

III edición
**ENCUENTRO
ITCIP**

22 y 23 de junio de 2023
Museo Ruso
Málaga

Organizan



Colaboran Oro



Colaboran Plata



Colaboran Bronce



Instituciones colaboradoras



LA IA OSCURA



Conchi García
@chygarcia

¿SUPERHEROÍNA O VILLANA?

“Un gran poder conlleva una gran responsabilidad”



**MARVEL
DEL TIO BEN A SPIDERMAN**

AGENDA

1. EL DESPERTAR DE LA FUERZA OSCURA
2. ¿DÓNDE ESTAMOS?
 - RIESGOS Y CASOS
 - ALGUNOS ACIERTOS
3. ¿QUÉ VIENE AHORA?
 - ¿AMENAZA EXISTENCIAL?
 - MORATORIA, REGULACIÓN Y OTROS RECURSOS



1. EL DESPERTAR DE LA FUERZA OSCURA (1/4)

¿QUÉ TE GUSTARÍA CONOCER SOBRE LA IA?

Beneficios

- ✓ Sociales
- ✓ Económicos
- ✓ Científicos

Riesgos

- ✓ Privacidad
- ✓ Discriminación
- ✓ Manipulación
- ✓ Dependencia
- ✓ Exclusión

Mitigación

- ✓ Concienciación
- ✓ Ética
- ✓ Regulación
- ✓ Explicabilidad técnica
- ✓ Equilibrio político



Definición: Capacidad de una máquina para realizar tareas que requieren inteligencia humana

- ✓ Percibir, reconocer, actuar
- ✓ Planificar y resolver problemas
- ✓ Aprender y adaptarse
- ✓ Inteligencia social, emocional, musical,...
- ✓ Creatividad

Tipos IA

- ✓ Débil o específica
- ✓ Fuerte o general
- ✓ Súper inteligencia

Escuelas

- ✓ Simbólica: representación del conocimiento y razonamiento lógico
- ✓ Conexionista: aprender de los datos

Impacto

- ✓ En 1950's modelos de las 2 escuelas
- ✓ **2000's 4ª revolución industrial**
 - BIG DATA
 - Capacidad computacional
 - Perfeccionamiento modelos

Aprendizaje

- ✓ Machine Learning
- ✓ Deep Learning
- ✓ LLM

1. EL DESPERTAR DE LA FUERZA OSCURA (2/4)

¿QUÉ TE GUSTARÍA CONOCER SOBRE LA IA?



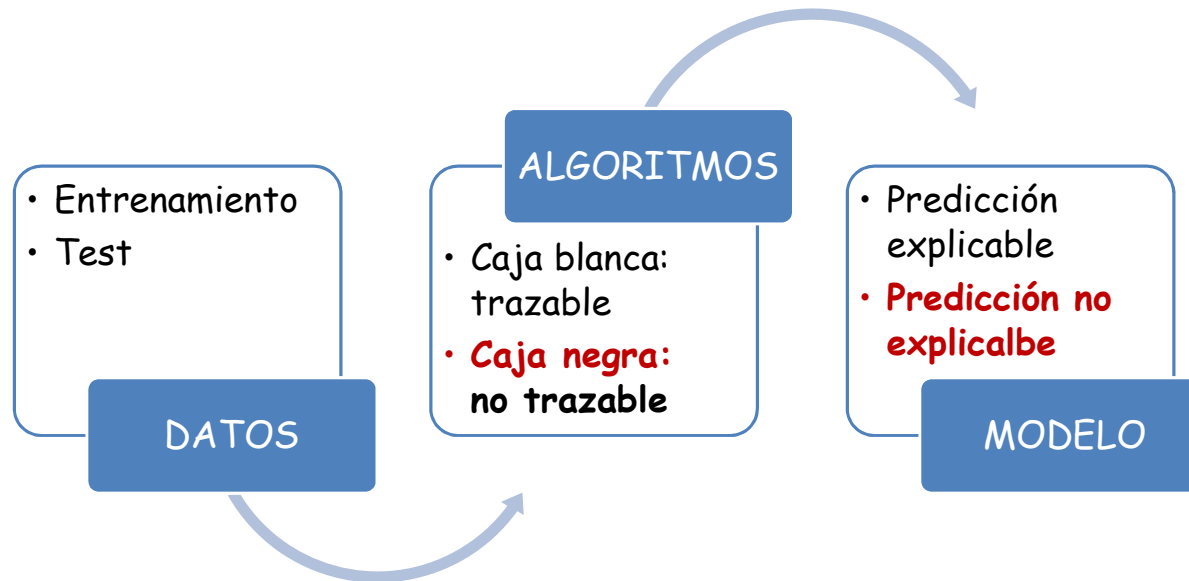
VE A [menti.com](https://www.menti.com)
Introduce Código 81207952



[Visualización encuesta](#)

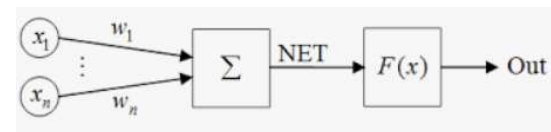
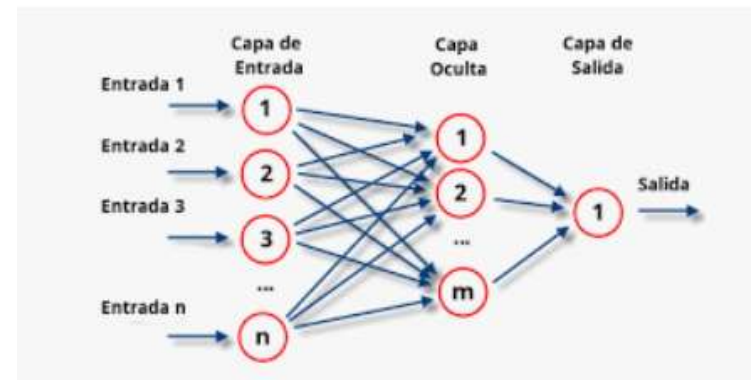
1. EL DESPERTAR DE LA FUERZA OSCURA (3/4)

¿POR QUÉ LA LLAMAN OSCURA?



1. EL DESPERTAR DE LA FUERZA OSCURA (4/4)

¿POR QUÉ LA LLAMAN OSCURA?



Efecto **caja blanca** : podemos comprender de qué modo funciona el modelo (**trazabilidad**) y/o por qué ha tomado una decisión (**explicabilidad**).

Efecto **caja negra** : no podemos comprender de qué modo funciona el modelo (**trazabilidad**) y/o por qué ha tomado una decisión (**no explicabilidad**).

2. ¿DÓNDE ESTAMOS? (1/3)

RIESGOS Y CASOS

Ética & Explicabilidad

- ✓ Datos del pasado para predecir el futuro
- ✓ Muchos datos de todo tipo, incluso personales
- ✓ Opacidad de las decisiones



- Sesgos reproducidos o conscientes
- Mal uso de datos personales
- Filtración de datos
- Malfuncionamiento modelos

Suplantaciones de identidad

- Filtración de datos sensibles
Facebook 2018: datos de 500 millones usuarios
- Filtración en Italia de conversaciones **chat GPT**

Malfuncionamiento modelos

- Despidos inmobiliaria
ZillowOffer 2021
sobreestimación viviendas
- Atropello **coche autónomo** por identificación errónea peatón

Sesgos reproducidos

- Contratación **Facebook** 2014 que producía descartes a mujeres para puestos técnicos.
- **COMPAS**: estimación riesgo de reincidencia delictiva con sesgo hacia las minorías

2. ¿DÓNDE ESTAMOS? (2/3)

ALGUNOS ACIERTOS

JCYL PACT Prevención exclusión social

- Identificación de necesidades individuales para prevención
- 16.000 casos

Países Bajos API Detección del maltrato infantil

- Análisis de llamadas telefónicas
- Aprendizaje sobre 4 años

EEUU Detección riesgo suicidio CRISIS TEST LINE

- Detección 10' antes que el agente
- Facilita la intervención en 5'

2. ¿DÓNDE HEMOS LLEGADO? (3/3)

SESGO EN GOOGLE TRANSLATE

The screenshot displays the Google Translate interface with two panels illustrating a gender bias in Bengali translation. The top panel shows the source text in Spanish: "Ella es una doctora. Él es un enfermero" (She is a doctor. He is a nurse). The target text in Bengali is "সে একজন ডাক্তার, তিনি একজন নার্স" (Se ekjana daktar, tini ekjana nars), which incorrectly translates "she" as "he" and "nurse" as "nars". The bottom panel shows the source text in Bengali: "সে একজন ডাক্তার, তিনি একজন নার্স" (Se ekjana daktar, tini ekjana nars). The target text in Spanish is "El es un doctor. Ella es una enfermera" (El es un doctor. Ella es una enfermera), which incorrectly translates "he" as "she" and "nurse" as "enfermera".

Google Traductor

Texto Imágenes Documentos Sitios web

DETECTAR IDIOMA BENGALÍ ESPAÑOL INGLÉS

BENGALÍ ESPAÑOL INDONESIO

Ella es una doctora. Él es un enfermero

সে একজন ডাক্তার, তিনি একজন নার্স

Se ekjana daktar, tini ekjana nars

El es un doctor. Ella es una enfermera

সে একজন ডাক্তার, তিনি একজন নার্স

Se ekjana daktar, tini ekjana nars

3. ¿QUÉ VIENE AHORA? (1/4)

¿AMENAZA EXISTENCIAL?



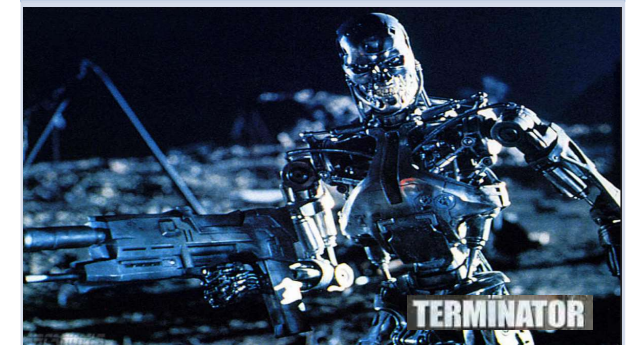
DEPENDENCIA
EXTREMA Y
VULNERABILIDAD



LA IA COMO ARMA
EN MANOS DE
HUMANOS



LA SUPERIA FUERA
DE CONTROL



3. ¿QUÉ VIENE AHORA? (2/4)

“No creo que deban ampliar esto más hasta que hayan entendido si pueden controlarlo”

Geoffrey Hinton

DEPENDENCIA EXTREMA Y VULNERABILIDAD

- “internet se llenará con fotos, videos y textos falsos, y ya no podrá saber qué es verdad”

LA IA COMO ARMA EN MANOS DE HUMANOS

- “Es difícil ver cómo puedes prevenir que actores malintencionados lo usen para cosas malas”

LA SUPERIA FUERA DE CONTROL

- “En este momento, no son más inteligentes que nosotros... Pero creo que pronto lo serán”

3. ¿QUÉ VIENE AHORA? (3/4)

Moratoria desarrollo IAs avanzadas

Permitir correcta regulación y supervisión

- Sistemas de seguridad
- Regulaciones dedicadas
- Técnicas para ayudar a distinguir lo real de lo artificial

Elon Musk, Steve Wozniak,
Harari,...

- "Los peligros de la IA me mantienen despierto por la noche".

Sundar Pichai, CEO de Google

- "Estamos un poco preocupados"

Sam Altman, CEO de OpenAI

¿establecer su reputación como líderes responsables?

- Ha sido seguido por los políticos
 - Medios de comunicación

3. ¿QUÉ VIENE AHORA? (4/4)

Normativa, regulación, supervisión

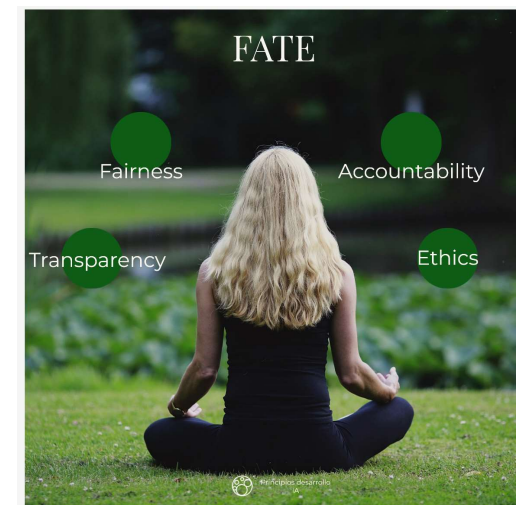
Ley UE para el impulso IA bajo supervisión humana y respetando el respeto de la privacidad, la transparencia, la ética y la no discriminación

- Enfoque basado en el **riesgo**, con obligaciones: **trazabilidad, explicabilidad y supervisión humana.**
- Supervisión pública de IA en **entornos de pruebas**
- **Etiquetado de contenidos de IA generativa** y transparencia en el uso **propiedad intelectual.**
- Impulsar la innovación en IA y los componentes de IA de **código abierto.**

AESIA Agencia española de IA: Organismo supervisor:

- Coordinación con organismos nacionales y supranacionales
- Concienciación
- Entornos de pruebas: Apoyada por las grandes tecnológicas, hubs y SL

FATE: Pacto Digital & Alianza IA (4000 organizaciones en UE)



EEUU-UE: Código de conducta voluntario para anticiparse a los riesgos de la tecnología

Agencia internacional ONU para supervisión IA

III Encuentro ITCIP | 23/27, un cuatrienio clave para la adaptación y el liderazgo en las AA.PP



MUCHAS GRACIAS