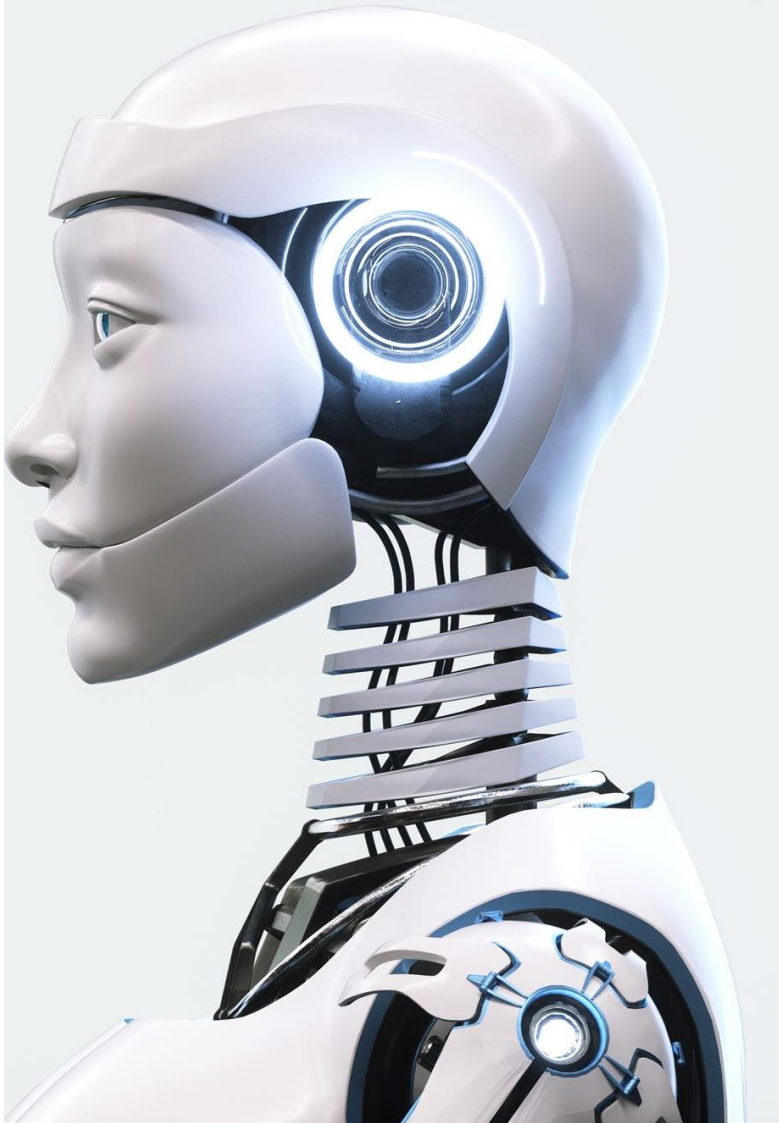

DESAFÍOS Y OPORTUNIDADES DE LA IA



ORDEN DEL DÍA

- Introducción a la IA
- Sesgos en la IA
- Desafíos Éticos de la IA
- Daños Involuntarios de la IA
- Necesidad de Intervención Humana
- Importancia de la Ética en la IA
- Pilares de la Ética de IA
- Ética en la Enseñanza



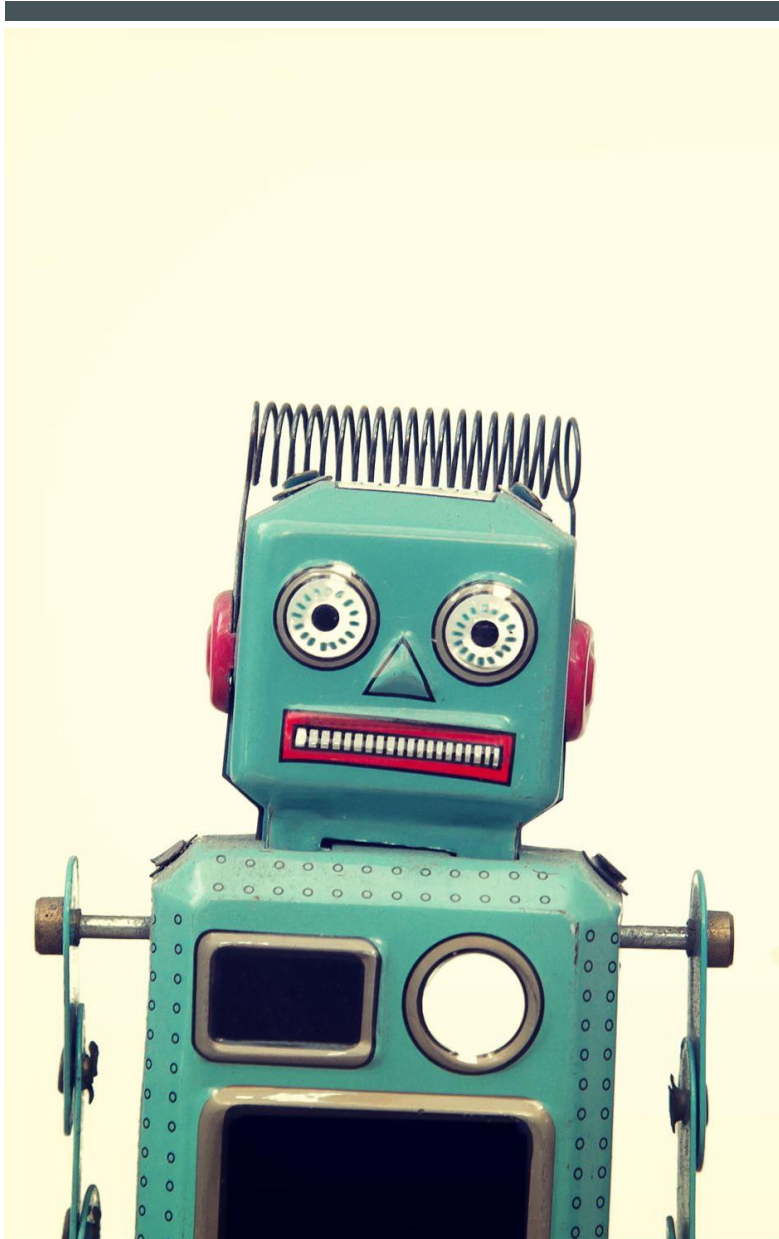
INTRODUCCIÓN A LA IA

- Capacidades Humanas
 - Creatividad
 - Lógica
 - Compasión
 - Razonamiento crítico
- Limitaciones de la IA
 - No puede hacer comentarios sobre empleados
 - No puede emitir juicios sobre candidatos a contratar
 - No puede proporcionar terapia a pacientes
 - Funciona como complemento de nuestras capacidades humanas

ORIGEN DE LOS SESGOS

- Sesgos Sistémicos
 - Tendencia sostenida por instituciones
 - Favorece o perjudica a determinados grupos
- Sesgos de Datos
 - Errores o prejuicios en la información
 - Genera datos sesgados

Para Limitar los Sesgos, la IA necesita la Intervención Humana ajustando



DESAFIOS CLAVES PARA UNA APROXIMACION DE ETICA IA

- Antropomorfismo
 - Asumir que la IA puede ser amable o cruel
- Alucinaciones
 - Errores en la generación de información
- Deuda Ética
 - Responsabilidad en el desarrollo y uso de la IA.
 - Sesgos
 - Ethical Washing

DAÑOS INVOLUNTARIOS SI SE USA LA IA DE MANERA IRRESPONSABLE

Distributiva: es cuando la IA niega oportunidades de información o recursos de bienestar a un sector, ejemplo evaluación crediticia

Calidad de Servicio. Herramientas de funcionan tan bien para grupos de personas, ejemplo reconocimiento de voz personas con afasia

Daño de representación refuerzo de la subordinación de grupos sociales basados en su identidad, Ejemplo traducción a de palabras asociadas a un genero

Daño de sistema social: efectos sociales que amplifican disparidades por clase, poder o privilegios "Deefakes".

IMPORTANCIA DE LA ÉTICA EN LA IA



- Imparcialidad y No Discriminación
 - Evitar sesgos
 - Acceso equitativo
- Transparencia y Explicabilidad
 - Decisiones comprensibles
- Responsabilidad
- Privacidad y Seguridad
 - Protección de datos
 - Seguridad de sistemas
- Impacto Socioeconómico
 - Desempleo y cambio laboral
 - Desigualdades económicas

5 PILARES DE LA ÉTICA IA

IMPARCIALIDAD

- Definición de Imparcialidad en IA
 - Tratamiento equitativo de individuos o grupos
 - Mitigación de sesgos no deseados
- Atributos Protegidos
 - Dividen la población en grupos
 - No hay un conjunto definido de atributos protegidos
- Grupos Privilegiados y No Privilegiados
 - Grupos privilegiados reciben resultados más favorables
 - Grupos no privilegiados reciben menos o ningún resultado favorable
- Importancia de la Imparcialidad
 - Evitar decisiones injustas
 - Mitigar sesgos sistemáticos



- Porque es importante la Imparcialidad
 - Equidad
 - Confianza
 - Ética
- Como se puede lograr la imparcialidad
 - Datos de alta calidad
 - Algoritmos transparentes
 - Evaluación Constante
 - Diversidad en los equipos
 - Regulaciones
- Ejemplos de sesgos de imparcialidad
 - Reconocimiento Facial
 - Contrataciones
 - Sistema de Justicia penal-Reincidencia





Solidez

- Importancia de la Solidez
 - Comportamiento correcto en condiciones excepcionales
 - Manejo eficaz de anomalías y ataques maliciosos
- Ataques de Adversario
 - Realizados intencionadamente para aprovechar vulnerabilidades
 - Tipos: envenenamiento y evasión
- **Ataques de envenenamiento:** El atacante corrompe el conjunto de datos de entrenamiento para que el modelo aprenda a cometer errores
- Un **ataque adversarial por evasión** es una técnica en la que se manipulan sutilmente los datos de entrada a un modelo de aprendizaje automático con el objetivo de que el modelo cometa un error específico en su clasificación



Explicabilidad

- Importancia de la Explicabilidad
 - Confianza en los resultados del modelo
 - Depuración de errores en el modelo
 - Justificación de decisiones en campos regulados
 - Garantía de ética y no discriminación
- Técnicas para Lograr la Explicabilidad
 - Modelos lineales para relaciones directas
 - Árboles de decisión para visualización de decisiones
 - Extracción de reglas simples
 - Técnicas de visualización con gráficos y diagramas
 - Métodos de atribución para puntuación de características

Transparencia

- Importancia de la Transparencia
 - Confianza: Aumenta la confianza en los resultados del modelo
 - Depuración: Ayuda a identificar y corregir errores
 - Regulaciones: Necesario para justificar decisiones en ciertos sectores
 - Ética: Garantiza que los modelos no discriminen
- Implicaciones de la Transparencia en IA
 - Interpretabilidad: Comprender la lógica interna del modelo
 - Explicabilidad: Explicar cómo se llegó a una decisión
 - Auditabilidad: Verificar y auditar el proceso de toma de decisiones
 - Justificación: Proporcionar una justificación clara de las decisiones
- Desafíos de la Transparencia
- Técnicas para Mejorar la Transparencia





Privacidad

- Privacidad en IA
 - Derecho del usuario sobre su información personal
 - Protección de datos en creación, validación y control del modelo
- Información Personal y Confidencial
 - Datos personales y confidenciales utilizados en el entrenamiento de modelos
 - Aplicación de técnicas de privacidad para proteger los datos
- Técnicas de Privacidad
 - Anonimización del modelo
 - Privacidad diferencial
- Ejemplo de Uso
- Soluciones para Proteger la Privacidad
- Pautas para Verificar la Seguridad de una IA

IMPLEMENTACION DE IA EN LA DOCENCIA

- Implementación de IA en la Labor Docente
 - Mejora de métodos de enseñanza
 - Optimización de recursos educativos
- Identificación de Sesgos y Equidad
 - Promoción de la igualdad en el aprendizaje
 - Reducción de prejuicios en la educación
- Evaluación Dirigida por IA
 - Fomento del pensamiento crítico
 - Evaluación de un espectro más amplio de habilidades
- Investigación y Neurodiversidad
 - Uso de IA para apoyar la diversidad estudiantil
 - Adaptación a diferentes estilos de aprendizaje



ACCIONES PARA INTEGRAR LA IA EN LA EDUCACIÓN

- A nivel Individual
 - Capacitación y espacios de experimentación de uso de IA
- Áreas Específicas
 - Respaldo a docentes en el uso de IA
 - Emplear expertos en IA para establecer normativas de seguridad y privacidad institucional
- Colaboración Transversal
 - Crear centros y comités interdisciplinarios sobre políticas de IA
- Colaboración Institucional
 - Establecer centros de investigación colaborativos



Gracias