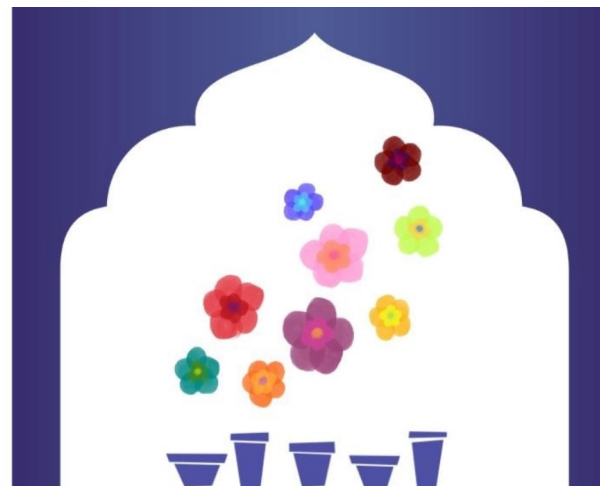


## IV Encuentro de profesionales de la edición científica y académica

*Inteligencia Editorial: optimización y herramientas  
prácticas para la ciencia en abierto*

17 y 18 de septiembre Universidad de Córdoba



## Situaciones conflictivas generadas tras la irrupción de la IA:

- derechos de autor
- autorización de uso de contenidos por motores de IA
- control de protocolos de recolección

Nacho Pérez Alcalde



## ExperienciaS de Editorial CSIC con “la IA”

Todos trabajamos para instituciones que quieren “estar en esto de la IA” y las editoriales generamos recursos de gran valor para ello, **somos un activo importante**

**Pero somos lo que somos:** servicios pequeños, con poco personal y pocos recursos y necesitamos ayuda, fundamentalmente, en dos aspectos: jurídico y técnico

Y a pesar de ello, hacemos lo que podemos:

# Mayo de 2025: publicamos nuestra Política sobre uso de IA

Hay un resumen en la web de cada revista: “Acerca de”

Trabajos de Prehistoria

Registrar Entrar

Acerca de

Equipo editorial

Indexación

Envíos

Anuncios

Últimos números

Archivo

Revistas CSIC

BUSCAR

Enfoque y alcance

Secciones de la revista

Frecuencia de publicación

Proceso de evaluación por pares

Revisores/as externos/as

Política antiplagio

Política de acceso y reutilización de contenidos

Política de depósito de datos de investigación

Política de desautorización de artículos

Recomendaciones lenguaje inclusivo y no sexista

Política de preservación de archivos digitales

Política sobre uso de inteligencia artificial

Estadísticas de producción editorial

Propiedad y gobernanza de las revistas

Protocolos de interoperabilidad

Distribución y venta

Declaración de privacidad

Política de cookies

Diseño de cubierta

Contacto

CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Enviar un artículo

EDITORIAL

CSIC

eISSN: 1988-3218

ISSN-L: 0082-5638

DOI: 10.3989/tp

Idioma

English

Español (Español)

Indexación y calidad

Journal Citation Reports®

Clarivate®

2024

Journal Impact Factor (JIF) 2024

JIF (2 años): 1.3

JIF (5 años): 1.2

Q2 (41/141)

Anthropology

Fuente: Clarivate Analytics®, Journal Citation Reports®

Journal Citation Indicator (JCI) 2024

JCI: 1.40

Q1 (19/142)

Anthropology

Q1 (17/168)

Archaeology

Fuente: Clarivate Analytics®, Journal Citation Reports®

sta científica publicada por el CSIC, editada en el Instituto de Historia del CCHS. La revista obre Prehistoria y Protohistoria de la península ibérica y sus relaciones con Europa y el e otros ámbitos geográficos de interés general por su enfoque teórico o metodológico o por unidos arqueológicos. Además, promueve temas de teoría, metodología, arqueometría y ibién por aspectos conectados con la gestión del Patrimonio Arqueológico. Contiene una científica.

r. Martín Almagro Basch, tuvo hasta 1968 forma de serie monográfica y, entre esa fecha y l se convierte en publicación semestral. Comienza a estar disponible online en 2007, en sn impresa hasta 2023, año en el que pasa a ser revista electrónica publicando en formato dos anteriores están igualmente disponibles en formato PDF.

a en Web of Science: Journal Citation Reports (JCR), Social Sciences Citation Index (SSCI) y -HCI); SCOPUS, CWTS Leiden Ranking (Journal Indicators), ERIH Plus, REDIB, DOAJ y otras acionales. Está incluida en el Catálogo Latindex 2.0 y cuenta con el Sello de Calidad de la

Revisión por pares

Código ético

Detección de plagio

Identificadores digitales

Interoperabilidad

Preservación digital

Depósito de datos

Edición impresa y PDF

Pr

2025-09-30

El Dr. Ignacio de la Torre Salazar recibe el Premio Nacional de Investigación 'Ramón Menéndez Pidal', en el área de Humanidades.

2025-06-20

Fallecimiento de la Doctora Rubí Sanz Gamo

LEER MÁS

## Trabajos de Prehistoria

Acerca de ▾ Equipo editorial Indexación Envíos ▾ Anuncios Últimos números ▾ Archivo ▾ Revistas CSIC

INICIO / Política sobre uso de inteligencia artificial

### Política sobre uso de inteligencia artificial

Editorial CSIC, tomando como principal referencia las recomendaciones del [Committee on Publication Ethics](#) (COPE), entiende que el uso de inteligencia artificial (IA) y de herramientas asistidas por IA puede suponer una tecnología auxiliar, un apoyo útil, siempre que se utilice con responsabilidad y prudencia, se declare su uso de forma detallada, y sus resultados sean revisados, corregidos y aceptados por seres humanos. Este deber de transparencia, además, está en consonancia con los principios de ciencia abierta, que fomentan la trazabilidad y reproducibilidad de la investigación.

El uso ético y responsable de estas tecnologías obliga a establecer siempre un control y supervisión por parte de quienes asumen la autoría, la edición o la revisión. Estas personas deben verificar la exactitud, precisión, validez, pertinencia e integridad del material generado por la IA, prestando especial atención (aunque no solo) a las referencias inventadas y a los sesgos que pudiera introducir el diseño de la herramienta (inconsistencias, falta de información o de fuentes, etc.), para evitar la desinformación. La rendición de cuentas sobre el manejo de herramientas de IA debe garantizar el análisis y declarar las limitaciones que se puedan detectar.

Lo más común en el ámbito de las publicaciones científicas son las IA generativas y los modelos de lenguaje de gran tamaño. Sin embargo, hay otros muchos tipos de IA que pueden utilizarse en los procesos de redacción del manuscrito o de edición y producción de la publicación, e incluso antes, en el diseño o el desarrollo de la investigación.

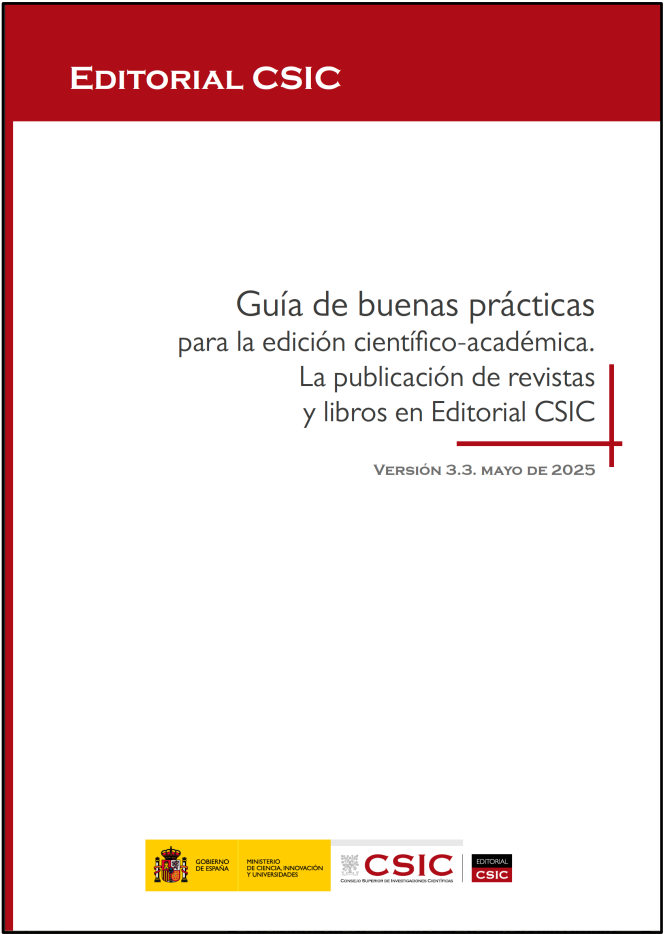
Existe una gran diversidad de herramientas y tecnologías que utilizan la IA para la edición, y muchas permiten proteger el contenido que agregamos en ellas. Se deben utilizar aquellas que permitan encriptar la comunicación o la información almacenada, denegar la autorización para que el modelo se entrene con nuestros contenidos o trabajar en entornos controlados, ya sea en una instalación local o en servidores con garantías de privacidad, así como cualquier otra medida de protección.

Para conocer en detalle la política de Editorial CSIC sobre uso de inteligencia artificial consulte la [Guía de buenas prácticas para la edición científico-académica](#) en la que se describen los conceptos de **autoría en relación con la IA**, quiénes y cómo se debe **declarar el uso de IA** y cómo se deben **referenciar y citar las herramientas de IA**.

Desarrollada en:

Guía de buenas prácticas para la edición científico-académica

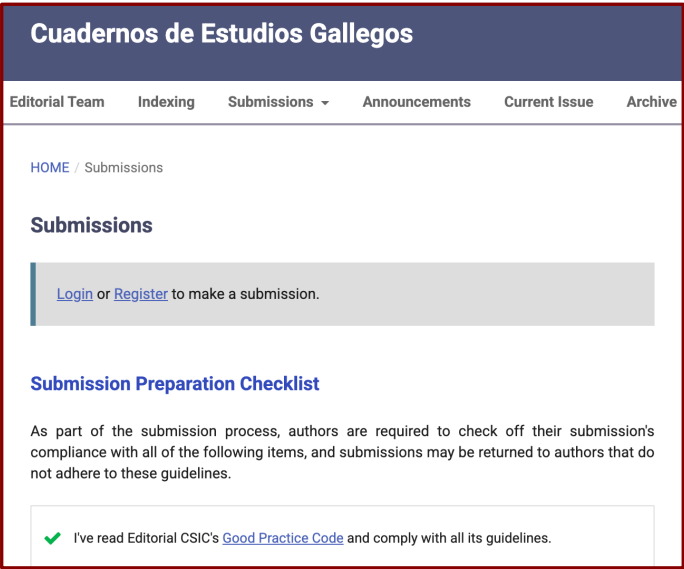
https://revistas.csic.es/public/guia\_buenas\_practicas\_CSIC.pdf



Español

Inglés

Gallego



Esta política editorial nace como respuesta a una demanda de nuestros equipos editoriales, para saber qué pueden hacer con IA pero, sobre todo, por el miedo a qué pueden hacer los autores

El personal de la editorial también empieza a utilizarla, y tiene esa misma necesidad

Pero antes de regular este uso, de hecho, unos años antes, nos enfrentamos ya a la posibilidad de ceder nuestros contenidos para el entrenamiento de herramientas de IA por parte de terceros:



Colaboración con el Barcelona Supercomputing Center y la SEDIA

## CONTEXTO

- En 2021 nos contacta la *Unidad de Text Mining* del BSC
- Estaban trabajando, en el marco del Plan-TL, en el desarrollo de aplicaciones de IA y tecnologías del lenguaje
- Para poder aplicar técnicas de aprendizaje profundo necesitaban **datos masivos**
- Disponían de datos suministrados por la BN y otras fuentes, pero les faltaba "**literatura científica**" y nos identificaron como **fuentes fiables** que, además, incluía títulos en las lenguas oficiales del estado
- También querían **utilizar nuestros datos** para entrenamiento de clasificación de textos y como **Gold Standard** para evaluación de modelos de la lengua

## “NEGOCIACIÓN”

- Varias reuniones con el BSC y la SEDIA (confianza mutua y ánimo de colaboración)
- Fue hace cinco años... nos planteamos:
  - El uso que darían a nuestros contenidos: ¿sólo para entrenar los modelos?, ¿en entorno controlado y limitado?...
  - ¿Era necesario firmar algún tipo de acuerdo o convenio?
  - ¿Cómo se haría la transferencia de archivos?

## “NEGOCIACIÓN”

- Se encauzó como una colaboración entre ministerios (Transformación Digital y de la Función Pública y Ciencia e Innovación)
- Se limitó el uso al entrenamiento de sus herramientas en un entorno controlado
- Plantearon el envío de PDF y les ofrecimos cosecha de PDF, HTML y XML-JATS + envío de CSV de control
- Nos ofrecieron su colaboración para hacer algún desarrollo para nosotros

## ¿Colaborar con otras entidades?

- Hemos cedido contenidos para entrenar otros modelos de IA dentro del CSIC pero, ¿qué debemos hacer si lo demandan otras entidades?
- Sin duda, empresas privadas han utilizado ya nuestros (y vuestros) contenidos sin pedir permiso ni dar nada a cambio
- Han pasado unos años y se conoce un poco más el contexto, pero seguimos teniendo grandes carencias en conocimientos jurídicos (a pesar de que ya existe legislación)
- También a nivel técnico. La IA avanza muy rápido, es muy complejo conocerlas en profundidad y esto hace que se genere desconfianza.
- En otras revoluciones previas pudimos tener más control (paso del papel al electrónico); esta revolución es global y mucho más rápida (no hay más que ver las noticias)

**Y a pesar de todo, las editoriales debemos estar ahí, tenemos mucho que aportar y llevamos años trabajando en un entorno favorable:**

- ✓ **somos fuentes confiables y de prestigio**
- ✓ **publicamos en electrónico (en varios formatos)**
- ✓ **aplicamos selección de contenidos y curación de datos**
- ✓ **generamos y difundimos metadatos estandarizados**
- ✓ **trabajamos con protocolos técnicos y científicos estandarizados**
- ✓ **utilizamos identificadores persistentes**
- ✓ **utilizamos plataformas especializadas homologadas**
- ✓ **etc.**

**Pero debemos aclarar muchas dudas. A modo de ejemplo, comento un par de ellas:**

## **1. Licencias Creative Commons**

- Aunque no lo digan expresamente, ¿permiten el uso de los contenidos para entrenar IA?
- ¿Se cumple con la atribución solamente con mencionar la obra entre las fuentes utilizadas?
- ¿Se comprueba si la licencia aplica a toda la obra (incluidas, por ejemplo, las imágenes)?
- El contenido licenciado, ¿podría aparecer ampliamente reproducido en las respuestas de la herramienta?

## **2. Cómo aplica la “excepción de minería de textos y datos”, ¿es suficiente con declarar expresamente su limitación?**

# Ideas sobre el control de protocolos de recolección

Todos debemos planificar y desarrollar una estrategia sobre recolección para **TDM (Text and data mining)** que permita proteger el patrimonio, pero también la indexación

**Antes de la planificación surge la urgencia. Nuestro caso (compartido):**

- Aumento de la actividad que ralentiza e incluso bloquea páginas web
- Cortamos el acceso a rangos de IPs de los bots más agresivos (Google, Microsoft, Meta/Facebook)
- Google contacta con nosotros > ajustamos acceso IPs por días y horas
- No pueden/quieren ajustar sus bots > ajustamos acceso IPs limitando solo número de accesos y peticiones por minuto
- Algunos bots dejan abierta la petición para reutilizarla > ajustamos el Apache para servir la petición y cerrar el proceso sin petición del cliente
- **Problema añadido:** la actividad del bot interfiere con el plugin OJS de “artículos publicados por un mismo autor”

# Una vez atendida la urgencia podemos planificar

## 1. Usos de IA que queremos permitir o limitar

- Qué contenidos queremos reservar
- ¿Permitimos indexación para respuestas de IA generativa, entrenamiento de sistemas de IA, extracciones masivas para *data mining*, etc.? teniendo en cuenta que “denegar la indexación” no es lo mismo que “denegar el uso para entrenar modelos”
- Qué licencias aplican y quién gestiona la concesión de esas licencias

Equilibrio para evitar una política excesivamente genérica o una excesivamente restrictiva

## 2. Declaraciones y avisos legibles máquinas

### ➤ **Robots.txt**

texto plano con indicaciones genéricas o bastante concretas sobre lo que el bot puede o no hacer. NO es una barrera técnica, pero puede ayudar a agilizar el tráfico y es una declaración a la que remitirnos en caso de reclamación

### ➤ **Otras señales TDM** en varios niveles:

- metadatos en el servidor
- metadatos en el archivo del artículo (propiedades del PDF o etiquetas HTML/XML)
- sección en la web con “términos y condiciones de uso”

# Planificar

## 3. Asociar las condiciones legales a cada obra, no solo de forma genérica a todo el sitio web

- Distintas limitaciones dentro de nuestra web:

libros ≠ revistas

contenido actual ≠ contenido histórico

- Tras su recolección la obra puede quedar descontextualizada del sitio web

# Planificar

## 4. Declarar usos permitidos, prohibidos o sujetos a una licencia específica

Podemos establecer excepciones y licencias particulares, por ejemplo:

- Uso educativo: permitido
- Uso comercial: prohibido salvo licencia específica
- Entrenamiento de modelos: prohibido salvo licencia específica
- Indexación convencional: permitida
- Acceso mediante API: permitido bajo condiciones
- Uso para RAG: permitido bajo condiciones

“Bajo condiciones” o “salvo licencia específica” obligan a desarrollar cuáles son esas condiciones y licencias (y cómo obtenerlas)

## 5. Adoptar una política activa de vigilancia y rastreo

- Nos permitiría conocer datos importantes: qué obras interesan más a “nuestros lectores”, pero también a “nuestros bots”
- Ayuda a mejorar la seguridad y estabilidad de nuestros sitios web
- Podría ser importante si queremos demostrar que un tercero ignoró una reserva de derechos
- La normativa europea exige a los desarrolladores de IA publicar un resumen suficientemente detallado del contenido utilizado para el entrenamiento; deberíamos consultarlos

**Muchas gracias**

[ji.p.alcalde@csic.es](mailto:ji.p.alcalde@csic.es)