



TIES Revista de Tecnología e Innovación en Educación Superior

DERIVACIONES DEL USO DE LA INTELIGENCIA ARTIFICIAL (IA) EN LA PRESERVACIÓN DIGITAL

Perla Olivia Rodríguez Reséndiz,
*Instituto de Investigaciones Bibliotecológicas y de la
Información, UNAM*
ORCID: 0000-0001-6517-8895

<http://www.ties.unam.mx/>

Fecha de recepción: marzo de 2023 • Fecha de publicación: febrero de 2024

Febrero 2024 | número de revista 9 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

DERIVACIONES DEL USO DE LA INTELIGENCIA ARTIFICIAL (IA) EN LA PRESERVACIÓN DIGITAL

Resumen

La capacidad de generar datos es constante e imparable. Cada año producimos más información que el anterior y su preservación digital es un problema social que afrontan las instituciones de la memoria, entre las que destacan los archivos sonoros y audiovisuales. Ante esta situación, la gestión y el manejo de grandes volúmenes de información digital a través de la Inteligencia Artificial (IA) podría significar una alternativa, cuyas derivaciones deben ser analizadas desde diferentes perspectivas.

En este artículo se comparte una reflexión en torno al empleo de la IA en la preservación digital, cuya incorporación resulta paradójica. La IA ofrece oportunidades de mejora en los procesos documentales y, al mismo tiempo, podría derivar en consecuencias adversas. Para establecer las derivaciones de la IA en los archivos sonoros y audiovisuales, se realizó una investigación cualitativa. Como resultado, se advierten costos medioambientales, patrimoniales y sociales, cuyo impacto en la preservación digital debe ser considerado, a fin de proponer estrategias de mitigación.

Palabras clave:

Inteligencia artificial, archivo, datos masivos, preservación digital.

ARTIFICIAL INTELLIGENCE (AI) DERIVATIONS IN DIGITAL PRESERVATION

Abstract

The generation capacity of data is constant and unstoppable. Each year we produce more information than the previous one and its digital preservation is a social problem facing sound and audiovisual archives, among other heritage institutions. In this situation, the management and handling of large volumes of digital information through Artificial Intelligence (AI) could represent an alternative whose derivations should be analyzed from different perspectives.

This article reflects on the use of AI in digital preservation, the incorporation of which seems paradoxical. It offers opportunities for improvement in documentary processes and at the same time could have adverse consequences. Qualitative research was carried out to establish the implications of AI in sound and audiovisual archives. As a result, environmental, patrimonial and social costs were identified, whose impact on digital preservation should be considered in order to suggest mitigation solutions.

Keywords:

Artificial intelligence, archiving, massive data, digital preservation.

DERIVACIONES DEL USO DE LA INTELIGENCIA ARTIFICIAL (IA) EN LA PRESERVACIÓN DIGITAL

1. Introducción

La inteligencia artificial (IA) se ha insertado en nuestra vida cotidiana como una herramienta para el procesamiento de grandes volúmenes de datos. Por medio de la IA se identifica, clasifica, predice y recupera información registrada en una amplia gama de formatos y lenguajes (textos, imágenes y sonidos). Se augura que el desarrollo científico se verá beneficiado, por ejemplo, en la salud, la biología y la genética para el tratamiento de enfermedades. Por otra parte, su inserción en la industria de los datos proyecta un crecimiento continuo en la economía mundial [1].

En fechas recientes, la IA ha devenido en una tecnología popular y polémica. Millones de personas utilizan programas como *Chat GPT (Open AI)*, *Bard (Google)* y *Bing (Microsoft)*, para realizar, desde una tarea escolar, hasta un discurso político. En contraste con la exaltación social que produce el uso de la IA, se han formulado preocupaciones y visiones críticas en torno a que podría degradar la ciencia y empobrecer la ética por ofrecer una visión errónea del lenguaje y del conocimiento [2].

Además, se ha señalado que la IA genera información a partir de datos publicados, sin hacer referencia a la propiedad intelectual, incurriendo con ello en una forma de plagio. Se alerta que puede ser una herramienta para la manipulación, a través de la generación de información falsa y *fake news*. Por ello, cada vez más foros y publicaciones ofrecen disertaciones sobre las implicaciones éticas del uso de la IA.

En este trabajo se abordará la relación de la IA con las instituciones de la memoria, es decir, los archivos, bibliotecas y museos, cuya misión es salvaguardar el patrimonio documental. De estas, se decantó por los archivos, donde se preservan diferentes tipos de documentos, como los sonoros y los audiovisuales, que salvaguardan importantes volúmenes de información, digitalizada o de origen digital, en los cuales la incorporación de la IA se percibe como una herramienta potencial que puede contribuir a la preservación digital, aunque su aplicación no está exenta de posibles efectos adversos.

El objetivo de este artículo es identificar las derivaciones asociadas con el uso de la inteligencia artificial en la preservación digital. Para ello, se optó por el empleo de la metodología cualitativa. Se presentan los antecedentes del uso de la IA en los archivos sonoros y audiovisuales. Como ejemplo, se abordan las primeras experiencias de investigación aplicada en el uso de IA en este tipo de instituciones de la memoria. Además, se definen y relacionan las nociones preservación digital, datos masivos e inteligencia artificial. Finalmente, se formulan algunas de las posibles implicaciones o los costos de carácter medioambiental, patrimonial y social, derivados del uso de la IA en la preservación digital.

2. La IA en la preservación digital

Los archivos se han erigido en espacios vinculados al poder político y al conocimiento de toda sociedad.

Preservan el pensamiento, las creaciones y los hechos que dan cuenta del acontecer de la sociedad. En la era de la información digital, los archivos han adquirido notoriedad por el acceso abierto y la transparencia en la información pública, cualidades esenciales de toda sociedad democrática.

Aun cuando en muchos casos subsiste el rezago en la digitalización de las colecciones analógicas, cada día se incrementa el volumen de documentos de origen digital, cuya preservación es una tarea inaplazable y compleja. Aplicar procesos documentales a grandes volúmenes de información es una tarea titánica. Por ello, desde hace más de una década, la IA ha despertado el interés de los archivistas para mejorar procesos como la catalogación, así como para recuperar contenidos de manera más eficiente.

De todos los tipos de archivos, destacan los que preservan contenidos sonoros y audiovisuales, con los cuales se han iniciado proyectos que ponen a prueba las posibilidades de la IA. Un ejemplo de ello es la experiencia del *American Archive of Public Broadcasting (AAPB)* de Estados Unidos, donde se han probado las posibilidades de usar el aprendizaje automático, en inglés *Machine Learning (ML)*, y la inteligencia artificial [3]. Otro caso es la iniciativa de Radio y Televisión Española (*RTVE*) para la generación automática de metadatos, a través de la transcripción automática de texto y segmentación de contenidos a fin de dar visibilidad a los materiales de archivo [4]. Asimismo, conviene destacar el desarrollo de *AIA Media* consorcio de investigación financiado por la Unión Europea y donde participan universidades, centros de investigación, archivos y medios públicos.

El primer intento por utilizar IA en los archivos sonoros y audiovisuales en México inició hace una década. La doctora Mireya Saraí García, del Centro de Investigación y Desarrollo de Tecnología Digital del Instituto Politécnico Nacional, lideró *MEX CULTURE*, un proyecto de investigación aplicada, encaminado a probar herramientas de inteligencia artificial para la recuperación de contenidos guardados en archivos sonoros [5]. Este propósito fue complejo porque hace una década apenas habían comenzado las tareas de digitalización. Por ello, la preservación de documentos de origen digital era una práctica desconocida. No había suficientes contenidos para aplicar y probar la IA en México. Además, para los responsables de los archivos la idea de aplicar algoritmos en una copia de los contenidos digitales era incomprensible y suponía un riesgo.

Así, las primeras aplicaciones de inteligencia artificial en archivos, promovidas por investigadores mexicanos,

florecieron en Colombia. En Radio y Televisión Nacional de Colombia se creó el Laboratorio de Inteligencia Artificial de RTVC, impulsado por la maestra Dora Braussin. Desde entonces esta iniciativa es un referente en América Latina.

En 2019, el Instituto de Investigaciones Bibliotecológicas y de Información de la UNAM, en colaboración con la Dirección General de Cómputo y de Tecnologías de Información y Comunicación, organizó el Congreso Internacional de Archivos Digitales, con el tema *Inteligencia Artificial y Big Data*. En este Congreso se presentaron algunas de las experiencias más relevantes de Europa y Estados Unidos en la aplicación de la IA en archivos audiovisuales.

Por la relevancia del tema, se emprendieron investigaciones en el ámbito de las Ciencias Sociales y las Humanidades en la UNAM. Entre otras, conviene señalar que inició la primera tesis de doctorado sobre el uso de la IA en archivos sonoros, a cargo de la maestra Georgina Sanabria Medina. En 2020 se creó, por iniciativa del doctor Pedro Salazar Ugarte, la Línea de Investigación en Derecho e Inteligencia Artificial (*LIDIA*) del Instituto de Investigaciones Jurídicas de la UNAM. En marzo de 2023, *LIDIA* y el Instituto Nacional de Acceso a la Información (*INAI*), organizaron el foro “Retos del uso de Inteligencia Artificial en la actividad archivística de México”.

En la actualidad, la IA y los archivos son un ámbito de interés científico y profesional.

3. Las derivaciones

Para garantizar la permanencia de los objetos digitales y los metadatos, la preservación digital no debe ser interrumpida. Los datos se almacenan en archivos o centros de datos, a través de los cuales se tiene acceso a la información de un determinado ámbito o negocio. Un ejemplo de ello son los servicios de información que ofrecen *Alexa* y *Siri* como asistentes virtuales. Cuando se formulan consultas, entra en operación un complejo y vasto sistema de aplicaciones informáticas, que procesan, en cuestión de segundos, millones de datos para arrojar de manera inmediata la información solicitada. En este proceso intervienen dos términos: el *big data* y la inteligencia artificial. El *Big data* procesa y recupera grandes volúmenes de información, gracias al uso de la inteligencia artificial [6]. El empleo de la IA en el procesamiento de grandes volúmenes de datos para la preservación digital, tiene al menos tres consecuencias o costos: medioambiental, patrimonial y social, que a continuación se desarrollan.

4. Medioambiental

Los centros de datos son instalaciones para la conservación y la gestión de sistemas informáticos y componentes asociados, como telecomunicaciones y sistemas de almacenamiento. Cuentan con un entorno seguro y una temperatura controlada [7]. En concordancia con la definición anterior, los archivos digitales constituyen un tipo de centro de datos.

Los centros de datos son parte de la sociedad contemporánea, porque preservan grandes volúmenes de información digital, tarea para la cual se requiere del uso de energía eléctrica de manera ininterrumpida, no sólo para la conservación a través del almacenamiento de los datos, sino también para llevar a cabo los procesos de búsqueda y recuperación de información. En otros términos, que la búsqueda de información desde cualquier dispositivo conectado a Internet genera un impacto en el medio ambiente.

Los centros de datos dejan una huella en el medio ambiente porque consumen energía y agua constantemente. Utilizan entre 10 y 50 veces más energía que una oficina. En 2021, emplearon el 1% de la demanda mundial de energía y millones de galones de agua para enfriar los sistemas de almacenamiento y generar energía [7].

Estudios recientes señalan que, para 2040, el sector de las tecnologías de la información emitirá el 14% de las emisiones de carbono y que la demanda de energía en los centros de datos se multiplicará por 15 en 2030 [8]. Este incremento está asociado al uso de la inteligencia artificial, porque para el procesamiento de grandes volúmenes de datos se emplean arquitecturas basadas en redes neuronales, con millones de parámetros que se deben entrenar [8].

Las empresas que mantienen los grandes centros de datos no proporcionan información sobre la huella de carbono, el uso de agua para enfriar los sistemas de almacenamiento y la producción de basura tecnológica que se genera al entrar en desuso la tecnología cada determinado tiempo [9]. Diseñan planes para reducir el impacto en el medio ambiente y afirman que contribuirán a que la información proporcionada, mediante los centros de datos, ayude a disminuir las emisiones mundiales de gases de efecto invernadero, entre un 6 y 12%, en 2030 [7].

No obstante, las estimaciones actuales proyectan que el impacto ambiental de las tecnologías de la información se acrecentaría con el uso de la IA, al ser una herramienta utilizada por millones de personas en el

mundo y porque, precisamente, durante el proceso de entrenamiento para perfeccionar su desempeño, requiere de mejores procesadores de información [8].

5. Patrimonial

Para garantizar la preservación de grandes volúmenes de datos, se requieren sistemas robustos de almacenamiento digital que emplean la nube como una solución. La preservación en la nube es una decisión de los responsables de los archivos, que tiene implicaciones patrimoniales y se relaciona con la noción de la soberanía digital.

El cómputo en la nube se refiere al uso de servidores, ubicados en lugares remotos, que a través de Internet almacenan, gestionan y procesan datos, en lugar de utilizar un servidor local o personal. Con esta infraestructura, las aplicaciones compartidas a través de Internet pueden ser utilizadas por múltiples usuarios. Los servicios de cómputo en la nube pueden ser proporcionados por proveedores externos, que ofrecen y mantienen el *software* y el *hardware* necesarios [7] para la preservación de grandes volúmenes de información digital.

Esta solución puede ser atractiva porque resuelve de manera práctica la gestión y la conservación de datos en un archivo. Sin embargo, se observa cada vez más resistencia a la dependencia de empresas como *Amazon*, *Google* o *Microsoft* en materia de almacenamiento y gestión de grandes volúmenes de datos. Este tema cobró relevancia en la agenda política de Europa desde 2019 [10].

El Parlamento Europeo reconoció que los datos y la inteligencia artificial son elementos para la innovación y para resolver problemas sociales, desde la salud y la agricultura, hasta la seguridad. Por ello, asignaron 20 millones de euros anuales para el desarrollo de la inteligencia artificial [10]. Con esta determinación se reconoció la importancia de invertir para crear soluciones de almacenamiento digital que no vulneren la soberanía digital. La experiencia europea debería motivar la inclusión del tema en la agenda política de todas las naciones, para sentar las bases de la soberanía digital y garantizar la protección del patrimonio.

6. Social

La manipulación en el procesamiento de grandes volúmenes de datos conlleva riesgos sociales. A los patrones de entrenamiento de la IA se pueden trasladar los prejuicios y las discriminaciones que ya existen en la sociedad, por razón de sexo, género, religión, opiniones políticas,

discapacidad, orientación sexual y edad, entre otras [11]. Esto supone imponer la visión de una persona o de un grupo de personas sobre los modos de interpretación y recuperación de un conjunto de datos.

La preocupación por el uso indebido de la IA motivó a investigadores y tecnólogos a firmar una carta para advertir sobre los riesgos que para la sociedad y la humanidad acarrearía el desarrollo sin control de la IA [12]. Se señala en la misiva que el desarrollo de la tecnología está fuera de control y podría crear mentes digitales cada vez más poderosas, que ni siquiera sus creadores podrían entender o predecir. En consecuencia, se formuló un llamado para detener, al menos seis meses, el desarrollo de la tecnología, a fin de poner en marcha protocolos de seguridad y crear un marco regulatorio en el que se involucren los legisladores de cada país para la supervisión y el seguimiento de los grandes sistemas de información que emplean IA [12].

Las posibilidades de manipulación y uso indebido de la IA en grandes volúmenes de datos dejan al descubierto el costo social. El tratamiento de la información puede incidir en la toma de decisiones de la sociedad e incluso vulnerar la democracia [6].

La discusión de la IA en la preservación digital debe basarse en la transparencia del procesamiento de la información, la mitigación de la discriminación de cualquier tipo y la garantía de que no se vulneren los derechos individuales en torno a la posible toma de decisiones automatizadas. El uso erróneo de la tecnología afectaría la confianza y la credibilidad en el archivo. La IA es una herramienta que puede proteger y crear modos creativos, que faciliten el acceso sencillo y más amplio a los datos de los bienes culturales digitales que forman parte del patrimonio digital [13].

7. Conclusiones

En la preservación digital de grandes volúmenes de datos el uso de IA es paradójico. Podría tener aplicaciones que beneficien los procesos documentales, para asegu-

rar la permanencia, el valor y el acceso a los contenidos, pero al mismo tiempo acarrear consigo costos o implicaciones adversas. En relación con lo cual, se identifican tres posibles derivaciones: medioambiental, patrimonial y social. En referencia a la medioambiental, se advierte que el uso de la IA podría acrecentar, en la preservación digital, el uso de energía eléctrica, grandes cantidades de agua y la generación de basura tecnológica, y con ello incidir en la afectación de las tecnologías de la información en el medio ambiente.

El costo patrimonial alude a que el volumen de datos preservados es tan grande que en algunos casos se emplean soluciones comerciales como la nube. Su utilización podría significar un peligro que vulnere la soberanía y sitúe en condición de riesgo al patrimonio digital. De ahí que sea necesario invertir en tecnología y en el desarrollo de infraestructura nacional para la preservación y almacenamiento digital de grandes volúmenes de datos. Asimismo, es necesario considerar que el tema debe ser parte de la agenda política nacional de cada país. El costo social se refiere a los riesgos asociados al uso indebido y la manipulación de la información. El uso indebido de la IA en el tratamiento documental afecta la credibilidad y la confianza, valores sobre los que un archivo se sustenta. Además, ubica en condición de vulnerabilidad a la sociedad porque pone en peligro los derechos humanos.

La disertación en relación con las derivaciones o costos, medioambiental, patrimonial y social, debe ser parte de la discusión pública. Es necesaria para diseñar y poner en marcha estrategias que aminoren el posible impacto negativo de la IA en la preservación digital. Entre otros fundamentos, se advierte el establecimiento de los éticos en el uso de la IA, que prevengan un uso social inadecuado; la creación de alternativas que aminoren el futuro impacto de la tecnología en el medio ambiente, y la puesta en marcha de estrategias tecnológicas que protejan el patrimonio nacional de los pueblos y no vulneren su soberanía.

BIBLIOGRAFÍA

[1] Comisión Europea. 2020. "Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza" [COM (2020) 65 final]. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

[2] N. Chomsky, I. Roberts y J. Watumull. "Noam Chomsky: The False Promise of ChatGPT". *The New York Times*, 23 de marzo de 2023. En línea: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

[3] K. Cariani, Karen y D. Ives. "Using Computational Tools and Experts to Improve Access to Digital Media Archive". En *Inteligencia artificial y datos masivos en archivos digitales sonoros y audiovisuales*, pp. 101-115. México: UNAM-IIBI, 2020.

[4] V. Bazán Gil. "Reflexiones en torno a la IA: el caso de RTVE". En *Inteligencia artificial y datos masivos en archivos digitales sonoros y audiovisuales*, pp. 117-131. México: UNAM-IIBI, 2020.

[5] M. García Vázquez, Y. Montiel Pérez, J. Benois- Pineau y M. Crucianu, Michel. "Plataforma multimedia Mex-Culture". En *Archivos Digitales Sustentables. Conservación y acceso a las colecciones sonoras y audiovisuales para las sociedades del futuro*, pp. 141-155. IIBI UNAM, 2017.

[6] D. Kusnetzky. "What is Big Data?" *ZDNet*. 2023. Disponible en: <https://www.zdnet.com/article/what-is-big-data/>

[7] Digital Climate Alliance (DCA). "Sustainable data centers: Powering the digital revolution". *Digital Climate Alliance*. 2023. Disponible en: <https://www.digitalclimate.io/data-center-white-paper>

[8] M. Pascual. "El sucio secreto de la inteligencia artificial". *El País*. 2023. Disponible en: <https://elpais.com/tecnologia/2023-03-23/el-sucio-secreto-de-la-inteligencia-artificial.html>

[9] L. Tadic. "La relación entre archivos audiovisuales y el medioambiente". En *Cambio climático y preservación digital sonora y audiovisual*. Quito, EC: Universidad Andina Simón Bolívar, Sede Ecuador / Red Iberoamericana de Preservación Digital de Archivos Sonoros y Audiovisuales, RIPDASA / Ciencia y Tecnología para el Desarrollo, CYTED, pp. 11-28. 2022.

[10] LL. Pellicer e I. Fariza. "Europa ultima un plan para dar la batalla en el negocio de los datos". *El País*, 16 de noviembre de 2019. Disponible en: https://elpais.com/economia/2019/11/16/actualidad/1573926886_318836.html

[11] UK Information Commission Officer. "Guidance on AI and data protection" *ICO, UK*. 2023. Disponible en: <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/>

[12] Institute Future of life. "Pause Giant AI Experiments: An Open Letter", 30 de marzo de 2023. Disponible en: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

[13] Parlamento Europeo. "La inteligencia artificial en los sectores educativo, cultural y audiovisual". *Resolución del Parlamento Europeo*, 19 de mayo de 2021. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_ES.html

Fecha de recepción: 30 de marzo, 2023

Fecha de publicación: febrero, 2024

Cómo se cita

P. O. Rodríguez Reséndiz, "Derivaciones del uso de la Inteligencia Artificial (IA) en la preservación digital", *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 9, febrero, 2024. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].