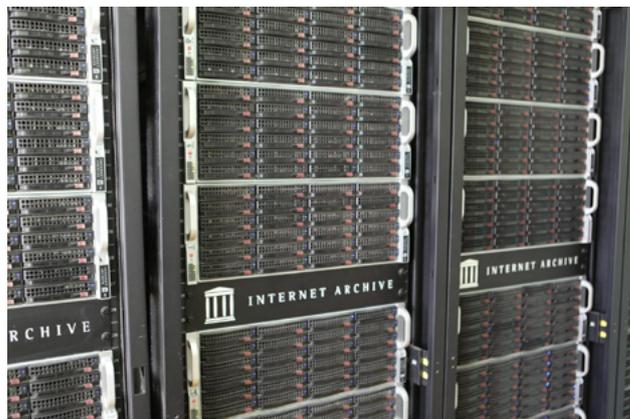
A comienzos de siglo, la [Unesco](#)¹ hizo un llamado sobre la importancia —y las dificultades— de preservar el «patrimonio digital», es decir, esos «materiales informáticos de valor perdurable dignos de ser conservados para las generaciones futuras» que nacieron como recursos meramente digitales:

Para las instituciones culturales que tienen a su cargo el acopio y la preservación del patrimonio cultural se ha convertido en un problema apremiante definir qué elementos deben conservarse para las generaciones futuras y cómo proceder en su selección y conservación. El enorme tesoro de información digital que se produce hoy día en prácticamente todas las áreas de las actividades humanas y concebida para ser consultada con computadoras podría perderse si no se elaboran técnicas y políticas específicas para su conservación.²

De cara a la inestabilidad de internet, causada por su exposición a la obsolescencia técnica y

¹ <https://es.unesco.org/themes/information-preservation/digital-heritage/background>

² Unesco, «El patrimonio digital: antecedentes». <https://es.unesco.org/themes/information-preservation/digital-heritage/background>



Servidores de Internet Archive en 2023. Crédito: Jason «Textfiles» Scott, vía Wikimedia Commons.

al deterioro físico,³ la Unesco sostiene que se debe garantizar que ese patrimonio se mantenga disponible para las generaciones actuales y venideras. Diversas plataformas buscan que sus usuarios puedan acceder y explorar versiones antiguas o capturas de contenidos de sitios web que han sido eliminados, entre las cuales están, por ejemplo, [The Wayback Machine](#), de Internet Archive,⁴ [Archive.today](#),⁵ [Time Travel](#), de Memento,⁶ o [Archive.is](#),⁷ entre otras.⁸

Bajo esa misma premisa trabajan bibliotecas, repositorios, archivos e instituciones culturales alrededor del mundo. En el caso español, el Archivo de la Web Española, creado en 2009, recolecta contenidos digitales para capturar la cultura online de España, en la que se engloban recursos como blogs, webs de servicios públicos, páginas de partidos políticos que ya no existen e incluso [memes virales](#).⁹ El objetivo de este archivo digital es que la información que era relevante en un punto de la historia española no se pierda en algún limbo del ciberespacio, sino

³ «Carta de la Unesco sobre la preservación del patrimonio digital». <https://es.unesco.org/themes/information-preservation/digital-heritage>

⁴ <https://web.archive.org/>

⁵ <https://archive.ph/>

⁶ <https://timetravel.mementoweb.org/>

⁷ <https://archive.is/>

⁸ <https://techcabal.com/2023/08/24/retrieve-lost-content-from-deleted-websites-2023/>

⁹ <https://ethic.es/2023/11/biblioteca-de-memes-para-la-posteridad/>

que, al igual que los periódicos, revistas y caricaturas de décadas pasadas, pueda ser consultada en el futuro.

Y es que, así como un error de etiquetado en una biblioteca hace que encontrar un libro específico se convierta en una tarea hercúlea, que haya errores en los metadatos de un archivo digital puede llevar a que quede extraviado en la vastedad de la web. Los metadatos son la información contextual que identifica y describe un archivo para facilitar su rastreo y acceso. En otras palabras, proporcionan *datos sobre los datos* de un contenido digital: indican su tipo, ya sea un documento, una imagen, un video o un audio; su formato, como PDF, JPG o MP3; el tamaño del archivo; su fecha de creación o modificación; su ubicación y los permisos de acceso e incluso detalles técnicos como la resolución de una imagen, la duración de un audio o la configuración de color.

Los metadatos son la información contextual que identifica y describe un archivo para facilitar su rastreo y acceso.

Que estén incorrectos puede responder a motivos muy diversos, que van desde un ataque de *malware*, manipulación intencionada, un cambio de software o errores en la migración o transferencia de archivos hasta una simple falta de estándares para la gestión uniforme o un error humano al introducir manualmente la información. Que existan metadatos sólidos y precisos es fundamental para una búsqueda de archivos correcta —además de eficiente—, pues los errores, daños o ausencia de estos afectan la ubicación de los contenidos e impiden su clasificación y, en consecuencia, el acceso a ellos y su posterior uso. Además, el etiquetado erróneo de un archivo literario, musical o audiovisual puede llevar a implicaciones graves de autoría, pues dificulta no solo la gestión y la distribución del contenido digital, sino también el reconocimiento del creador original de la obra.

2. Indiana Jones artificial

El desarrollo exponencial del *big data*, *machine learning* y aprendizaje profundo (*deep learning*) —que está logrando articular cada vez más capas de redes neuronales para llegar a mayores niveles de abstracción en el cómputo de datos— está llevando a que la IA sea utilizada en el ámbito de las industrias culturales para facilitar la gestión y distribución de los contenidos. Las llamadas Humanidades Digitales son una disciplina en ascenso que se caracteriza por la interdisciplinariedad, la colaboración y la retroalimentación entre el campo humanístico y las ciencias informáticas. En estas, a la investigación tradicional se le suma la investigación computacional para abordar nuevos enfoques en la recopilación, categorización y análisis de la información. Justamente, esta hibridez entre el ámbito cultural y el tecnológico está permitiendo que se expanda la creación de conocimiento a través de hallazgos y resultados que habrían tomado años a equipos tradicionales.

Actualmente, la IA está haciendo las veces de Indiana Jones en prácticamente todos los ámbitos de la cultura y el patrimonio. Las herramientas digitales se están usando para la visualización de colecciones en plataformas virtuales, la traducción, la interacción y el análisis de audiencias, desde la restauración arquitectónica hasta la catalogación de manuscritos, películas caseras e incluso de música enterrada en la web. Diversos softwares están siendo entrenados para excavar en la historia —ya sea antigua o relativamente reciente—, haciendo de la IA una herramienta clave para la búsqueda de tesoros patrimoniales que parecían perdidos para siempre, incluso tras haber quedado carbonizados, como es el caso de los *rollos de papiro*¹⁰ calcinados en Herculano tras la erupción del Vesubio hace dos mil años y que la tecnología está logrando salvar hoy.

¹⁰ <https://www.economist.com/science-and-technology/2023/10/18/ai-could-help-unearth-a-trove-of-lost-classical-texts>

Por ejemplo, gracias a un nuevo programa de IA llamado *Ithaca*¹¹ —desarrollado por Alphabet, de DeepMind—, que ayuda a llenar espacios y determinar su origen geográfico y su fecha aproximada de creación, los historiadores han logrado descifrar textos maltrechos provenientes de la antigua Grecia. Para lograrlo, usaron los datos de más de 70 000 textos, algunos tan antiguos que se remontan al año 700 a. C. Como explica la *Smithsonian Magazine*,¹² los historiadores del programa solo lograban restaurar los textos con el 25 % de exactitud, pero con el nuevo sistema fueron capaces de reconstruir partes de inscripciones dañadas o faltantes con una exactitud hasta del 72 %. Además, *Ithaca* podría servir como herramienta para reconstruir textos de otras civilizaciones, como la maya o la escritura cuneiforme.

Precisamente, un *equipo* de arqueólogos y científicos computacionales creó un programa de IA que ha logrado traducir las inscripciones de antiguas tablas cuneiformes mediante un modelo de traducción de aprendizaje automático neuronal (NMT, por sus siglas en inglés).¹³ La colaboración humano-máquina traduce del acadio —una lengua antigua de regiones mesopotámicas como Asiria, Isin y Babilonia— al inglés de forma automática.¹⁴ Ahora, los investigadores pueden usar las primeras versiones de la aplicación online del modelo, que recibe el nombre de *Babylonian Engine*.¹⁵ Pero las nuevas posibilidades que está abriendo la IA no se han quedado solo en la transcripción y traducción de textos antiguos.

¹¹ <https://ithaca.deepmind.com/>

¹² <https://www.smithsonianmag.com/smart-news/a-new-ai-can-help-historians-decipher-damaged-ancient-greek-texts-180979736/>

¹³ <https://www.heritagedaily.com/2023/06/archaeologists-use-artificial-intelligence-ai-to-translate-5000-year-old-cuneiform-tablets/147816>

¹⁴ <https://academic.oup.com/pnasnexus/article/2/5/pgado96/7147349?login=false>

¹⁵ <https://digitalpasts.github.io/docs/BEn.html>

3. Arqueología literaria

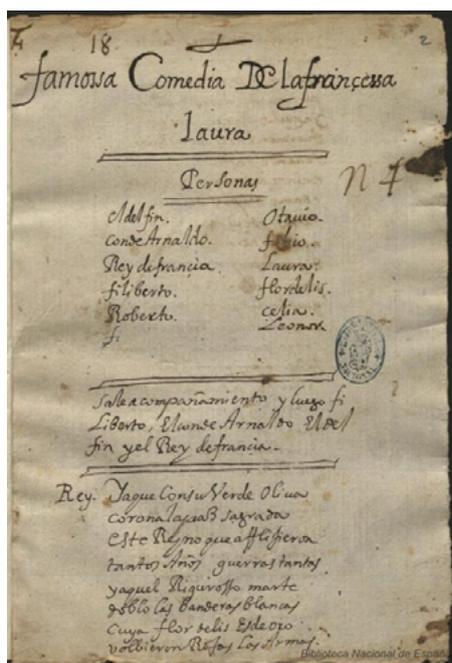
A finales de noviembre de 2023, los espectadores en Madrid pudieron ver por primera vez el estreno de *La francesa Laura*, una obra del Siglo de Oro que yacía extraviada entre los manuscritos anónimos de la Biblioteca Nacional de España (BNE). Gracias a un equipo híbrido de expertos en filología y sofisticados modelos de inteligencia artificial, pudo atribuirse a Lope de Vega,¹⁶ uno de los grandes referentes de la época. Las tecnologías utilizadas —y los procesos para atribuir la autoría del manuscrito— fueron múltiples.

Gracias a un equipo híbrido de expertos en filología y sofisticados modelos de inteligencia artificial, pudo atribuirse a Lope de Vega un manuscrito anónimo de la BNE.

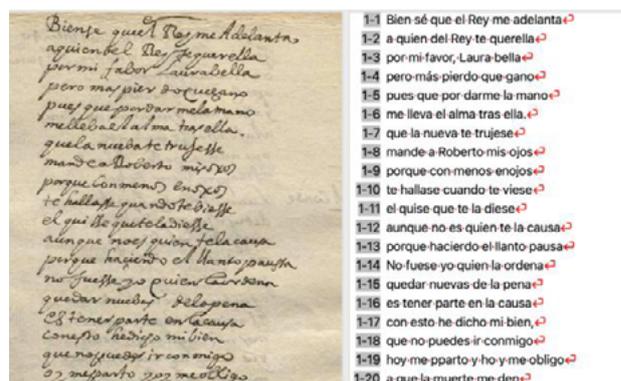
En primer lugar, los investigadores utilizaron la herramienta Transkribus, creada por la Universidad de Innsbruck, a la que entrenaron con tres millones de palabras para que pudiera reconocer cada letra de forma autónoma: unas 1300 obras del Siglo de Oro español fueron procesadas por la herramienta para convertirlas en texto digital. Tras la tarea de transcripción —que habría tomado años a un equipo tradicional—, esta se pasó por la plataforma ETSO (Estilometría aplicada al Teatro del Siglo de Oro), una base de datos compuesta por 2800 comedias de 250 dramaturgos. Por último, los investigadores de las universidades de Valladolid y Viena, la BNE y el Grupo Prolope de la Universitat Autònoma de Barcelona usaron el programa Stylo para comparar los textos introducidos en la base de datos ETSO y que clasificara las obras como escritas o no por Lope de Vega. Luego vino el trabajo filológico tradicional, en el que a través de la versificación, la ortología y el estudio de ecos lopescos —un estudio de la métrica y las estructuras léxicas y conceptuales— se llegó finalmente a la conclusión de que el manuscrito encajaba con la etapa final de la obra del poeta.

¹⁶ <https://ethic.es/2023/05/tecnologia-para-la-arqueologia-literaria/>

«Han tenido que ser los fríos ojos digitales, inasequibles a las meras apariencias, los que han puesto ante nosotros la evidencia de que no es una obra huérfana más», escribieron los investigadores Álvaro Cuéllar y Germán Vega en el *Anuario Lope de Vega*.¹⁷ Pero lo cierto es que, en efecto, la IA fue la herramienta, pero no habría sido posible sin todo el trabajo previo de entrenamiento. Como dice el propio Cuéllar: «Tienes que enseñar a la máquina a hacer algo y va a llegar un punto en el que pueda hacerlo mejor que un ser humano: al principio es un robot que hace muy mal todo, pero a la que va bien te revoluciona todo el campo». La catalogación realizada por la IA es solo una «primera pista», luego, como dice el investigador, «es conveniente que un humano experto revise que todo está bien».



Primera página del manuscrito anónimo de *La francesa Laura* (BNE. MSS/16946). Digitalización a cargo de la BNE. Disponible en <http://bdh-rd.bne.es/viewer.vm?id=0000216882>



Ejemplo de transcripción automática de unos versos de *La francesa Laura* gracias a la herramienta Transkribus y un modelo entrenado por Álvaro Cuéllar con tres millones de palabras. Disponible en <https://www.bne.es/es/noticias/inteligencia-artificial-ayuda-descubrir-obra-desconocida-lope-vega-fondos-bne>

Precisamente, el año pasado surgió el Coloso Español, un *modelo de IA*¹⁸ gratuito desarrollado en colaboración con la comunidad de Transkribus. Fue diseñado para transcribir todo un abanico de textos en español, desde manuscritos medievales hasta textos del siglo pasado, sin importar que estén escritos a mano o en máquina de escribir. Este supermodelo de transcripción podría contribuir a potenciar el campo de la arqueología literaria y ayudar a atribuir miles de obras huérfanas en España y América Latina —solo en el fondo de manuscritos de la BNE hay 85 000 obras anónimas—.

4. Cofres audiovisuales

Y es que, por su gran capacidad de procesamiento, la IA está siendo ampliamente usada en el ámbito de lo textual, pero ¿qué hay de lo visual? La creciente digitalización en el terreno cultural ha llevado incluso a que la IA pueda utilizarse en el rastreo y la catalogación de películas, cortometrajes y documentales.

Esto es a lo que se dedica, por ejemplo, Chicago Film Archives, una organización sin ánimo de lucro fundada en 2003 que identifica, recopila,

¹⁷ <https://revistes.uab.cat/anuariolopedevega/article/view/v29-cuellar-vega-garcia-luengos>

¹⁸ <https://readcoop.eu/model/coloso-espanol/>

preserva y brinda acceso a más de 40 000 películas y medios audiovisuales que representan al Medio Oeste estadounidense. Y no solamente de grandes filmes, sino que la institución adquiere continuamente colecciones de películas caseras que contribuyen a la preservación de la historia y la cultura de la región y a las que, seguramente, sería muy difícil acceder, no solo por la falta de digitalización, sino de estándares para gestionarlas.

En España, el fondo audiovisual de la Filmoteca Española¹⁹ fue lanzado en 2012 de la mano de RTVE para convertirse en el mayor fondo histórico audiovisual del país. Desde él, internautas e investigadores²⁰ pueden consultar cientos de horas de imágenes del Archivo Real de Alfonso XIII, de la Guerra Civil, primeras proyecciones del noticiario No-Do, grabaciones de las vanguardias, así como archivos internacionales como el fondo «Europa 1914», un proyecto de colaboración de las filmotecas europeas sobre la Primera Guerra Mundial.



Fotografía de la Filmoteca Española. Jorge Franganillo, 2019, CC BY 2.0. <https://creativecommons.org/licenses/by/2.0>, vía Wikimedia Commons.

Todo esto se ha logrado no solo mediante una enorme labor de digitalización, sino gracias a la

¹⁹ <https://www.rtve.es/noticias/20121220/archivo-filmoteca-no-do-mayor-fondo-historico-audiovisual/590521.shtml>

²⁰ <https://www.rtve.es/filmoteca/no-do/investigadores/>

catalogación en la web por cronología, semana, tema o búsqueda libre. Además, diversas donaciones, tanto de productoras conocidas como anónimas, han ido engrosando con los años el acervo de la Filmoteca Española. Por nombrar solo uno, en febrero de 2023 la filmoteca recibió la [donación de la mayor colección de cine documental colonial por parte de la productora Hermic Films](#).²¹ Gracias a ella, ahora se pueden consultar series rodadas en Guinea Ecuatorial y Marruecos y más de cinco mil negativos fotográficos en soporte nitrato grabados en 1944.

Las iniciativas también se han expandido en el ámbito local. Por ejemplo, el proyecto gallego [Cinema Náufrago](#)²² está rescatando videos en Super-8 de las Islas Atlánticas, con el objetivo de preservar el patrimonio audiovisual filmado por la población local en los años cincuenta, sesenta y setenta del siglo xx. A través de un proceso de digitalización y etiquetado, estas películas domésticas de valor cultural, que estaban «olvidadas ya en cajones y trasteros y en alto riesgo de extinción», según dicen los responsables del proyecto,²³ podrán usarse tanto para la investigación como para otros usos artísticos.

Y es que, como afirma la directora de la Fundación Patrimonio Fílmico Colombiano, Alexandra Falla:

La salvaguarda de la memoria audiovisual es fundamental para la identidad cultural de una nación. Preservar, restaurar y circular estos acervos históricos implica no solo un punto de partida donde conceptualmente esté claro su necesidad y contexto, sino que también conlleva procesos técnicos altamente especializados, dentro de los cuales los avances tecnológicos han jugado un papel muy

²¹ <https://www.europapress.es/cultura/cine-00128/noticia-filmoteca-espanola-recibe-donacion-mayor-coleccion-cine-colonial-20230201131043.html>

²² <https://cinemanaufrago.gal/>

²³ <https://www.lespanol.com/treintayseis/articulos/cultura/cinema-naufrago-el-proyecto-gallego-que-rescata-videos-en-super-8-de-las-islas-atlanticas>

importante. En este sentido, los archivos audiovisuales no pueden estar al margen de los escenarios y análisis actuales frente a temas como la inteligencia artificial y las aplicaciones que esta tiene en el campo del trabajo de preservación, conservación y restauración del patrimonio audiovisual.²⁴

Así, las herramientas tecnológicas se han convertido en una excelente ayuda para la recopilación y la gestión del contenido audiovisual, permitiendo que miles de usuarios accedan a películas a las que quizá nunca habrían podido llegar por sus propios medios. Y lo mismo sucede en materia sonora.

A mediados de 2023, un Proyecto de la Comisión Europea logró recuperar y catalogar miles de cantos gregorianos olvidados desde hacía un milenio.

5. Sonido bajo tierra

En 2019, los internautas recibieron la noticia de que, debido a un fallo en la migración de datos del servidor, todo el material multimedia subido a la red social MySpace entre 2003 y 2015 había desaparecido para siempre. Pero Internet Archive²⁵ —dedicada a almacenar y digitalizar música, películas, aplicaciones, libros y videojuegos, especialmente cuando estos desaparecen de su alojamiento original— hizo lo que en un principio se creía imposible. La organización sin ánimo de lucro logró recuperar casi 500 000 canciones que se consideraban perdidas. Entre ellas, la colección de una década de canciones de bandas y solistas de metal, emo, punk o música alternativa que, al no tener contratos con disqueras, usaron la plataforma de MySpace para compartir sus demos musicales. Actualmente, [Internet Archive](https://archive.org/details/myspace-Music?tab=collection)²⁶

²⁴ <https://patrimoniofilmico.org.co/inteligencia-artificial-en-el-manejo-de-archivos/>

²⁵ <https://archive.org/details/myspace-Music?tab=collection>

²⁶ <https://archive.org/about/>

guarda 14,7 millones de grabaciones de audio, incluidos 240 000 conciertos en vivo.

Pero la tecnología no solo sirve para excavar audios perdidos en alguna de las cuerdas del universo digital. A mediados de 2023, un proyecto de la Comisión Europea encabezado por investigadores de la Universidad de Jaén logró recuperar y catalogar miles de cantos gregorianos²⁷ que estaban olvidados desde hacía un milenio. Tras ser entrenada con 127 000 cantos, la IA desarrollada por el proyecto «Repertorium» pudo recobrar y digitalizar 4000 archivos musicales de la abadía de San Pedro de Solesmes. Gracias a ello, el proyecto ofrecerá no solo conciertos y óperas, sino también la tecnología en código abierto para que pueda ser usada en otras modalidades.

6. Hacia dónde apunta la brújula

La esencia misma de la IA es tomar enormes cantidades de datos para procesarlos en tiempo récord y posibilitar su clasificación y análisis. La aplicación de sus herramientas «puede transformar el análisis de colecciones digitales a través de un ciclo virtuoso de interacción entre las personas expertas y las metodologías computacionales», afirma Luis Miguel García Velázquez.²⁸ En efecto, las nuevas tecnologías han creado nuevas perspectivas de usabilidad e interacción con los objetos culturales, especialmente en materia de digitalización, consulta, procesamiento de texto, imagen, video y audio y sistematización del acervo cultural.

Las tendencias apuntan a un fortalecimiento de los buscadores y los algoritmos de clasificación automática, además de la sofisticación del

²⁷ <https://elpais.com/cultura/2023-07-22/inteligencia-artificial-para-recuperar-4000-cantos-gregorianos-olvidados-desde-hace-mil-anos.html>

²⁸ Luis Miguel García Velázquez, «Inteligencia artificial y patrimonio cultural: una aproximación desde las Humanidades Digitales», *Dicere*, n.º 4, julio-diciembre de 2023, pp. 180-191.

procesamiento de lenguaje natural (NLP), los sistemas de reconocimiento óptico, la simulación computarizada y el análisis de audiencias, entre otras proyecciones que ya están mostrando sus primeros visos.

A través de tecnologías conversacionales, los motores de búsqueda están cambiando la forma en que los usuarios interactúan con ellos. En los últimos tiempos, se han puesto a disposición de los internautas herramientas como ChatGPT, [You](https://you.com/),²⁹ [Bard](https://bard.google.com/?hl=es)³⁰ o [Perplexity.ai](https://www.perplexity.ai/)³¹; así, los chatbots podrán hacer más fácil y accesible la búsqueda en internet. En el futuro próximo, la optimización de las bases de datos con tecnologías de IA y la mejora en las técnicas de búsqueda llevarán no solo a una mayor interacción con los usuarios, sino también a la personalización de los resultados y a la búsqueda predictiva. Los metabuscadores no solo entenderán mejor qué necesita el usuario —por lo general sabemos qué queremos encontrar, pero no cómo buscarlo—, sino que además serán más eficientes a la hora de excavar entre los millones de archivos digitales disponibles, incluso esos que permanecían ocultos.

El uso de la IA facilitará los procesos de identificación, descripción y categorización de los contenidos, reduciendo también los obstáculos en la búsqueda de archivos.

Los avances en el procesamiento del lenguaje natural harán posibles las consultas de texto cada vez más complejas, así como las búsquedas semánticas o conceptuales precisas. Actualmente, los investigadores ya están avanzando en la detección caligráfica y la datación de manuscritos, lo cual posibilitará saber de qué autor es el trazo o ubicar una obra literaria en un momento concreto de la historia. Cuéllar lo compara con una prueba de carbono 14 textual: a través de análisis de palabras y estilo se llegará a situar los textos rápidamente en un momento dado o en

²⁹ <https://you.com/>.

³⁰ <https://bard.google.com/?hl=es>

³¹ <https://www.perplexity.ai/>.

otro. Además, el avance del reconocimiento óptico de caracteres (OCR) hará cada vez más fácil y precisa la digitalización de documentos físicos, lo cual facilita no solo el acceso y el análisis, sino también la preservación.

El robustecimiento de los algoritmos de clasificación automática permitirá la organización cada vez más eficiente de ingentes cantidades de material. El uso de la IA facilitará los procesos de identificación, descripción y categorización de los contenidos, reduciendo también los obstáculos en la búsqueda de archivos. Además, tener algoritmos cada vez más eficaces permitirá no solo el análisis automático, sino también la detección y corrección de metadatos erróneos. Dice García Velázquez que «otras formas de enriquecer metadatos se apoyan en la implementación de tecnologías como el reconocimiento de entidades nombradas (NER, por sus siglas en inglés), que permite identificar palabras que designan personas, organizaciones, lugares, tiempos o medidas en documentos de texto, para ubicarlas dentro de una clasificación preestablecida». La adición de metadatos relevantes y de calidad permitirá que organizaciones como la Asociación Europea —el mayor buscador de elementos culturales en Europa— faciliten el acceso a los datos abiertos de millones de objetos digitales.

Además, la democratización de las colecciones digitales ayudará a hacer frente a la inequidad en el acceso. El aumento de los ejercicios colectivos de clasificación y curaduría podrá llevar a que, por ejemplo, las comunidades de usuarios asignen etiquetas colaborativas y faciliten la búsqueda de contenidos según los atributos que consideren más relevantes.³²

Por otro lado, a pesar de que uno de los mayores desafíos que plantea la IA radica en el respeto de la propiedad intelectual, las herramientas tecnológicas podrían ser utilizadas precisamente para detectar el uso indebido de contenidos sin permiso o que violan las licencias de autor. Los

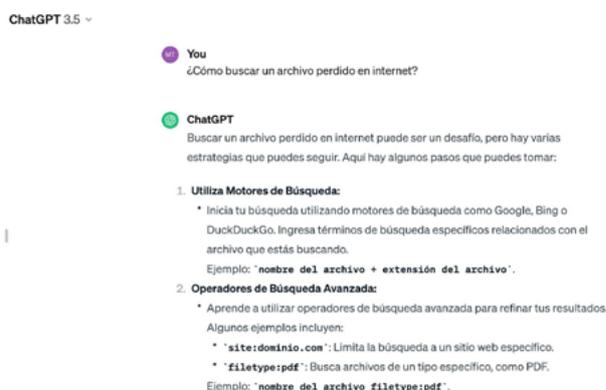
³² Luis Miguel García Velázquez, art. cit.

algoritmos de detección de plagio más sofisticados podrán detectar con facilidad si hay semejanzas no autorizadas en los textos, las películas o los elementos sonoros, cosa que contribuye al respeto de los derechos patrimoniales y la correcta atribución de los creadores.

De hecho, los artistas y los gestores culturales también podrían echar mano de herramientas digitales para añadir valor a su labor profesional y, como expone [Nuria Lloret](#),³³ prever cómo funcionarán sus productos culturales entre su audiencia. La IA puede ser usada para entender mejor a los espectadores —por ejemplo, a través de algoritmos de reconocimiento y análisis de gestos y emociones—, o para el diseño de propuestas personalizadas y la simulación de resultados de forma computarizada.

La IA debe desarrollarse como una herramienta para potenciar la creatividad humana y fomentar el consumo cultural. Ahí entra justamente el rol que jugarán —y están jugando ya— las instituciones culturales en el uso de tecnologías digitales y la conservación del patrimonio digital. Los repositorios digitales, las bibliotecas virtuales y las organizaciones de agregación de datos serán lugares de almacenamiento que marcarán la pauta en cuanto a los estándares de catalogación y preservación, además de posibilitar eventuales experiencias inmersivas en un posible «metaverso cultural» en el que se pueda acceder a gemelos digitales de los objetos culturales. Su papel es clave para formalizar buenas prácticas para crear redes de navegación en el contenido

cultural,³⁴ así como para el monitoreo de los archivos digitales que garantice su integridad en el largo plazo.



«Puede ser un desafío». Respuesta de ChatGPT 3,5 a la pregunta de cómo buscar un archivo perdido en internet.

Las nuevas tecnologías están revolucionando la cadena de valor en la creación y el consumo de los contenidos culturales, pero también pueden llevar a grandes avances en materia de accesibilidad, formación de nuevos públicos y alcance. Por ejemplo, un adecuado desarrollo de las tecnologías de traducción automática podría facilitar el acceso a objetos culturales en otras lenguas y fomentar la diversidad. Ahora bien, la principal perspectiva es que el desarrollo de los algoritmos de aprendizaje profundo y por transferencia podrá llevar a que las herramientas tecnológicas provenientes de otros ámbitos u otras investigaciones puedan aplicarse en campos diversos, fomentando así la creación de conocimiento en las industrias culturales.

7. Conclusión: el poder excavatorio de la IA

La IA evoluciona por minutos. Las herramientas tecnológicas permiten la relectura (¿la reescritura?) arqueológica y antropológica de nuestro patrimonio cultural, tome las formas que tome: ya sean inscripciones en lenguas antiguas,

33 Nuria Lloret Romero, «La aplicación de la inteligencia artificial en las industrias culturales y su aportación al desarrollo de nuevas formas de creación y gestión de la cultura», en Clara Román y José Javier González (eds.), *Informe sobre el estado de la cultura en España 2023. La presencia cultural de España en Europa*, Fundación Alternativas, 2023, pp. 147-155. https://fundacionalternativas.org/wp-content/uploads/2023/11/informe_cultura_2023_web.pdf

34 [https://www.europarl.europa.eu/thinktank/en/document/IPOL_BRI\(2020\)629220](https://www.europarl.europa.eu/thinktank/en/document/IPOL_BRI(2020)629220)

manuscritos anónimos, películas caseras o canciones extraviadas en redes sociales extintas, los algoritmos están descifrando los códigos y marcando el camino para la gestión cultural del futuro. Pero su poder para excavar en el pasado —lejano o reciente— es también lo que plantea cuestionamientos de cara a los próximos años.

En abril de 2023, el Caffè Letterario de la Via Ostiense en Roma fue la sede de *ChatXBT: The Lost Archives*, la primera exhibición de arte curado por IA, en la que estudiantes de la John Cabot University seleccionaron fotos *vintage* que luego fueron mejoradas y acompañadas por una historia creada por la IA. Su idea era observar las percepciones de los visitantes a través de esos «archivos perdidos» *reinventados* por una historia artificial y cuestionar cómo la IA puede afectar la integridad del arte y la creatividad.

Y es que, en los últimos años, el mundo creativo ha visto crecer las dudas de cara al futuro, debido al desarrollo intempestivo de la IA. Su poder nanosegúndico para crear textos, imágenes, videos, canciones y hasta «obras de arte» ha planteado un debate que se juega no solo en el ámbito cultural, sino también en el filosófico —y ético—.

Las redes neuronales de generación de lenguaje que funcionan a través del aprendizaje profundo serán cada vez más sofisticadas: ChatGPT 4 ya puede escribir literariamente y describir qué hay en una imagen; DALL-E y Stable Diffusion están mejorando sus codificadores de texto para interpretar las instrucciones de los usuarios; el AI Film Festival está premiando los cortometrajes creados por inteligencia artificial;³⁵ Midjourney ha ganado galardones de ilustración.

Al tiempo que dibuja nuevas fronteras culturales y creativas, la evolución vertiginosa de la IA

³⁵ <https://www.infobae.com/america/tecno/2022/12/14/las-razones-de-guillermo-del-toro-para-criticar-las-peliculas-hechas-con-inteligencia-artificial/>.

plantea difíciles debates éticos sobre la creatividad, la privacidad y la propiedad intelectual. Solo hay que ver las noticias del año pasado. En 2023, un juez federal falló en contra de Internet Archive —el salvador de los miles de archivos de audio de MySpace— por escanear y prestar libros electrónicos como si se tratara de una biblioteca;³⁶ directivos de grupos editoriales afirmaron no tener ningún medio de corroborar que un texto no haya sido escrito por IA, mientras que cientos de autores, entre los que se encuentran Margaret Atwood y George R. R. Martin, han exigido a las compañías de IA que dejen de usar sus obras sin su consentimiento para entrenar sus herramientas;³⁷ y recientemente se sumó una demanda de *The New York Times* contra OpenAI por infringir los derechos de autor al usar sin permiso sus artículos para entrenar sus modelos de lenguaje.

En los últimos años, el mundo creativo ha empezado a cuestionar cómo la IA puede afectar la integridad del arte y la creatividad de cara al futuro.

Si se crea una canción con IA, que ha registrado los datos de miles de artistas, ¿quién es el autor? Si una IA es entrenada con cientos de miles de obras, ¿quién recibe el pago por las licencias y los derechos de un texto, una película, una canción? Los cuestionamientos se ramifican, incluso llegando al punto de poner sobre la mesa la posibilidad de que la IA acabe con múltiples profesiones como traductores, *copywriters*, ilustradores o diseñadores gráficos, entre otros. ¿Cómo hacer que las nuevas tecnologías sirvan a los creadores en vez de perjudicarlos?

Como en toda herramienta, lo «bueno» o lo «malo» radica en el uso que se haga de ella.

³⁶ <https://www.theverge.com/2023/3/24/23655804/internet-archive-hatchette-publisher-ebook-library-lawsuit>

³⁷ <https://www.aljazeera.com/economy/2023/7/19/thousands-of-writers-demand-ai-stop-using-work-without-permission>

Y, como en toda irrupción tecnológica, podría haber espacios para el optimismo, sobre todo si se piensa en la IA como un apoyo más que como una competencia, como una habilitadora y no como una sustituta.

En el tema que nos compete, una IA bien entrenada podría ser una buena Indiana Jones, pero, al igual que cualquier creación humana, no es infalible (¿queremos que lo sea?). Por ejemplo, la clasicista Eleanor Dickey, de la Universidad de Reading, dijo a *The Verge*³⁸ sobre el software Ithaca de Deep Mind que la gente debe tener en cuenta que esa IA obtiene resultados incorrectos «aproximadamente un tercio de las veces». Este caso demuestra que no es recomendable simplemente entregarse a los hallazgos de una máquina, pues, además de que para tener resultados óptimos necesita de ingentes cantidades de datos y entrenamiento, también carga con los sesgos de su desarrollador.

«La IA no es una tecnología capaz de examinar un conjunto de datos automáticamente sin intervención humana en ninguno de los pasos del proceso. La figura del académico es fundamental para revisar los resultados producidos por el software con el fin de modificar, completar o ampliar los resultados», sostienen Diego Chapinal Heras y Carlos Díaz Sánchez.³⁹

En suma, necesitamos que, en medio del extenso mapa de internet, las tecnologías de IA cumplan su papel de recursos y funjan como asistentes para localizar, reconocer, catalogar, atribuir y preservar los tesoros de relevancia histórica. Que den pistas en la búsqueda de los archivos

³⁸ <https://www.theverge.com/2022/3/9/22968773/ai-machine-learning-ancient-inscriptions-texts-deepmind-ithaca-model>

³⁹ Diego Chapinal Heras y Carlos Díaz Sánchez, «A review of AI applications in Human Sciences research», *Digital Applications in Archaeology and Cultural Heritage*, vol. 30, septiembre de 2023. Disponible en https://repositorio.uam.es/bitstream/handle/10486/709204/review_chapinal_DAACH_2023.pdf?sequence=1&isAllowed=y

perdidos u ocultos, que sirvan de aliadas y facilitadoras para los creadores, artistas y gestores culturales. Que contribuyan a preservar la memoria del mundo, a conservar el legado cultural. Y que, en medio de la aceleración y del vértigo actuales, ayuden a ahorrar tiempo y optimizar procesos, claro, pero también permitan que la creación —antigua y reciente— perdure, que no se convierta en polvo.

Referencias

«A New A.I. Can Help Historians Decipher Damaged Ancient Greek Texts», *Smithsonian Magazine*, 16 de marzo de 2022. <https://www.smithsonianmag.com/smart-news/a-new-ai-can-help-historians-decipher-damaged-ancient-greek-texts-180979736/>.

«AI could help unearth a trove of lost classical texts», *The Economist*, 18 de octubre de 2023. <https://www.economist.com/science-and-technology/2023/10/18/ai-could-help-unearth-a-trove-of-lost-classical-texts>

«AI reads text from ancient Herculaneum scroll for the first time», *Nature*, 13 de octubre de 2023. <https://www.nature.com/articles/d41586-023-03212-1>

«Archaeologists use artificial intelligence (AI) to translate 5,000-year-old cuneiform tablets». <https://www.heritagedaily.com/2023/06/archaeologists-use-artificial-intelligence-ai-to-translate-5000-year-old-cuneiform-tablets/147816>

Biblioteca Digital Hispánica. <https://www.bne.es/es/catalogos/biblioteca-digital-hispanica>

Biblioteca Nacional de España, *La inteligencia artificial ayuda a descubrir una obra desconocida de Lope de Vega en los fondos de la BNE*. Consultado el 9 de diciembre en <https://www.bne.es/es/noticias/inteligencia-artificial-ayuda-descubrir-obra-desconocida-lope-vega-fondos-bne>

Caramiaux, Baptiste, «Research for CULT Committee - The Use of Artificial Intelligence in the Cultural and Creative Sectors», Policy Department for Structural and Cohesion Policies, European Parliament, 2020. [https://www.europarl.europa.eu/thinktank/en/document/IPOLE_BRI\(2020\)629220](https://www.europarl.europa.eu/thinktank/en/document/IPOLE_BRI(2020)629220)

Chapinal Heras, Diego, y Carlos Díaz Sánchez, «A review of AI applications in Human Sciences research», *Digital Applications in Archaeology and Cultural Heritage*, vol. 30, 2023. https://repositorio.uam.es/bitstream/handle/10486/709204/review_chapinal_DAACH_2023.pdf?sequence=1&isAllowed=y

Cinema Náufrago. Consultado el 26 de diciembre de 2023. <https://cinemanaufrago.gal/>

Coloso Español. Free Public AI Model for Handwritten Text Recognition with Transkribus. <https://readcoop.eu/model/coloso-espanol/>

Cuellar, Álvaro, y Germán Vega, «La francesa Laura. El hallazgo de una nueva comedia del Lope de Vega último», *Anuario Lope de Vega*, vol. 29, 2023. <https://revistes.uab.cat/anuariolopedevega/article/view/v29-cuellar-vega-garcia-luengos>

Falla, Alexandra, «¿El futuro restaura el pasado? La inteligencia artificial en el manejo de archivos», Fundación Patrimonio Fílmico Colombiano. Consultado el 27 de diciembre de 2023. <https://patrimoniofilmico.org.co/inteligencia-artificial-en-el-manejo-de-archivos/>

García Velázquez, Luis Miguel, «Inteligencia artificial y patrimonio cultural: una aproximación desde las Humanidades Digitales», *Dicere*, n.º 4, julio-diciembre de 2023, pp. 180-191.

Gutherz, Gai, Shai Gordin, Luis Sáenz, Omer Levy y Jonathan Berant, «Translating Akkadian to English with neural machine translation», *PNAS Nexus*, vol. 2, n.º 5, mayo de 2023, pgad096. <https://doi.org/10.1093/pnasnexus/pgad096>

Heredia Santana, Sofía, «UE: El Parlamento Europeo establece los desafíos de la inteligencia artificial en el contexto del patrimonio cultural y los museos», Instituto Autor, 9 de mayo de 2023. <https://www.institutoautor.com/ue-el-parlamento-europeo-establece-los-desafios-de-la-inteligencia-artificial-en-el-contexto-del-patrimonio-cultural-y-los-museos/>

«JCU Students Organize *ChatXBT: The Lost Archives*, First AI-curated Art Exhibit in Rome», John Cabot University. <https://news.johncabot.edu/2023/04/chatxبت-the-lost-archives-first-ai-curated-art-exhibit-in-rome/>

Lloret Romero, Nuria, «La aplicación de la inteligencia artificial en las industrias culturales y su aportación al desarrollo de nuevas formas de creación y gestión de la cultura», en Clara Román y José Javier González (eds.), *Informe sobre el estado de la cultura en España 2023. La presencia cultural de España en Europa*, Fundación Alternativas, 2023, pp. 147-155. https://fundacionalternativas.org/wp-content/uploads/2023/11/informe_cultura_2023_web.pdf

Makinde, Damilola, «Retrieve lost content from deleted websites 2023», *TechCabal*, 24 de agosto de 2023. <https://techcabal.com/2023/08/24/retrieve-lost-content-from-deleted-websites-2023/>

Morales, Manuel, «Inteligencia artificial para recuperar 4000 cantos gregorianos olvidados desde hace mil años», *El País*, 21 de julio de 2023. <https://elpais.com/cultura/2023-07-22/inteligencia-artificial-para-recuperar-4000-cantos-gregorianos-olvidados-desde-hace-mil-anos.html>

Pérez-Bouzada, Andrea, «Cinema Náufrago, el proyecto gallego que rescata vídeos en Super-8 de las Islas Atlánticas», *El Español*, 26 de diciembre de 2023. <https://www.lespanol.com/treintayseis/articulos/cultura/cinema-naufrago-el-proyecto-gallego-que-rescata-videos-en-super-8-de-las-islas-atlanticas>

Peters, Jay, y Sean Hollister, «The Internet Archive has lost its first fight to scan and lend e-books like a library», *The Verge*, 24 de marzo de 2023. <https://www.theverge.com/2023/3/24/23655804/internet-archive-hatchette-publisher-ebook-library-lawsuit>

Pico, Raquel C., «Biblioteca de memes para la posteridad», *Ethic*, 8 de noviembre de 2023. <https://ethic.es/2023/11/biblioteca-de-memes-para-la-posteridad>

«RTVE y Filmoteca Española lanzan en internet el mayor fondo histórico audiovisual de España», 20 de diciembre de 2012. <https://www.rtve.es/noticias/20121220/archivo-filmoteca-no-do-mayor-fondo-historico-audiovisual/590521.shtml>

The MySpace Archive Project. <https://archive.org/details/myspace-Music?tab=collection>

«Thousands of writers demand AI stop using work without permission», *Aljazeera*, 19 de julio de 2023. <https://www.aljazeera.com/economy/2023/7/19/thousands-of-writers-demand-ai-stop-using-work-without-permission>

Toro Nader, Mariana, «Tecnología para la arqueología literaria», *Ethic*, 17 de mayo de 2023. <https://ethic.es/2023/05/tecnologia-para-la-arqueologia-literaria/>

Unesco, *Cutting edge. Protecting and preserving cultural diversity in the digital era*, 2020. <https://www.unesco.org/en/articles/cutting-edge-protecting-and-preserving-cultural-diversity-digital-era>

Unesco, *El patrimonio digital: antecedentes*. <https://es.unesco.org/themes/information-preservation/digital-heritage/background>

Unesco, *Recommendation on the ethics of Artificial Intelligence*, 2022. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

Vincent, James, «DeepMind's new AI model helps decipher, date, and locate ancient inscriptions», *The Verge*, 9 de mayo de 2022. <https://www.theverge.com/2022/3/9/22968773/ai-machine-learning-ancient-inscriptions-texts-deepmind-ithaca-model>

World Economic Forum, *Creative disruption: The impact of emerging technologies on the creative economy*, 2018. <https://weforum.org/whitepapers/creative-disruption-the-impact-of-emerging-technologies-on-the-creative-economy>

Recursos

Archivo de la Web Española. <https://www.bne.es/es/colecciones/archivo-web-espanola/buscador>

Biblioteca Digital Hispánica. <http://bdh.bne.es/>

Cetso. Estilometría aplicada al Teatro del Siglo de Oro. <https://etso.es/cetso>

Chicago Film Archives. <https://www.chicagofilmmarchives.org/>

Europeana. <https://www.europeana.eu/es>

Filmoteca Española Online. <https://www.rtve.es/filmoteca/>

Humanidades Digitales Hispánicas. <https://humanidadesdigitaleshispanicas.es>

Internet Archive. <https://archive.org/>

Ithaca. <https://ithaca.deepmind.com/>

The Babylonian Engine. <https://digitalpasts.github.io/docs/BEn.html>

The Wayback Machine. <https://web.archive.org/>