

# Guía liberación Intelixencia artificial



AXENCIA PARA A  
MODERNIZACIÓN  
TECNOLÓXICA DE GALICIA

2025. AMTEGA Guía liberación  
Intelixencia Artificial.

Esta obra distribúese cunha licenza Atribución Compartir  
Igual 4.0 Internacional de Creative Commons. Para ver  
unha copia da licenza pulse [aquí](#)

# Contenido

01

[Que é unha intelixencia artificial?](#)

02

[Intelixencia artificial xenerativa ao detalle](#)

03

[Intelixencia artificial Open Source](#)

04

[Como liberar unha Intelixencia artificial  
Open source?](#)

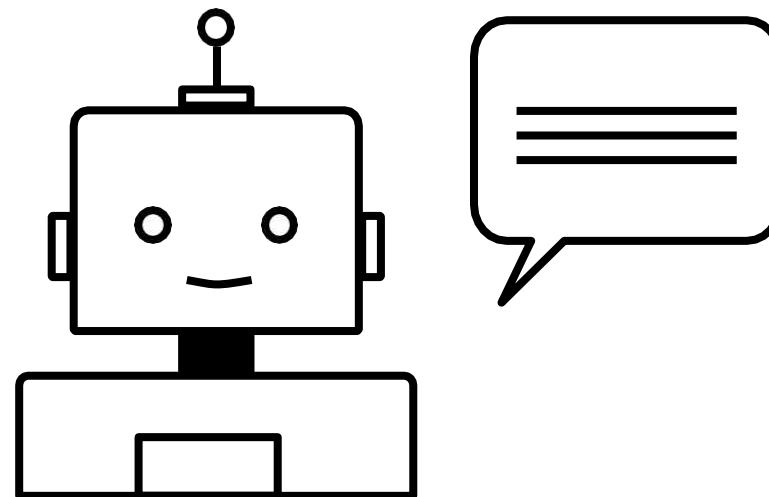
# Que é unha intelixencia artificial?

A Intelixencia Artificial é unha tecnoloxía que permite aos computadores e as máquinas simular a aprendizaxe humana, comprensión, solución de problemas, toma de decisións, creatividade e autonomía.

As aplicacións e dispositivos que usan Intelixencia artificial realizar todo tipo de tarefas como:

- Ver e identificar obxectos.
- Entender e responder a linguaxe humano
- Aprender da información e as experiencias novas.
- Facer recomendacións detalladas a usuarios e expertos.
- Poden identificar patróns e extraer información relevante.
- Poden actuar de forma autónoma, substituíndo a necesidade de intelixencia ou intervención humana (como nos coches autónomos).

**Leva sendo investigada desde 1950!**



# Conceptos e tipos de intelixencia artificial.

## Intelixencia Artificial (IA) – O campo xeral

Área da informática que busca crear sistemas capaces de realizar tarefas que normalmente requiren intelixencia humana, como o recoñecemento de voz, a toma de decisións e o procesamento da linguaxe natural.

## Deep Learning (Aprendizaxe Profunda) – Un tipo especializado de Aprendizaxe Automática

Dentro da aprendizaxe automática, o Deep Learning refírese a técnicas que usan redes neuronais profundas, con moitas capas de nodos (similares a neuronas), para modelar e resolver tarefas complexas.

## Aprendizaxe Automática (Machine Learning) – Un subconjunto da IA

Dentro da Intelixencia Artificial, a aprendizaxe automática céntrase en desenvolver algoritmos que permiten ás máquinas aprender dos datos e facer predicións ou tomar decisións sen ser programadas explicitamente para cada tarefa.

## Intelixencia Artificial Xenerativa – Un tipo avanzado de aprendizaxe automática

A IA xenerativa é un tipo de aprendizaxe automática que non só analiza datos, senón que tamén xera contido novo (texto, imaxes, música etc.)/ etc.). Baséase en modelos como redes neuronais profundas e técnicas como os modelos xenerativos adversarios (GANs) ou os transformadores (como GPT).

Década  
1950

# Intelixencia artificial

Simulación ou exhibición de intelixencia polas máquinas

Década  
1980

## Machine learning

Sistemas de Intelixencia artificial que aprenden de datos históricos

Década  
2010

## Deep learning

Os modelos de Machine Learning copian o funcionamento do cerebro con "neuronas", por iso chámanlles "sistemas neuronais"

Década  
2020

## Intelixencia artificial xenerativa

Modelos de Deep Learning (modelos fundacionais) que poden xerar contidos

# Intelixencia artificial xenerativa ao detalle

Aínda que a Intelixencia Artificial leva connosco desde 1950, desde 2022 tomaron especial relevancia entre os investigadores e o público xeral as coñecidas como “Intelixencia Artificial Xenerativa”

As Intelixencias Artificiais Xenerativas caracterízanse por ser unha tecnoloxía que xera textos, imaxes, vídeos e outros contidos.

Estes modelos aprenden os patróns e estruturas dos datos cos que foron adestrados e úsanos para xerar novos datos baseados na entrada que reciben, xeralmente en forma de instrucións en linguaxe natural.

Para comprender plenamente a Intelixencia Artificial xenerativa, primeiro debemos entender que a súa tecnoloxía baséase na aprendizaxe automática (ML) e a aprendizaxe profunda, só que vai un paso máis aló.

## Concepto clave

As Intelixencias Artificiais Xenerativas requiren grandes conxuntos de datos para ser adestradas e cheguen a certos resultados.

# Datasets para o adestramento de IA.

**Un dataset ou conxunto de datos é unha colección organizada de información que se utiliza para realizar análise, adestrar modelos de intelixencia artificial, facer investigacións, entre outras aplicacións.**

A información utilizada é moi variada aínda que se englobe como "datos" ou información", xa que, en realidade, trátase dun conxunto de números, textos, imaxes ou sons, algúns datos poden ser especialmente sensibles, por ser información persoal ou poden estar protexidos por propiedade intelectual.

Para poder adestrar ás Intelixencias artificiais dátaos sets estrutúranse, xeralmente, en forma de táboas ou matrices.

Cada fila dun dataset representa unha instancia ou exemplo dos datos (por exemplo, unha persoa, unha transacción, unha canción), e cada columna representa unha característica ou atributo (por exemplo, nome, idade, salario, notas ou os píxeles dunha imaxe). Dependendo do tipo de problema, os datasets poden ser etiquetados (con información de saída coñecida) ou non etiquetados (sen información de saída).

## **Por que son importantes?**

Dátaos sets é a clave detrás da precisión dunha intelixencia artificial e necesitamos saber que algúns datos espertan preocupacións, como os datos persoais (nome, salario, dirección) ou por estar protexidos por propiedade intelectual.

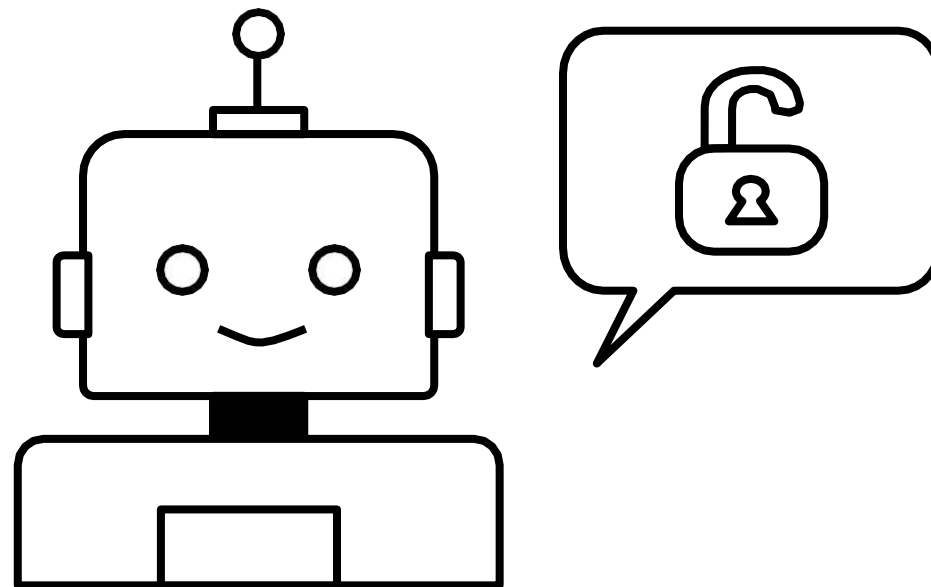
# Intelixencia artificial Open source

O desenvolvemento de Programas de Computador conta con código que pode ser usado, estudado, distribuído, modificado e mellorado de maneira libre, o que chamamos Software libre ou Open source.

Recentemente a Open Source Initiative (OSI) lanzou a primeira definición de Intelixencia artificial Open source.

A modo de resumo podemos dicir que unha Intelixencia Artificial Open source é aquela cuxo código, datos e modelos están dispoñibles para que calquera os use, estude, modifique e comparta libremente.

Isto permite que a comunidade contribúa ao seu desenvolvemento e mellora, fomentando a transparencia e a innovación.



# Beneficios da Intelixencia artificial Open source

## Transparencia e seguridade

As Intelixencias artificiais Open source permite obter información esencial para auditar o sistema ou mitigar os prexuízos heredados dos datos, permite obter responsabilidade sobre os modelos e a transparencia sobre a fonte de datos, o que, ao final, tradúcese nunha mellora na investigación sobre seguridade da IA

## Competencia e diversidade

A IA Open source , estimula a innovación e a calidade xa que, cuantos máis modelos haxa e maior sexa a competencia, maior será diversidade de solucións haberá. Isto é beneficioso para evitar os monopolios e os prezos elevados das solucións, proporcionando a máis partes interesadas acceso á tecnoloxía fundacional.

## Aplicacións variadas

A IA Open source ofrece aos desenvolvedores acceso a recursos cruciais para desenvolver aplicacións específicas para cada contexto, ademais, permite representar unha verdadeira diversidade cultural e lingüística, evitando actos discriminatorios no modelo aliñado con distintos sistemas de valores.

# Serven as licenzas tradicionais de Software libre?



A resposta require pensar en como funcionan as Intelixencias artificiais, xa que ao ter distintos compoñentes, como os data sets, as licenzas tradicionais de Open source poida que non sexan suficientes para garantir o dereito para usar, estudar, compartir ou modificar os sistemas de intelixencia artificial

Pensa que, sen os datos de adestramento, non poderíamos auditar verdadeiramente á intelixencia artificial.

Se se libera a Intelixencia artificial, debería incluír os datos para que sexa verdadeiramente libre.

# Requisitos para Intelixencia artificial Open source

Segundo a Open Source Initiative (OSI), para garantir as liberdades do Open source nun sistema de Intelixencia Artificial débese ter licenciar forma preferida para realizar modificacións no sistema, o que implica ter acceso a:

## **Código fonte accesible:**

A licenza libre ou Open source afecta o código fonte utilizado para adestrar e executar o sistema ao completo.

O código incluírá a especificación completa de como se procesaron e filtraron os datos e como se realizou o adestramento.

## **Parámetros**

Os parámetros e pesos do modelo deben poder descargarse e utilizarse libremente.

## **Información dos Datos de adestramento**

A información sobre os datos utilizados no adestramento debe ser clara e accesible. De modo que unha persoa capacitada poida construír un sistema substancialmente equivalente.

A información sobre os datos facilitarase incluíndo:

1. A descrición completa de todos os datos utilizados para o adestramento, incluídos os datos non compartibles,
2. A procedencia dos datos, o seu alcance e características, como se obtiveron e seleccionaron os datos, l
3. Os procesos de selección, filtrado e etiquetaxe deben ser transparente;
4. Unha listaxe de todos os datos de adestramento dispoñibles publicamente e onde obtelos; e
5. Unha listaxe de todos os datos de adestramento obtenibles de terceiros e onde obtelos, mesmo previo pago.

Como comentamos anteriormente, a Intelixencia Artificial ten o reto especial do conxunto de datos de adestramento, xa que estes son fundamentais para o seu funcionamento.

# Como liberar unha Intelixencia artificial Open source?

Xa habemos visto que liberar unha intelixencia artificial (IA) como Open source implica garantir que tanto o sistema completo como os seus compoñentes individuais estean dispoñibles baixo termos que outorguen as catro liberdades fundamentais, é dicir, que calquera persoa poida usar, estudar modificar e distribuír o sistema.

Para liberar á intelixencia artificial, realizaremos uns pasos extra aos xa comentados noutras Guías da Xunta de Galicia.

Resumiremos os pasos e daremos por feito que aplicarás a mesma licenza e as súas condicións.

# 1 Documentar a información do conxunto de datos

Proporcionar unha descrición completa de todos os datos utilizados para adestrar a IA, incluíndo a súa procedencia, alcance, características, métodos de obtención e selección, procedementos de etiquetaxe e metodoloxías de procesamento e filtrado. Tamén se debe incluír unha lista de todos os datos de

adestramento dispoñibles publicamente e onde obtelos, así como dos datos obtenibles de terceiros e onde acceder a eles, mesmo se requiren pago.

# 2 Liberar o código fonte

Facer dispoñible o código fonte utilizado para procesar e filtrar os datos, adestrar a IA (incluíndo argumentos e configuracións utilizadas), validación e probas, bibliotecas de soporte como tokenizadores e código de procura de hiperparámetros, código de inferencia e a arquitectura do modelo. Este código debe estar baixo licencias Open source.

# 3 Compartir os parámetros do modelo

Proporcionar os parámetros do modelo, como os pesos ou outras configuracións, incluíndo puntos de control de etapas intermedias clave do adestramento, así como o estado final do optimizador. Estes parámetros deben estar dispoñibles baixo licencias Open source

# 4 Seleccionar unha licenza adecuada

Elixir unha licenza libre ou Open source que garanta as catro liberdades e sexan compatibles co teu proxecto. Algunhas licenzas comúns inclúen a Licenza Pública Xeral de GNU (GPL), a Licenza MIT e a Licenza Pública de Mozilla (MPL). No noso caso, recomendaremos utilizar a EUPL.

# 5

## Proporcionar documentación

Incluír documentación que explique como utilizar, modificar e contribuír ao proxecto de IA. Isto facilita que outros comprendan e colaboren no desenvolvemento e mellora do sistema.

# 6

## Establecer directrices de contribución

Definir claramente como outros poden contribuír ao proxecto, incluíndo pautas para reportar problemas, enviar melloras e participar en discusións.

# 7

## Crear unha comunidade activa

Fomentar a creación dunha comunidade ao redor do proxecto para promover a colaboración, o intercambio de ideas e a mellora continua do sistema de IA.



# Paso a paso para aplicar a licenza

Unha vez que se analizou o software e seleccionouse a licenza máis adecuada, é recomendable realizar os seguintes actos preparatorios encamiñados á distribución:

1. Revisar a lista de compoñentes e as súas correspondentes licenzas e obrigacións sobre a distribución
  2. Prepararemos as cabeceiras e os ficheiros que acompañarán ao software tradicionalmente, estes foron:
    - COPYRIGHT, deberá indicar o ano e o dono do copyright
    - LICENSE, onde se incluírá o texto completo da licenza
    - AUTHORS, onde se faga unha mención ao autor ou autores do software
    - README, o ficheiro principal onde se coloca a información sobre o proxecto
    - NOTICE, que incluírá a listaxe de licenzas das dependencias de terceiros utilizadas, e de ser o caso, outros avisos legais.
- » A nota de copyright incluírase, como dicimos, en cada un dos ficheiros fontes. Deberá indicar o ano e o dono do copyright. Algo con este formato

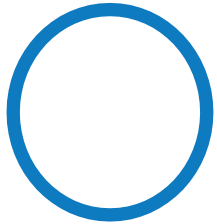
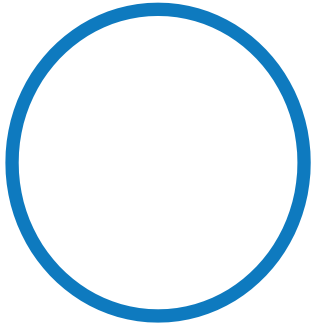
resulta adecuado: Copyright (ano) (nome da empresa).

» Utilízase a palabra copyright por convención internacional, mesmo para materiais que están en idiomas diferentes ao inglés

A autorización de copia incluírase en cada un dos ficheiros fontes e establece, con palabras sinxelas, cal é a licenza baixo a cal se distribúe o programa e onde se pode ler o texto íntegro da mesma.

3. Non modifiques as cabeceiras dos ficheiros de fonte aberta que non foron modificados.
4. Se modificaches ficheiros de fontes abertas, fai unha anotación na cabeceira e especifica a natureza da modificación, data e titular dos dereitos da modificación así como o contacto que poida facilitar información a futuros colaboradores.

Tamén é recomendable publicar na páxina web do proxecto cal é a licenza que escollemos e enlazar ao texto completo da licenza en cuestión.



AXENCIA PARA A  
MODERNIZACIÓN  
TECNOLÓXICA DE GALICIA