

Política de Cumplimiento de Derechos de Autor para Modelos de IA de Propósito General

Fecha de entrada en vigor: 00.00.0000

1. Definiciones

En la presente Política de Cumplimiento de Derechos de Autor, los siguientes términos tendrán el significado que se indica a continuación:

“Nosotros” o “Nuestra” se refiere a [NOMBRE DE LA EMPRESA].

“Modelo de IA de propósito general” o “Modelo GPAI” tiene el mismo significado que en el Reglamento (UE) 2024/1689, artículo 3(63) (Ley de IA).

“Contenido accesible legalmente” significa aquel contenido al que puede accederse de forma legal sin eludir medidas de protección tecnológica.

“Reserva de derechos” significa la manifestación por parte de los titulares de derechos que prohíbe la minería de textos/datos o la reutilización mediante medios legibles por máquina, como, entre otros, los archivos *robots.txt*.

“Lista de infractores persistentes” significa la lista de sitios identificados por las autoridades o tribunales de la UE/EEE como dedicados de forma reiterada a la infracción a escala comercial.

2. Finalidad

La presente política establece la gobernanza, los procedimientos y los controles técnicos necesarios para garantizar que los Modelos de IA de Propósito General (GPAI) de [NOMBRE DE LA EMPRESA] se desarrollan, entrenan, despliegan y mantienen en plena conformidad con la legislación de derechos de autor de la Unión Europea y de los Estados miembros de la UE, en particular, aunque no de forma exclusiva, con la Directiva 2001/29/CE, la Directiva (UE) 2019/790 y la Directiva 2004/48/CE, de conformidad con el artículo 53 del Reglamento (UE) 2024/1689.

Esta obra ha sido liberada al dominio público bajo la licencia Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication.

3. Ámbito de aplicación

La presente política se aplica a todos nuestros Modelos GPAI comercializados o puestos a disposición en el mercado único de la UE. Asimismo, se aplica a todos nuestros conjuntos de datos, herramientas y procesos utilizados para el entrenamiento, ajuste fino (*fine-tuning*), evaluación o inferencia de modelos, y resulta vinculante para todo nuestro personal, contratistas y socios externos implicados en el desarrollo o explotación de GPAI.

4. Uso exclusivo de Contenido Accesible Legalmente

Nuestros rastreadores web (*web crawlers*) y/o demás tecnologías únicamente accederán, reproducirán o extraerán contenido que sea legalmente accesible y no eludirán medidas de protección tecnológica, como aquellas implementadas por los titulares de derechos mediante mecanismos de control de acceso o protección, incluyendo cifrado, ofuscación u otras transformaciones de la obra u objeto protegido, o cualquier otro mecanismo de control de copia, siempre que cumplan la finalidad de protección definida en el artículo 6(3) de la Directiva 2001/29/CE.

[NOMBRE DE LA EMPRESA] excluirá del rastreo, acceso e indexación para fines de entrenamiento de sus Modelos GPAI cualquier sitio web reconocido por tribunales o autoridades públicas de la UE o del EEE como Infractor Persistente de derechos de autor a escala comercial.

5. Cumplimiento de Reservas de Derechos

Nuestros rastreadores web y/o demás tecnologías respetarán los protocolos de exclusión de robots (*robots.txt*) según lo especificado en la Request for Comments n° 9309 del *Internet Engineering Task Force* (IETF), así como otros protocolos similares que expresen reservas de derechos en virtud del artículo 4(3) de la Directiva (UE) 2019/790. Adicionalmente, supervisaremos y adoptaremos otros protocolos legibles por máquina, acordes con el estado de la técnica, a medida que surjan de los esfuerzos de estandarización, y divulgaremos públicamente información sobre los rastreadores web y/o demás tecnologías utilizados, su comportamiento frente a *robots.txt* y otros mecanismos relevantes para identificar y cumplir con las reservas de derechos.

6. Inalterabilidad de las Reservas de Derechos

Nada de lo dispuesto en la presente Política de Cumplimiento de Derechos de Autor se interpretará en el sentido de limitar o perjudicar los derechos de los titulares a reservar expresamente el uso de sus obras u otros contenidos protegidos por cualquier medio, incluidos, sin carácter limitativo, las medidas legibles por máquina aplicables a contenidos en línea o cualquier otro mecanismo lícito. Ello incluye el contenido recopilado o rastreado por terceros que eventualmente pueda ser utilizado por nosotros.

7. Mitigación de Resultados Infractores

Implementaremos salvaguardas técnicas adecuadas y proporcionales con el fin de reducir el riesgo de generar resultados que reflejen materiales protegidos por derechos de autor utilizados en el entrenamiento de nuestros Modelos GPAI. Asimismo, nuestros Términos de Servicio prohíben expresamente los usos que infrinjan derechos de autor.

8. Punto de contacto y presentación de reclamaciones

El punto de contacto designado para que los titulares de derechos de autor se dirijan a nuestra organización respecto a posibles infracciones es:

SR./SRA. XYZ

[NOMBRE DE LA EMPRESA]

DIRECCIÓN

CORREO ELECTRÓNICO

CIUDAD / PAÍS

Los titulares de derechos y sus representantes autorizados, incluidas las entidades de gestión colectiva, podrán presentar reclamaciones suficientemente precisas y debidamente fundamentadas sobre nuestro presunto incumplimiento utilizando el formulario de contacto designado, que permite además adjuntar documentación adicional. Actuaremos sobre las reclamaciones recibidas de manera diligente, no arbitraria y en un plazo razonable. Se ignorará toda reclamación manifiestamente infundada o que ya haya sido objeto de respuesta idéntica.

9. Exención de responsabilidad en materia de cumplimiento legal

Reconocemos que la presente política apoya, pero no sustituye, el cumplimiento de la legislación de derechos de autor de la UE o de las legislaciones nacionales de derechos de autor. El cumplimiento legal efectivo sigue siendo de nuestra responsabilidad.

10. Proporcionalidad y aplicabilidad

Nuestras medidas son proporcionales al tamaño y a la naturaleza de nuestras operaciones. Mediante la implementación de esta política, buscamos alinearnos con las expectativas voluntarias, pero jurídicamente relevantes, recogidas en el Código de Buenas Prácticas de la UE para Modelos GPAI.

ANNEX (SAMPLE)

Crawler & Rights Reservation Detection

A. Crawler Configuration Requirements

1. Crawler Identification

- All crawlers must:
 - Use a **unique User-Agent string** identifying [Company Name] and a contact URL/email.
 - Publish crawler technical specifications and robots.txt compliance behavior at a public endpoint (e.g., [https://\[company\]/crawler-info](https://[company]/crawler-info)).

Example (HTTP request header):

User-Agent: CompanyCrawler/1.2 (+<https://example.com/crawler-info>; contact@example.com)

2. Crawl Scope Enforcement

- Implement a **domain allowlist** and **infringer blocklist** at the crawler configuration level.
- Blocklist is synchronized daily from:
 - EUIPO or national IP authority “persistent infringer” lists.
 - Any internal takedown or complaint resolutions that require source blocking.

Example configuration:

crawler:

obey_robots_txt: true

allow_domains:

- example.edu

- example.gov

- example.org

block_domains:

- infringer-site1.com

- infringer-site2.net

max_depth: 3

rate_limit: 1 req/sec

3. Access Control Compliance

- Crawler must refuse access to:
 - Pages requiring authentication, CAPTCHAs, or tokens.
 - URLs returning HTTP 401, 402, or 403.
- Crawler may not employ headless browser automation to bypass restrictions without explicit license.

B. Rights-Reservation Detection Workflow

1. Primary Signal: robots.txt Parsing

- Detect User-agent: * rules with Disallow: / or specific path rules.
- Parse X-Robots-Tag HTTP header for nodm, noindex, or equivalent tags.

Sample Python snippet:

```
import requests
from urllib.parse import urljoin

def check_rights_reservation(domain):
    robots_url = urljoin(domain, '/robots.txt')
    resp = requests.get(robots_url, timeout=5)
    if resp.status_code == 200 and ('noai' in resp.text.lower() or 'disallow' in resp.text.lower()):
        return True
    return False
```

2. Secondary Signal: Embedded Metadata

- Check page HTML and media metadata for:
 - IPTC “AI Data Mining” prohibition flags.
 - C2PA manifest claims with “No AI Training” directives.
 - Dublin Core rights and license fields.

Example JSON from C2PA manifest:

```
{
  "assertions": [
    {
      "label": "c2pa.datamining",
      "value": "prohibited"
    }
  ]
}
```

3. Storage & Logging

- Log all rights-reservation detections with:
 - Timestamp
 - URL
 - Detected signal type (robots.txt, HTTP header, metadata)
 - Crawling decision (allow/deny)
- Retain logs for **5 years** in an immutable audit store.

C. API Endpoints for Rights-Reservation Verification

1. Check Single URL

GET /api/v1/rights-reservation?url={url}

Response:

```
{
  "url": "https://example.com/article",
  "rights_reservation": true,
  "signals": ["robots.txt:nodm", "X-Robots-Tag:noindex"]
}
```

2. Bulk Check

POST /api/v1/rights-reservation/bulk

Body:

```
{
  "urls": [
    "https://site1.com/page",
    "https://site2.com/page"
  ]
}
```

D. Audit & Review Procedures

- **Daily:** Automated scan of updated rights-reservation protocols from standardization bodies.
- **Quarterly:** Manual audit of crawler logs to confirm:
 - No overridden blocklist entries.
 - No training ingestion from rights-reserved sources.
- **Annually:** Third-party technical audit with reproducible test URLs.