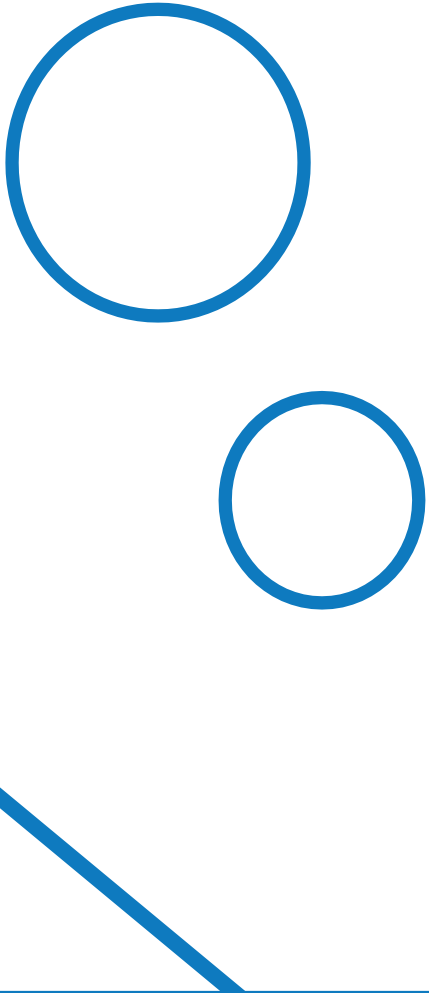




Guía liberación Inteligencia artificial



AXENCIA PARA A
MODERNIZACIÓN
TECNOLÓXICA DE GALICIA

2025. AMTEGA Guía liberación
Inteligencia Artificial.

Esta obra se distribuye con una licencia Atribución
Compartir Igual 4.0 Internacional de Creative Commons.
Para ver una copia de la licencia pulse [aquí](#)

Contenido

01

[¿Qué es una Inteligencia Artificial?](#)

02

[Inteligencia Artificial Generativa](#)

03

[Inteligencia artificial Open Source](#)

04

[¿Cómo liberar una Inteligencia artificial?](#)

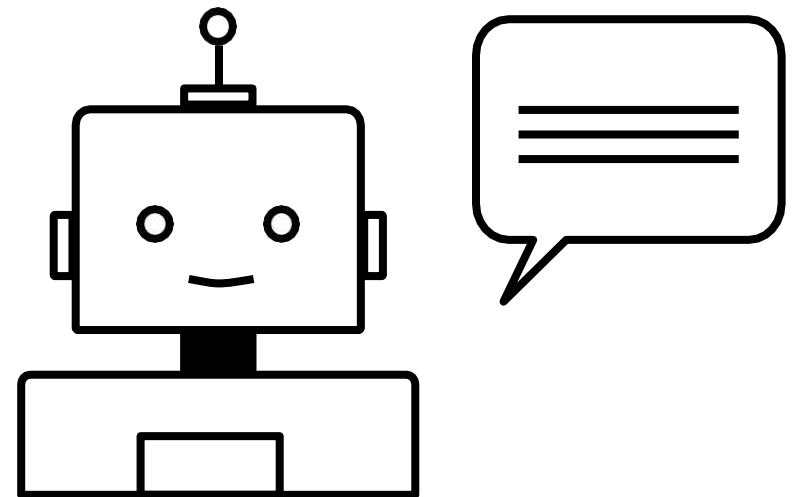
¿Qué es una inteligencia artificial?

La Inteligencia Artificial es una tecnología que permite a los ordenadores y las máquinas **simular** el aprendizaje humano, comprensión, solución de problemas, toma de decisiones, creatividad y autonomía.

Las aplicaciones y dispositivos que usan Inteligencia artificial realizar todo tipo de tareas como:

- Ver e identificar objetos.
- Entender y responder al lenguaje humano
- Aprender de la información y las experiencias nuevas.
- Hacer recomendaciones detalladas a usuarios y expertos.
- Pueden identificar patrones y extraer información relevante.
- Pueden actuar de forma autónoma, sustituyendo la necesidad de inteligencia o intervención humana (como en los coches autónomos).

¡Lleva siendo investigada desde 1950!



Conceptos y tipos de inteligencia artificial.

Inteligencia Artificial (IA) – El campo general

Área de la informática que busca crear sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de voz, la toma de decisiones y el procesamiento del lenguaje natural.

Deep Learning (Aprendizaje Profundo) – Un tipo especializado de Aprendizaje Automático

Dentro del aprendizaje automático, el Deep Learning se refiere a técnicas que usan redes neuronales profundas, con muchas capas de nodos (similares a neuronas), para modelar y resolver tareas complejas.

Aprendizaje Automático (Machine Learning) – Un subconjunto de la IA

Dentro de la Inteligencia Artificial, el aprendizaje automático se centra en desarrollar algoritmos que permiten a las máquinas aprender de los datos y hacer predicciones o tomar decisiones sin ser programadas explícitamente para cada tarea.

Inteligencia Artificial Generativa – Un tipo avanzado de aprendizaje automático

La IA generativa es un tipo de aprendizaje automático que no solo analiza datos, sino que también genera contenido nuevo (texto, imágenes, música, etc.). Se basa en modelos como redes neuronales profundas y técnicas como los modelos generativos adversarios (GANs) o los transformadores (como GPT).

**Década
1950**

Inteligencia artificial

Simulación o exhibición de inteligencia por las máquinas

**Década
1980**

Machine learning

Sistemas de Inteligencia artificial que aprenden de datos históricos

**Década
2010**

Deep learning

Los modelos de Machine Learning copian el funcionamiento del cerebro con "neuronas", por eso les llaman "sistemas neuronales"

**Década
2020**

Inteligencia artificial generativa

Modelos de Deep Learning (modelos fundacionales) que pueden generar contenidos

Inteligencia artificial generativa al detalle

Aunque la Inteligencia Artificial lleva con nosotros desde 1950, desde 2022 han tomado especial relevancia entre los investigadores y el público general las conocidas como "**Inteligencia Artificial Generativa**"

Las Inteligencias Artificiales Generativas se caracterizan por ser una tecnología que genera textos, imágenes, vídeos y otros contenidos.

Estos modelos aprenden los patrones y estructuras de los datos con los que fueron entrenados y los usan para generar nuevos datos basados en la entrada que reciben, generalmente en forma de instrucciones en lenguaje natural.

Para comprender plenamente la Inteligencia Artificial generativa, primero debemos entender que su tecnología se basa en el aprendizaje automático (ML) y el aprendizaje profundo, sólo que va un paso más allá.

Concepto clave

Las Inteligencias Artificiales Generativas requieren grandes conjuntos de datos para ser entrenadas y lleguen a ciertos resultados.

Los dataset para el entrenamiento de IA.

Un dataset o conjunto de datos es una colección organizada de información que se utiliza para realizar análisis, entrenar modelos de inteligencia artificial, hacer investigaciones, entre otras aplicaciones.

La información utilizada es muy variada aunque se englobe como "datos" o información", ya que, en realidad, se trata de un conjunto de números, textos, imágenes o sonidos, algunos datos pueden ser especialmente sensibles, por ser información personal o pueden estar protegidos por propiedad intelectual.

Para poder entrenar a las Inteligencias artificiales los data sets se estructuran, generalmente, en forma de tablas o matrices.

Cada fila de un dataset representa una instancia o ejemplo de los datos (por ejemplo, una persona, una transacción, una canción), y cada columna representa una característica o atributo (por ejemplo, nombre, edad, salario, notas o los píxeles de una imagen). Dependiendo del tipo de problema, los datasets pueden ser etiquetados (con información de salida conocida) o no etiquetados (sin información de salida).

¿Por qué son importantes?

Los data sets es la clave detrás de la precisión de una inteligencia artificial y necesitamos saber que algunos datos despiertan preocupaciones, como los datos personales (nombre, salario, dirección) o por estar protegidos por propiedad intelectual.

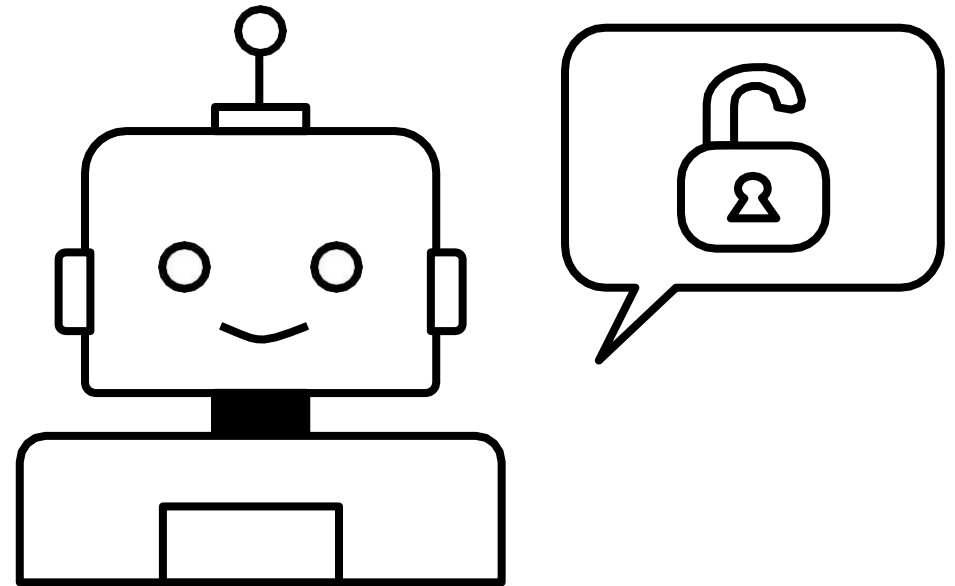
Inteligencia artificial Open Source

El desarrollo de Programas de Ordenador cuenta con código que puede ser usado, estudiado, distribuido, modificado y mejorado de manera libre, lo que llamamos Software libre u Open Source.

Recientemente la Open Source Initiative (OSI) ha lanzado la primera definición de Inteligencia artificial Open Source.

A modo de resumen podemos decir que una **Inteligencia Artificial Open Source** es aquella cuyo código, datos y modelos están disponibles para que cualquiera los use, estudie, modifique y comparta libremente.

Esto permite que la comunidad contribuya a su desarrollo y mejora, fomentando la transparencia y la innovación.



Beneficios de la Inteligencia artificial Open Source

■ **Transparencia y seguridad**

Las Inteligencias artificiales Open Source permite obtener información esencial para auditar el sistema o mitigar los prejuicios heredados de los datos, permite obtener responsabilidad sobre los modelos y la transparencia sobre la fuente de datos, lo que, al final, se traduce en una mejora en la investigación sobre seguridad de la IA

■ **Competencia y diversidad**

La IA Open Source , estimula la innovación y la calidad ya que, cuantos más modelos haya y mayor sea la competencia, mayor será diversidad de soluciones habrá. Esto es beneficioso para evitar los monopolios y los precios elevados de las soluciones, proporcionando a más partes interesadas acceso a la tecnología fundacional.

■ **Aplicaciones variadas**

La IA Open Source ofrece a los desarrolladores acceso a recursos cruciales para desarrollar aplicaciones específicas para cada contexto, además, permite representar una verdadera diversidad cultural y lingüística, evitando actos discriminatorios en el modelo alineado con distintos sistemas de valores.

¿Sirven las licencias tradicionales de Software libre?



La respuesta requiere pensar en como funcionan las Inteligencias artificiales, ya que al tener distintos componentes, como los data sets, las licencias tradicionales de Open Source puede que no sean suficientes para garantizar el derecho a usar, estudiar, compartir o modificar los sistemas de inteligencia artificial

Piensa que, sin los datos de entrenamiento, no podríamos auditar verdaderamente a la inteligencia artificial.

Si se libera la Inteligencia artificial, debería incluir los datos para que sea verdaderamente libre.

Requisitos para Inteligencia artificial Open Source

Según la Open Source Initiative (OSI), para garantizar las libertades del Open Source en un sistema de Inteligencia Artificial se debe tener licencia forma preferida para realizar modificaciones en el sistema, lo que implica tener acceso a:

Código fuente accesible:

La licencia libre u Open Source afecta al código fuente utilizado para entrenar y ejecutar el sistema al completo.
El código incluirá la especificación completa de cómo se procesaron y filtraron los datos y cómo se realizó el entrenamiento.

Parámetros

Los parámetros y pesos del modelo deben poder descargarse y utilizarse libremente.

Información de los Datos de entrenamiento

La información sobre los datos utilizados en el entrenamiento debe ser clara y accesible.

De modo que una persona capacitada pueda construir un sistema sustancialmente equivalente.

La información sobre los datos se facilitará incluyendo:

1. La descripción completa de todos los datos utilizados para el entrenamiento, incluidos los datos no compartibles,
2. La procedencia de los datos, su alcance y características, cómo se obtuvieron y seleccionaron los datos, l
3. Los procesos de selección, filtrado y etiquetado deben ser transparente;
4. Un listado de todos los datos de entrenamiento disponibles públicamente y dónde obtenerlos; y
5. Un listado de todos los datos de entrenamiento obtenibles de terceros y dónde obtenerlos, incluso previo pago.

Como hemos comentado anteriormente, la Inteligencia Artificial tiene el reto especial del conjunto de datos de entrenamiento, ya que estos son fundamentales para su funcionamiento.

¿Cómo liberar una Inteligencia artificial Open Source?

Ya hemos visto que liberar una inteligencia artificial (IA) como Open Source implica garantizar que tanto el sistema completo como sus componentes individuales estén disponibles bajo términos que otorguen las cuatro libertades fundamentales, es decir, que cualquier persona pueda usar, estudiar, modificar y distribuir el sistema.

Para liberar a la inteligencia artificial, realizaremos unos pasos extra a los ya comentados en otras Guías de la Xunta de Galicia.

Resumiremos los pasos y daremos por hecho que aplicarás la misma licencia y sus condiciones.

1 Documentar la información del conjunto de datos

Proporcionar una descripción completa de todos los datos utilizados para entrenar la IA, incluyendo su procedencia, alcance, características, métodos de obtención y selección, procedimientos de etiquetado y metodologías de procesamiento y filtrado. También se debe incluir una lista de todos los datos de entrenamiento disponibles públicamente y dónde obtenerlos, así como de los datos obtenibles de terceros y dónde acceder a ellos, incluso si requieren pago.

2 Liberar el código fuente completo

Hacer disponible el código fuente utilizado para procesar y filtrar los datos, entrenar la IA (incluyendo argumentos y configuraciones utilizadas), validación y pruebas, bibliotecas de soporte como tokenizadores y código de búsqueda de hiperparámetros, código de inferencia y la arquitectura del modelo. Este código debe estar bajo licencias Open Source.

3 Compartir los parámetros del modelo

Proporcionar los parámetros del modelo, como los pesos u otras configuraciones, incluyendo puntos de control de etapas intermedias clave del entrenamiento, así como el estado final del optimizador. Estos parámetros deben estar disponibles bajo licencias Open Source

4 Seleccionar una licencia adecuada

Elegir una licencia libre u Open Source que garantice las cuatro libertades y sean compatibles con tu proyecto. Algunas licencias comunes incluyen la Licencia Pública General de GNU (GPL), la Licencia MIT y la Licencia Pública de Mozilla (MPL). En nuestro caso, recomendaremos utilizar la EUPL.

5

Proporcionar documentación

Incluir documentación que explique cómo utilizar, modificar y contribuir al proyecto de IA. Esto facilita que otros comprendan y colaboren en el desarrollo y mejora del sistema.

6

Establecer directrices de contribución

Definir claramente cómo otros pueden contribuir al proyecto, incluyendo pautas para reportar problemas, enviar mejoras y participar en discusiones.

7

Crear una comunidad activa

Fomentar la creación de una comunidad alrededor del proyecto para promover la colaboración, el intercambio de ideas y la mejora continua del sistema de IA.



Paso a paso para aplicar la licencia

Una vez que se analizó el sistema y se seleccionó la licencia más adecuada, es recomendable realizar los siguientes actos preparatorios encaminados a la distribución:

1. Revisar la lista de componentes y sus correspondientes licencias y obligaciones sobre la distribución.
 2. Prepararemos las cabeceras y los ficheros que acompañarán al software tradicionalmente, estos fueron:
 - COPYRIGHT, deberá indicar el año y el titular de los derechos
 - LICENSE, donde se incluirá el texto completo de la licencia
 - AUTHORS, donde se haga una mención al autor o autores del software
 - README, el fichero principal donde se coloca la información sobre el proyecto
 - NOTICE, que incluirá el listado de licencias de las dependencias de terceros utilizadas, y de ser el caso, otros avisos legales.
- » La nota de copyright se incluirá, como decimos, en cada uno de los ficheros del sistema o en sitio visible. Deberá indicar el año y el dueño del

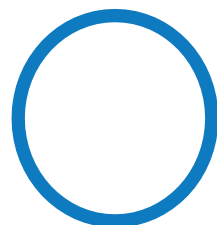
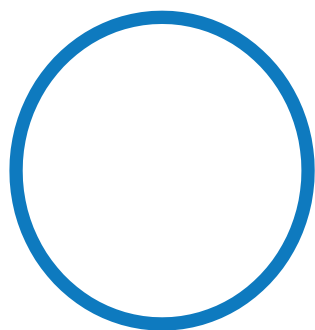
copyright. Algo con este formato resulta adecuado: Copyright (año) (nombre de la empresa).

» Se utiliza la palabra copyright por convención internacional, mismo para materiales que están en idiomas diferentes al inglés.

La autorización de copia se incluirá en cada uno de los ficheros fuentes y establece, con palabras sencillas, cuál es la licencia bajo a cuál se distribuye el programa y donde se puede leer el texto íntegro de la misma.

3. No modifiques las cabeceras de los ficheros de fuente abierta que no fueron modificados.
4. Se modificaste ficheros de fuentes abiertas, hace una anotación en la cabecera y especifica la naturaleza de la modificación, fecha y titular de los derechos de la modificación así como el contacto que pueda facilitar información a futuros colaboradores.

También es recomendable publicar en la página web del proyecto cuál es la licencia que escogemos y enlazar al texto completo de la licencia en cuestión.



AXENCIA PARA A
MODERNIZACIÓN
TECNOLÓXICA DE GALICIA