



Primeras experiencias con inteligencia artificial en la Biblioteca Nacional

JOSÉ LUIS BUEREN GÓMEZ-ACEBO
Biblioteca Nacional de España.

Este artículo presenta una panorámica de las primeras experiencias de la Biblioteca Nacional de España (BNE) en la aplicación de tecnologías asociadas a la inteligencia artificial. Partiendo de una breve contextualización histórica sobre la incorporación de la tecnología en las bibliotecas, se analizan distintos proyectos desarrollados o en fase experimental en la BNE, como el catálogo semántico datos.bne.es, el reconocimiento automático de imágenes, la catalogación automática de libros o la transcripción de manuscritos.

El texto aborda igualmente el papel singular que puede desempeñar la BNE en el desarrollo de una inteligencia artificial ética y legal, gracias al volumen y diversidad de los corpus documentales que custodia a través del depósito legal y la digitalización de colecciones. Se subrayan, no obstante, las limitaciones actuales derivadas de la escala de las colecciones, la complejidad de los procesos técnicos, la rápida evolución de las herramientas y las incertidumbres legales existentes.

Finalmente, se reflexiona sobre las perspectivas de futuro y sobre la necesidad de integrar estas tecnologías de forma prudente, sin perder de vista la misión esencial de la institución: la conservación y difusión del patrimonio bibliográfico español.

Inteligencia artificial; bibliotecas nacionales; catalogación; humanidades digitales; patrimonio documental

This article offers an overview of the first experiences of the National Library of Spain (BNE) in the application of artificial intelligence-related technologies. After a brief historical contextualisation of the role of technology in libraries, it analyses several projects developed or tested at the BNE, including the semantic catalogue datos.bne.es, image recognition, automatic book cataloguing and handwritten text transcription.

The paper also highlights the unique role that the BNE, and cultural heritage institutions in general, can play in the development of legal and ethical artificial intelligence, thanks to the scale and diversity of the documentary corpora preserved through legal deposit and large-scale digitisation programmes. At the same time, it addresses the current limitations affecting the widespread implementation of these technologies, such as the size and complexity of the collections, infrastructural constraints, the rapid evolution of AI tools and existing legal uncertainties.

The article concludes with a reflection on future perspectives, stressing the importance of a cautious and sustainable integration of artificial intelligence that supports, rather than compromises, the core mission of the National Library: the preservation and dissemination of Spain's bibliographic heritage.

Artificial intelligence; national libraries; cataloguing; digital humanities; documentary heritage

Introducción

Desde su origen, las bibliotecas han integrado la tecnología en su sentido más amplio para el desarrollo de sus trabajos. Si nos centramos en la aparición de la informática, las bibliotecas fueron siempre permeables a la incorporación de todas las ventajas que ésta trajo consigo (catalogación automatizada, intercambio de registros, integración de funciones como el préstamo o la adquisición, digitalización de colecciones).

La propia Biblioteca Nacional de España (BNE) es ejemplo de este recorrido. Comenzó su automatización en los años ochenta; a finales del siglo XX inició la digitalización sistemática de sus colecciones (antes empleaba la microfilmación, que es también una tecnología), y ya en el siglo XXI arrancaron proyectos como el depósito legal de publicaciones en línea, el desarrollo del catálogo semántico datos.bne.es o el impulso de la reutilización de colecciones. El resultado de estos procesos ha sido extraordinariamente positivo: el control y el acceso a las colecciones es hoy mayor que en ningún otro momento de su historia. No hay que olvidar que la informática nace, en buena medida, con el fin de mejorar la gestión de la información.

En este contexto, la aparición de las tecnologías asociadas a la inteligencia artificial no deja de ser un paso más en este proceso evolutivo¹. Conviene recordar, además, que bajo el paraguas de la “inteligencia artificial” se agrupan herramientas y procesos muy diversos: desde la identificación de elementos en una imagen hasta la traducción automática, el resumen de textos o la transcripción de escritura manuscrita. Se da incluso la paradoja de que procesos técnicamente similares, como la transcripción automática de manuscritos y el reconocimiento óptico de caracteres (OCR), se adscriban hoy a categorías distintas, al considerarse este último una tecnología ya consolidada.

Por último, cabe recordar que, al igual que ocurrió con tecnologías anteriores, la BNE debe integrar estas nuevas herramientas sin dejar de atender sus funciones esenciales: asegurar el ingreso de cerca de medio millón de documentos físicos al año, procesar sus colecciones, prestar servicios presenciales y mantener una infraestructura digital cada vez más compleja.

Una consideración inicial: el papel de la BNE en el desarrollo de la IA

En este texto se exponen algunos trabajos y reflexiones que la BNE está desarrollando en torno a la inteligencia artificial. En tanto herramientas tecnológicas, su potencial es enorme; sin embargo, antes de continuar conviene subrayar que en torno a la IA existen importantes

¹ En 2023 la BNE publicó un documento sobre estrategia digital en el que se resume un panorama general de estas cuestiones: https://www.bne.es/sites/default/files/repositorio-archivos/2023_estrategia_digital_bne_1.pdf

consideraciones éticas, legales y sociales. El propio desarrollo de muchas de estas herramientas está siendo cuestionado por haberse empleado, para su entrenamiento, contenidos sujetos a derechos de propiedad intelectual sin la autorización de sus titulares.

Precisamente en este punto la BNE —y, en general, las instituciones culturales— puede desempeñar un papel fundamental en el desarrollo de una inteligencia artificial legal y ética: recopilación y uso de corpus de forma legítima, conocimiento y corrección de sesgos, igualdad de acceso, transparencia, etc. En el caso de la BNE, la institución cuenta con el enorme corpus documental que le proporciona el depósito legal. Desde hace años recoge y preserva tanto la web española como publicaciones nacidas digitales y, más recientemente, las versiones digitales de libros y publicaciones periódicas impresas. Esta realidad, unida a la digitalización de colecciones históricas y a la diversidad tipológica de los fondos conservados, sitúa a la BNE en una posición singular como potencial proveedora de grandes conjuntos de datos.

Las incertidumbres legales, el compromiso de la BNE con el respeto a la propiedad intelectual y la necesidad de acuerdo con los agentes implicados hacen que, hasta el momento, no se haya establecido una forma de acceso generalizado a estos corpus. No obstante, cabe pensar que en el futuro la BNE pueda convertirse en uno de los grandes proveedores para el entrenamiento y desarrollo de herramientas de IA con las debidas garantías.

Primeras experiencias con IA en la BNE

Un portal semántico: datos.bne.es²

La posibilidad de emplear herramientas que contribuyan a mejorar los procesos internos, la gestión de la información o los servicios a los usuarios resulta de gran interés para la Biblioteca. En este sentido, un proyecto ya consolidado como el catálogo semántico datos.bne.es —presentado en 2011— puede considerarse un antecedente relevante en la incorporación de tecnologías asociadas a la inteligencia artificial.

Aunque existen importantes diferencias con el concepto de IA tal como se entiende hoy, datos.bne.es supuso un cambio cualitativo en la forma de interrogar el catálogo, al basarse en modelos conceptuales como FRBR y en la estructuración de los datos en formato MARC. Por primera vez, una base de datos de la BNE era capaz de manejar conceptos como “obra”, “edición” y “ejemplar”, y de ofrecer resultados que priorizaban las obras frente a la enumeración exhaustiva de ediciones. Asimismo, el sistema comprende que un autor es una persona que mantiene distintas relaciones con las obras.

Así, la búsqueda de “Emilia Pardo Bazán” en datos.bne.es ofrece un resultado más claro y estructurado que el del catálogo bibliográfico tradicional, que sigue siendo imprescindible como soporte de los procesos técnicos, pero que no estaba concebido para este tipo de exploración conceptual.

² <https://datos.bne.es/inicio.html>

En la BNE

Autor de
145
Obras

Tema en
46
Obras

Participa en
70
Ediciones

Sus obras (Ver todas las ediciones)

Los pazos de Ulloa Pardo Bazán, Emilia, (68 ediciones) Obra	Insolación Pardo Bazán, Emilia, (35 ediciones) Obra	La cocina española antigua Pardo Bazán, Emilia, (10 ediciones) Obra
La cuestión palpitante Pardo Bazán, Emilia, (8 ediciones) Obra	Cuentos de amor Pardo Bazán, Emilia, (8 ediciones) Obra	La madre naturaleza Pardo Bazán, Emilia, (16 ediciones) Obra

Imagen de los resultados de “Emilia Pardo Bazán” en datos.bne.es.

emilia pardo bazán

Catálogo de la biblioteca

BÚSQUEDA AVANZADA

Todos Libros Periódicos Revistas Material gráfico Registros sonoros Partituras Más

0 seleccionado PÁGINA 1 1-10 de 1237 Resultados

Filtrar los resultados

Ordenar por Relevancia

Disponibilidad

Disponibilidad en la biblioteca

Disponibilidad en línea

Autor/Creador

Materia

Género-Forma

Datos de publicación

Lugar

Fecha de creación

Tipo de recurso

Lengua de la publicación

Lengua original

Ubicación

- LIBRO

Emilia Pardo Bazán / Cristina Fernández Cubas

Fernández Cubas, Cristina (1945-)

2001

Disponibilidad en Sede Central (Paseo de Recoletos) Salón General. Referencia (SB-4 929 PAR) y otras ubicaciones
- LIBRO

Emilia Pardo Bazán / selección y prólogo de Carmen Castro

Pardo Bazán, Emilia, Condesa de (1851-1921)

Castro, Carmen (1912-1997)

1945

Disponibilidad en Sede Central (Paseo de Recoletos) Salón General (4/235390) y otras ubicaciones
- MONOGRAFÍA

Emilia Pardo Bazán / Ramón Gómez de la Serna ; traducción de María Fe González Fernández

Gómez de la Serna, Ramón (1888-1963)

2009

Disponibilidad en Sede de Alcalá de Henares Salón General-Petición anticipada (12/668476) y otras ubicaciones
- MONOGRAFÍA

Emilia Pardo Bazán / guion, alicia Palmer ; dibujo, Weronika Hańczyc ; artículo, Patricia Angulo Miró y Xulia Saantiso

Imagen de los resultados de “Emilia Pardo Bazán” en el catálogo de la BNE.

Reconocimiento de imágenes

Centrándonos en herramientas recientes, asociadas a lo que hoy se entiende como IA, la BNE en los últimos años ha desarrollado algunos proyectos piloto para evaluar la posibilidad de implantar diferentes tecnologías en algunos procesos.

El primero de ellos fue un prototipo de aplicación de un algoritmo de reconocimiento facial, originalmente entrenado para su uso en material audiovisual moderno. El proyecto se enmarcó en un contexto pre-IA generativa (último trimestre de 2022), anterior a la expansión masiva de aplicaciones y herramientas basadas en inteligencia artificial. El algoritmo —basado en patrones biométricos— se aplicó a un conjunto acotado de revistas de cine de la década de 1930, seleccionadas por la abundancia y calidad de los retratos, así como por la homogeneidad del estilo visual y de los soportes materiales. En total, se analizaron aproximadamente 8.500 páginas, logrando definir el patrón biométrico de más de 221 actores y actrices mediante la creación de una base de datos de referencia a la que las imágenes remitían para establecer coincidencias.

Los resultados fueron muy satisfactorios y permitió vislumbrar el potencial de las técnicas algorítmicas para el enriquecimiento de los sistemas de búsqueda y recuperación de información. No obstante, las limitaciones se hicieron evidentes al intentar escalar el modelo a materiales menos homogéneos, así como por el elevado grado de trabajo manual requerido para la creación de los patrones de referencia. A ello se sumó, poco después, un punto de inflexión clave en el desarrollo de la IA: en noviembre de 2022 se produjo el lanzamiento de ChatGPT y, a comienzos de 2023, la inteligencia artificial generativa se integró de forma masiva en múltiples aplicaciones, desde herramientas de búsqueda hasta sistemas de creación e identificación de imágenes. Este nuevo escenario abrió vías mucho más amplias para el desarrollo de estas aplicaciones, con mayores posibilidades de automatización y escalabilidad, que se espera poder abordar en el futuro.

Catalogación automática

Otro aspecto que la BNE ha estado estudiando en los últimos años es la catalogación automática de libros, una actividad esencial para la BNE, que recibe anualmente en torno a 70.000 títulos por depósito legal. Las pruebas realizadas han demostrado que existen herramientas capaces de generar registros bibliográficos en formato MARC21 con un nivel razonable de calidad³.

La dificultad no reside tanto en la capacidad técnica de los modelos —que previsiblemente seguirá mejorando⁴— como en la integración de estos procesos en el flujo ordinario de trabajo de la Biblioteca. La modificación de la Ley de Depósito Legal en 2022, que incorpora las versiones digitales de libros impresos, ofrece a la BNE una oportunidad singular. A diferencia de otras bibliotecas, no sería necesario recurrir a la captura fotográfica de los ejemplares, aunque sí deben resolverse cuestiones clave, como el respeto a la propiedad intelectual.

³ El proyecto elaborado por la Biblioteca Nacional de Bélgica (<https://repository.ifla.org/rest/api/core/bitstreams/20c641b3-b285-4be1-ac35-12540150f208/content>) es uno de los más conocidos

⁴ Se puede ver la presentación de Ricardo Santos en Bibliotecas, datos, inteligencia artificial: las nuevas rutas del conocimiento: <https://www.youtube.com/watch?v=76uQN8jMenU> a partir de las cuatro horas y catorce minutos.

La dificultad no reside tanto en la capacidad técnica de los modelos —que previsiblemente seguirá mejorando— como en la integración de estos procesos en el flujo ordinario de trabajo de la Biblioteca. La modificación de la Ley de Depósito Legal en 2022, que incorpora las versiones digitales de libros impresos, ofrece a la BNE una oportunidad singular. A diferencia de otras bibliotecas, no sería necesario recurrir a la captura fotográfica de los ejemplares, aunque sí deben resolverse cuestiones clave, como el respeto a la propiedad intelectual.

Por este motivo, la BNE valora el desarrollo de herramientas instaladas localmente, sin transferencia de datos fuera de su entorno. Asimismo, se ha comprobado que no es imprescindible utilizar el texto completo, ya que la cubierta, la portada y algunas páginas iniciales suelen ser suficientes. En cualquier caso, la catalogación automática no se concibe como una sustitución del trabajo de los técnicos, sino como una fase previa de apoyo.

Transcripción de manuscritos

Más recientemente la BNE también ha comenzado a probar una herramienta de transcripción automática de manuscritos con el fin de que estos textos puedan ser indexados por las bases de datos y ampliar así las capacidades de recuperación de información, que ya no sería sólo sobre los metadatos. Desde la Biblioteca se ha colaborado en el pasado con diversos grupos de investigación que buscaban desarrollar este tipo de tecnología, pero ha sido en 2025 cuando se ha iniciado un proyecto propio de instalación de un programa que facilita esta transcripción. En este caso, la BNE ha optado por hacer este proyecto con eScriptorium, *software* libre que diversos grupos de investigación próximos a la BNE estaban empleando con bastante éxito.

Un factor clave en esta tecnología es elaborar modelos de transcripción que reconozcan los diferentes tipos de letra manuscrita. Esto requiere una inversión de tiempo en diseñar los *corpus* y entrenar los modelos. El propósito de la BNE es no sólo lograr indexar el texto de sus manuscritos sino compartir en abierto los modelos de transcripción que vaya elaborando para tal fin, de forma que otras instituciones o investigadores puedan aplicar estos modelos sobre otros manuscritos.

Como en los casos anteriores, la puesta en marcha de este proyecto se enfrenta al riesgo de la escala de las colecciones, si bien en este caso parece más sencillo lograr en un plazo más corto ir indexando conjuntos de manuscritos con tipos de letra similar e integrar esto en la biblioteca digital de la BNE (BNEdigital).

Apoyo a los investigadores

Aunque el acceso a grandes *corpus* de datos aún no se ha generalizado debido a las incertidumbres legales señaladas anteriormente, la BNE sí ha podido desempeñar en estos años un papel relevante de impulso y apoyo al desarrollo de la inteligencia artificial mediante la puesta a disposición de aquellas colecciones y datos que se encuentran en dominio público. Son ya numerosos e importantes los proyectos de investigación académica que se están desarrollando gracias a los contenidos digitales que la Biblioteca ofrece en abierto.

Quizá el caso más conocido sea la atribución a Lope de Vega de una obra teatral que hasta entonces se consideraba anónima, gracias al uso combinado de herramientas de transcripción de manuscritos y técnicas de estilometría, sin olvidar el imprescindible análisis filológico. Desde hace años, la BNE apuesta de forma decidida por la publicación abierta y reutilizable de sus datos

y colecciones, acompañándolos de licencias de uso amplias, con la única exigencia de citar la procedencia de los contenidos.

La relación de la BNE con la inteligencia artificial no se limita al uso directo de herramientas, sino que se concibe también como una institución que, a través de sus colecciones, servicios y políticas de acceso abierto, facilita que investigadores y otros agentes desarrollen sus propios estudios, metodologías y aplicaciones, contribuyendo así al avance del conocimiento en un marco legal y ético.

La BNE como espacio de reflexión

Finalmente, conviene destacar el papel que la BNE ha desempeñado como espacio de reflexión y debate en torno a la inteligencia artificial y, más ampliamente, a la transformación digital en el ámbito cultural. En los últimos años se han celebrado en la institución diversos encuentros profesionales, desde los vinculados a las humanidades digitales hasta otros centrados específicamente en la inteligencia artificial y sus implicaciones para bibliotecas, archivos y museos.

La BNE participa además en foros y estructuras de cooperación internacional, como el grupo de trabajo sobre inteligencia artificial de la Conferencia de Directores de Bibliotecas Nacionales Europeas (CENL⁵) o la infraestructura de investigación digital CLARIAH⁶, que integra las redes europeas CLARIN (recursos lingüísticos) y DARIAH (artes y humanidades). Estos espacios permiten compartir experiencias, identificar buenas prácticas y reflexionar de manera colectiva sobre los retos comunes.

Dadas las profundas implicaciones sociales, culturales y éticas que plantea la inteligencia artificial, resulta esencial que su desarrollo incorpore la perspectiva de instituciones culturales públicas, capaces de aportar una visión basada en la preservación del patrimonio, el acceso equitativo a la información y la defensa del interés general

Conclusión: perspectivas de futuro

La inteligencia artificial abre, sin duda, posibilidades muy interesantes para la BNE, si bien su impacto efectivo en los procesos y servicios es todavía limitado por diversos factores. Por una parte, la escala y complejidad de las colecciones, así como de la infraestructura digital de la institución, hacen que la implantación de proyectos de alcance general sea necesariamente compleja y costosa. Por otra, la rápida evolución de las herramientas aconseja actuar con prudencia a la hora de adoptar soluciones que pueden quedar obsoletas en un corto periodo de tiempo.

A ello se suman las incertidumbres legales y éticas, que condicionan el ritmo de avance y refuerzan la necesidad de un enfoque cauteloso. En este sentido, para la BNE resulta prioritario avanzar sobre bases sólidas y sostenibles, incluso si ello implica un desarrollo más gradual. Todo esto debe realizarse, además, en paralelo al mantenimiento y mejora del conjunto de los

⁵ Precisamente el año pasado la reunión anual trató el tema de la inteligencia artificial en bibliotecas y, como resultado se publicó un documento sobre la IA en instituciones culturales: <https://www.cenl.org/wp-content/uploads/2026/01/ArtificialIntelligenceMeetsCulturalHeritage.pdf>

⁶ <https://www.clariah.es/>

procesos y servicios de la institución. En los últimos años, la Biblioteca ha acometido la migración de su web, su catálogo, su biblioteca y hemeroteca digital, así como el desarrollo de nuevas plataformas para el depósito de versiones digitales de libros, prensa y revistas, entre otros proyectos. Estos entornos son, por otra parte, la infraestructura que permite el desarrollo y aplicación de herramientas de inteligencia artificial.

En el horizonte se vislumbran transformaciones profundas en los modos de procesamiento, búsqueda y recuperación de la información: catalogación y clasificación automáticas, búsqueda sobre el contenido de imágenes, textos manuscritos o palabra hablada, así como consultas en lenguaje natural sobre el contenido completo de los documentos. Estos avances ya comienzan a ser una realidad en otros ámbitos y previsiblemente se irán incorporando de forma progresiva a los sistemas de las bibliotecas.

Es posible, incluso, que el acceso a los datos y contenidos de la Biblioteca se produzca cada vez más a través de intermediarios externos, como buscadores generales u otras plataformas, lo que obliga a la BNE a reforzar su papel como fuente de referencia y autoridad en el ecosistema informativo digital.

Por último, en el futuro la BNE podrá desempeñar un papel determinante como proveedora de grandes volúmenes de contenidos para el entrenamiento y desarrollo de herramientas de inteligencia artificial en un entorno legal y controlado. Desde esta posición, su contribución al desarrollo de una IA ética, transparente y con una clara finalidad social puede resultar especialmente significativa.

La inteligencia artificial nos enfrenta a grandes retos profesionales e institucionales. La Biblioteca Nacional de España está llamada a desempeñar un papel activo en la construcción de una inteligencia artificial ética y sostenible, que amplíe las posibilidades de acceso al conocimiento sin comprometer los valores fundamentales de una institución pública dedicada a la preservación y difusión del patrimonio bibliográfico español.

Agradecimientos

Este artículo resume la intervención presentada en la XII Jornada Actualízate: Bajo la superficie de la IA: lo que no vemos pero sí nos afecta.

Sobre el autor



JOSÉ LUIS BUEREN GÓMEZ-ACEBO

Director Técnico de la Biblioteca Nacional de España

Licenciado en Filología Hispánica por la Universidad Complutense de Madrid en 2004, en 2009 obtuvo el Diploma de Estudios Avanzados. Perteneció al Cuerpo Facultativo de Bibliotecas y está destinado desde 2006 en la Biblioteca Nacional de España. Entre 2014 y 2017 fue Director de Biblioteca Digital y Sistemas de Información y desde entonces ocupa la Dirección Técnica.