

# **Bibliotecaria *in-the-loop*: oportunidades y desafíos de la Inteligencia Artificial en instituciones GLAM**

## **Librarian *in-the-loop*: opportunities and challenges of Artificial Intelligence in GLAM institutions**

Ana Fresco-Santalla. Bibliotecaria. ORCID iD: [0000-0001-8218-6408](https://orcid.org/0000-0001-8218-6408)

Texto del póster presentado en las XVIII Jornadas Españolas de Información y Documentación (JEID23).

### **Resumen**

---

La expresión Inteligencia Artificial fue elegida palabra del año en 2022 (FundéuRAE), lo que evidencia la popularidad que ha ido ganando tanto dentro como fuera de Internet. Más allá del imaginario colectivo, en el que es identificada como una máquina pensante, la Inteligencia Artificial (en adelante IA) es –en realidad– un conjunto de tecnologías con aplicaciones transversales en áreas diversas.

Este póster es una aproximación a la IA: qué es, qué tecnologías engloba y para qué sirven, prestando especial atención a sus aplicaciones en el trabajo que realizan los profesionales del sector GLAM (Galleries, Libraries, Archives and Museums).

Aunque la IA ofrece grandes oportunidades para mejorar tanto los servicios como los procesos de trabajo en nuestras instituciones, existen importantes barreras susceptibles de retrasar su adopción. Estos obstáculos, incluido el impacto que la IA podría suponer para la profesión, así como los dilemas éticos suscitados por su uso, también son planteados en el póster.

**Palabras clave:** inteligencia artificial (IA); aprendizaje automático; instituciones GLAM; bibliotecas; patrimonio cultural

### **Abstract**

---

*The popularity of Artificial Intelligence (AI) has been growing both within and outside the Internet, as evidenced by its selection as the word of the year in 2022 by FundéuRAE. Despite its identification as a thinking machine in the collective imagination, AI is a collection of technologies with diverse applications in various fields.*

*In this poster, we provide an approach to AI, its constituent technologies, and their applications. We focus on its applications in the GLAM (Galleries, Libraries, Archives, and Museums) sector, with a special emphasis on its potential to enhance the work processes and services provided by professionals in these institutions.*

*Despite the significant benefits that AI offers, its adoption may be hindered by various obstacles, including ethical dilemmas and concerns about the impact of AI on the profession. This poster addresses these barriers to adoption, along with ethical dilemmas raised by its use.*

**Keywords:** artificial intelligence (AI); machine learning (ML); GLAM institutions; libraries; cultural heritage

## **1. ¿Qué es la Inteligencia Artificial?**

Concepto complejo para el que no existe una única definición. En esencia, la IA es un campo de estudio centrado en el diseño de programas cuyas características y comportamiento imitan la inteligencia humana: sistemas que aprenden a tomar decisiones a partir de los datos y de la experiencia.

## **2. La IA está en todas partes, pero no es visible en ningún sitio**

Quizás no seamos conscientes, pero hace tiempo que la IA está presente en nuestro día a día.

Ejemplos de su uso cotidiano:

- Smartphones (reconocimiento facial).
- Asistentes de voz (Siri, Alexa).
- Chatbots o agentes conversacionales.
- Sistemas de recomendaciones (Netflix, Amazon, Spotify).
- Correo electrónico (filtro de spam).
- OCR (*Optical Characters Recognition*). Sí, el reconocimiento óptico de caracteres también es una tecnología de Inteligencia Artificial.
- Subtítulos automáticos (Youtube).
- Motores de traducción (DeepL).
- Coches autónomos (¿en el futuro próximo?)

### 3. ¿Una o muchas?

La IA es un campo general que engloba diversos subcampos: **Aprendizaje Automático**, **Aprendizaje Profundo**, **Procesamiento del Lenguaje Natural** o **Visión Artificial**, entre otros. Por lo tanto, la IA no es una sola tecnología, sino un conjunto de tecnologías con aplicaciones transversales en numerosos sectores: justicia, defensa, turismo, agricultura, finanzas, salud, educación, etc. Como veremos, las colecciones de las instituciones GLAM también son un campo de juego para la IA.

### 4. *No data, no IA*

Sin datos las tecnologías de IA no pueden extraer patrones y, por lo tanto, no es posible generar conocimiento: los datos pueden ser tanto información estructurada (datos tabulares) como no estructurada (imágenes, vídeo, documentos, audio).

Una tecnología de IA será tan buena como los datos usados en su entrenamiento: si la información seleccionada está incompleta, no es representativa (sesgada) o está mal anotada, el algoritmo reproducirá y amplificará los mismos sesgos y prejuicios humanos (raza, sexo, etc.) que estuvieran presentes en los conjuntos de datos utilizados para entrenarlo. Un dataset sesgado producirá un algoritmo sesgado que, a su vez, podrá afectar al comportamiento humano. Por ejemplo:

Sesgo de género:

- Un algoritmo de recomendación literaria entrenado con una colección de textos literarios escritos en su mayoría por hombres, seleccionará (recomendará) preferentemente obras de autores porque ha aprendido *erróneamente* que los escritores son hombres.

Conocer los datos que vamos a usar (saber dónde puede existir un sesgo) para crear conjuntos de datos de entrenamiento precisos y fiables, equilibrados, perfectamente anotados y etiquetados (*ground-truth data*), es fundamental para prevenir la **discriminación o sesgo algorítmico**. En este proceso, el control y la supervisión humana es fundamental (*human-in-the-loop*).

## 5. Pide la llave ¿Es la IA una caja negra?

Con frecuencia es muy difícil –cuando no imposible– explicar cómo y por qué un modelo genera unos determinados resultados: los datos entran y los resultados salen, pero el usuario desconoce cuál ha sido el proceso entre el input y el output. Cuando esto sucede, hablamos de algoritmos de **caja negra** o *black box*. El resultado: a mayor opacidad o falta de transparencia algorítmica, menor confianza en el sistema de IA.

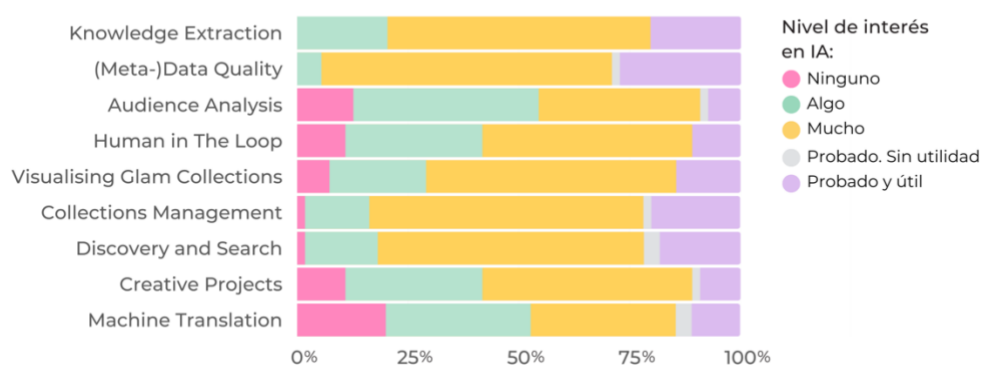
## 6. IA en el sector GLAM: del escepticismo al entusiasmo

Las tecnologías de IA ofrecen grandes oportunidades para mejorar tanto los procesos de trabajo como los servicios que prestan las instituciones GLAM. De un aparente escepticismo inicial respecto a su impacto en la profesión, hemos pasado –en la actualidad– a una etapa en la que el interés por comprender, experimentar y aprovechar el potencial de la IA parece ir en aumento.

En apenas tres años la situación ha cambiado; en 2017, el desinterés entre los profesionales – bibliotecarios en este caso– era dominante:

- “In a 2017 survey of librarians from across all sectors in the USA, Wood and Evans (2018) found that 56.3% of respondents thought supercomputers, like Watson, could transform librarianship. This still meant that 44% thought it would have nor or not much effect” (Cox, Pinfield & Rutter, 2019).
- “No university or university library mentions artificial intelligence in their strategic plan. While most plans were fairly recent, there are no initiatives to address the rise of artificial intelligence technology” (Wheatley & Hervieux, 2019).

En 2020, en cambio, los datos sugieren lo contrario:



**Gráfico.** Nivel de interés en IA. Encuesta realizada en 2020 por el Grupo de Trabajo EuropeanaTech AI in relation to GLAMs. Fuente: *AI in relation to GLAMS Task Force. Report and recommendations* (2021).

Asimismo, el creciente número de publicaciones profesionales, la puesta en marcha de AI Labs, la creación de grupos de trabajo, así como la celebración de foros y congresos son un buen indicador del interés por la IA en el sector GLAM:

#### **Comunidades y Grupos de trabajo:**

- AI4LAM
- AEOLIAN Network
- CENL AI in Libraries Network Group
- Cultural AI Lab
- Europeana Task Force: AI in relation to GLAMs
- IFLA Special Interest Group on AI
- The Museums + AI Network

#### **AI Labs:**

- The AI-lab (National Library of Norway)
- BL Labs (British Library)
- BNF DataLab (National Library of France)
- BNELab (Biblioteca Nacional de España)
- KB LAB (National Library of Netherlands)
- KB LAB (National Library of Sweden)
- LC Labs (Library of Congress)

#### **Foros y Congresos:**

- Fantastic Futures
- New Horizons in AI in Libraries (IFLA WLIC 2022)
- PLAID3 Summit

## **7. ¿Qué hay para mí? Oportunidades de la IA**

Sin datos no hay IA, y las instituciones del patrimonio cultural si son ricas en algo es, precisamente, en datos: sus colecciones digitales/digitalizadas (manuscritos, prensa, mapas,

fotografías, pinturas, audiovisuales, partituras, etc.) son un campo de juego ideal para que los profesionales del sector GLAM experimentemos y exploremos posibles usos y aplicaciones de las tecnologías de IA.

El sector GLAM engloba instituciones con colecciones, recursos y públicos diferentes; no obstante, podemos encontrar intereses y áreas de actividad comunes: catalogación y generación/enriquecimiento de metadatos (descriptivos, estructurales, etc.) que, a su vez, contribuirán a mejorar el descubrimiento y visualización de contenidos, servicios para fidelizar/atraer usuarios, y gestión e investigación sobre las colecciones. La mayoría de los proyectos –liderados principalmente por grandes centros de carácter nacional y/o académicos– están aplicando tecnologías de IA en dichas áreas. Aunque muchos proyectos son de carácter experimental, hay algunos que están en fase de producción.

### **7.1. Catalogación y generación/enriquecimiento de metadatos**

Proyectos y ejemplos de uso:

- **Catalogación automática** de libros procedentes del depósito legal (Bibliothèque Royale de Belgique). Han desarrollado una aplicación que ha permitido a la Biblioteca procesar con bastante rapidez 19.000 portadas digitalizadas, lo que a un catalogador medio le hubiera llevado cinco años de trabajo.
- **Cataloguing Machine** (Deutsche Nationalbibliothek). Indización por materias automatizada con ANNIF.
- **Norart Klassificator** (National Library of Norway). Clasificación automática (asignación automatizada de números DDC).
- **TIGER – Tagging Images Generically for Exploration and Research** (State Library New South Wales). Etiquetado automático de imágenes.
- **qUlc** (Arxiu Lliure Teatre Lliure). Reconocimiento facial para identificar personas en fotografías.
- **Saint George on a bike – SGoaB** (Europeana Foundation + BSC-CNS Barcelona Supercomputing Center). Enriquecimiento de metadatos asociados a obras pictóricas. En desarrollo.
- **LECTAUREP – Lecture automatique de répertoires** (Ministère de la Culture, Inria, Archives Nationales). Reconocimiento y transcripción automática de texto manuscrito: HTR (*Handwritten Text Recognition*) + NER (*Named Entity Recognition*).

- **PolifonIA** (Universidad de Alicante). OMR (*Optical Music Recognition*). Partituras musicales. En desarrollo.
- **Proyecto Carabela** (PRHLT Universitat Politècnica de València + Centro de Arqueologia Subacuática del Instituto Andaluz del Patrimonio Histórico). HTR e indexación probabilística. En desarrollo.

## 7.2. Usuarios

### Descubrir y explorar colecciones:

- **MAKEN** (National Library of Norway). Búsqueda por similitud de imágenes y texto (representación vectorial de imágenes / palabras – *vector embeddings*).
- **Gallica Pix** (Bibliothèque Nationale de France). Búsqueda de imágenes por similitud. Extracción y clasificación de imágenes (gravados, dibujos, mapas, etc.) contenidas en las colecciones digitales (prensa, libros, manuscritos, etc.).
- **Image-based similarity search** (Bayerische Staatsbibliothek).
- **MOSAIC** (MET + MIT). Descubrir conexiones ocultas entre obras de arte (color, textura, etc.).
- **Newspaper Navigator** (Library of Congress). Extracción de contenido visual (fotografías, ilustraciones, viñetas, mapas, etc.) de más de 16 millones de páginas de *Chronicling America* para crear una interfaz de búsqueda que facilita la exploración del material gráfico.
- **Línea del Tiempo** (Museo del Prado). NER + NEL para desarrollar una herramienta de lectura aumentada.
- **HAMLET – How About Machine Learning Enhanced Theses?** (Andromeda Yelton + MIT). Búsqueda por similitud semántica de textos (*vector embeddings*).

### Chatbots:

- **Bizzy** (University of Oklahoma Libraries).
- Chatbot del Centre Pompidou.
- **KB Lab Bot** (National Library of the Netherlands). Prototipo.
- **LAMA** (Library Ask Me Anything) (Health Library Ireland).
- **Litté Bot** (Bibliothèque Nationale de France). Prototipo.
- **Plácido** (Berlenga Film Museum, Valencia).

- **They Voice of Art** (Pinacoteca de São Paulo). Chatbot cognitivo. Conversa con los visitantes.

#### Robots:

- **Libby** (University of Pretoria). Robot asistente usuarios.
- **Pepper** (Museu Europeu d'Art Modern; Smithsonian Museums). Robot guía visitantes.

### 7.3. Gestión e investigación

#### Proyectos:

- **InterPARES Trust AI** (University of British Columbia). Soluciones de IA para la gestión documental. Riesgos y beneficios. Aplicaciones prácticas.
- **Qurator Project – Automated curation technologies for the digitised cultural heritage** (German Federal Ministry of Education and Research). Sistema inteligente para la gestión de patrimonio cultural digitalizado (análisis de imagen, OCR, NER, etc.).
- **AMP: Audiovisual Metadata Platform** (Indiana University Libraries + University of Texas Austin + New York Public Library + AVP). Plataforma para generar metadatos asociados a colecciones audiovisuales.
- **AIDA Image Analysis for Archival Discovery** (Digital Libraries Research Lab. University of Nebraska-Lincoln + Library of Congress). Segmentación de imágenes; extracción de imágenes/gráficos de páginas; extracción de texto de imágenes; clasificación de imágenes.
- **Museum Feedback: análisis de experiencia de usuarios con facial tracking** (Museo de la Evolución Humana). Mejorar la experiencia de usuario combinando *face tracking* (reconocimiento facial), realidad aumentada y asistencia de voz.
- **Investigación histórico–filológica**: caso de Lope de Vega y su obra *La francesa Laura* (ETSO: Estilometría aplicada al Teatro del Siglo de Oro + BNE + Grupo PROLOPE). HTR (*Transkribus*) + Estilometría para resolver problemas de autoría.

#### Datasets y modelos:

- **KB-BERT Model** (National Library of Sweden).
- **NCC** (*Norwegian Colossal Corpus*) y **NorNE** (*Norwegian Named Entity Recognition*) (National Library of Norway).



- **MarIA** (Biblioteca Nacional de España). Primer modelo de inteligencia artificial masivo de la lengua española.

## 8. El lado oscuro de la IA: desafíos

Las tecnologías de IA ofrecen grandes oportunidades para el sector GLAM, pero su aplicación plantea importantes desafíos que requieren atención.

### 8.1. Cuestiones éticas y legales

**Privacidad y vigilancia:**

- Cumplimiento **RGPD** (control de los datos, consentimiento, olvido etc.).
- Respetar el principio de **prevención del daño**.

Porque, que algo sea legal, no significa que sea ético:

- ¿Reconocimiento facial de las emociones de los visitantes de una exposición? (Por ejemplo, *Share Art* en los Museos de Bolonia).
- ¿Usamos datos de uso de recursos-e para combinarlos con datos personales (edad, sexo, raza, renta, etc.) y datos de rendimiento académico para personalizar sugerencias de lectura?

**Sesgos (datasets / algoritmos):**

- Respetar el principio de **equidad**.
- Evitar el *nudging* (influir en el comportamiento de los usuarios aprovechando sesgos). Respetar el principio de **autonomía humana**.

**Transparencia vs. Opacidad algorítmica:**

- “Pide la llave”. Mejor soluciones *open-source* que sistemas propietarios: *“the development team discovered late in testing that AWS Transcribe speech-to-text transcription was censoring what it deemed ‘offensive’ words in transcripts without opting into that feature and without providing the list of censored words”* (Dunn et al., 2021).
- **Explainable AI** (XAI): asegurar el principio de la **explicabilidad**.

**Copyright:**

- Reutilización de datos en el contexto IA.

## 8.2. Profesionales GLAM

### Impacto en la profesión:

- Desconfianza y temor.
- ¿Pérdida de empleos?
  - Es posible: “*Frey and Osborne estimate the probability of the replacement by computers of ‘library technicians’ as 99%, ‘Library assistants, clerical 95%, archivists 76% and librarians 65%’*” (Cox, Pinfield & Rutter, 2019).
  - Pero hay otras oportunidades: nuevas tareas, nuevos perfiles, etc.

### Nuevas competencias:

- Actualización formación: contenidos curriculares (Grados / Másteres) y acciones formativas asociaciones profesionales (SEDIC, FESABID, ANABAD, etc.).

¿Por dónde empezar?: *Elements of AI, Programming Historian, Intro to AI for GLAM – The Carpentries Incubator*, etc.

### Nuevas tareas:

- Boomers, X, Millenials, Z, alfa... y ¿Generación IA?: Alfabetización en IA / Alfabetización algorítmica.
- Selección y creación de **corpus y datasets** para entrenamiento (*ground-truth*).
- Practicar una **IA ética** (privacidad, explicabilidad, decisiones justas, seguridad, etc.).
- Enfoque *Human-in-the-loop*: usar tecnologías de IA cuyo diseño haya sido supervisado personas. Es decir: **profesionales GLAM**.

### Nuevos roles:

- Actualización perfiles profesionales.

## 9. Bibliografía

Cordell, R. (2020). Machine learning and libraries: a report on the state of the field. LC Labs, Library of Congress. <https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf>

Cox, A. (2021). The impact of AI, machine learning, automation and robotics on the information professions: a report for CILIP. <https://eprints.whiterose.ac.uk/174522/>

Cox, A., Pinfield, S., & Rutter, S. (2019) The intelligent library. *Library Hi Tech*, 37 (3). <https://doi.org/10.1108/LHT-08-2018-0105>

Darby, A., Coleman, C. N., Engel, C., van Strien, D., Trizna, M., & Painter, Z.W. (2022). AI training resources for GLAM: a snapshot. <https://doi.org/10.48550/arXiv.2205.04738>

Dunn, J.W., Feng, Y., Hardesty, J. L., Wheeler, B., Whitaker, M., Whittaker, T., Averkamp, S., Lyons, B., Rudersdorf, A., Clement, T., & Fischer, L. (2021). Audiovisual Metadata Platform Pilot Development (AMPPD). Final Project Report. <https://hdl.handle.net/2022/26989>

European Commission, Directorate-General for Communications Networks, Content and Technology. (2022). Opportunities and challenges of artificial intelligence technologies for the cultural and creative sectors. *Publications Office of the European Union*. <https://data.europa.eu/doi/10.2759/144212>

Gregory M. et al. (2021). AI in relation to GLAMs task Force. Report and Recommendations. Europeana Network Association. <https://pro.europeana.eu/project/ai-in-relation-to-glams>

Lorang, E., Soh, L.K., Liu, Y., & Pack, Ch. (2020). Digital Libraries, Intelligent Data Analytics, and Augmented Description: A Demonstration Project. UNL Libraries. <https://digitalcommons.unl.edu/libraryscience/396>

Murphy, O., & Villaespesa, E. (2022). La Red de Museos + Inteligencia Artificial: Guía para la aplicación de IA en museos. London: Goldsmiths, University of London. <https://themuseumsai.network/toolkit/>

Wheatley, A., & Hervieux, S. (2019). Artificial Intelligence in Academic Libraries: An Environmental Scan. *Information Services & Use*, 39. <http://dx.doi.org/10.3233/ISU-190065>