



# INTELIGENCIA ARTIFICIAL

BENEFICIOS, RIESGOS Y RESPONSABILIDAD



Ing. Omar Vicente Lozano

# ORDEN DEL DÍA

- Introducción a la IA
- Asistentes de Voz
- IA Responsable
- Beneficios de la IA
- Riesgos de la IA
- Principios Responsables de IA
- Conclusiones Claves



# DEFINICIÓN DE IA

- Definición de Inteligencia Artificial
  - Ciencia que permite a las máquinas realizar análisis avanzados de datos
  - Ayuda a resolver problemas complejos
- Componentes de la IA
  - Creación de programas informáticos
  - Desarrollo de algoritmos
  - Análisis y aprendizaje a partir de datos



# HABILIDADES DE LA IA

- Procesamiento de Lenguaje Natural
  - Capacidad para entender y generar texto humano
- Reconocimiento de Patrones de Imágenes
  - Identificación y clasificación de objetos en imágenes
- Adaptación a Nuevas Situaciones
  - Aprendizaje y ajuste en tiempo real





# DEFINICIÓN Y PRÁCTICA DE IA RESPONSABLE

- Práctica de IA Responsable
  - Diseñar, desarrollar, implementar y retirar IA de forma segura y ética
- Impacto de las Aplicaciones de IA
  - En la sociedad
  - En los individuos
  - En el medio ambiente
- Abordar Riesgos Potenciales
  - Garantizar el uso responsable de las aplicaciones de IA



# DEBATES INTERDISCIPLINARIOS

- Participación de expertos de diversos campos
  - Informática
  - Filosofía
  - Derecho
  - Ciencias sociales
- Abordaje de desafíos complejos
  - Derivados del avance de la IA



# MEJOR ACCESO A LOS SERVICIOS

- Mejora de la accesibilidad
  - Brindar servicios a personas con acceso limitado
- Ejemplo de uso de IA
  - Diagnóstico de enfermedades respaldado por IA
  - Áreas con insuficientes profesionales de la salud



# AUTOMATIZACIÓN Y EFICIENCIA



- Automatización de Tareas Repetitivas
  - IA puede manejar tareas rutinarias
  - Aumenta la eficiencia y productividad
- Ejemplo de Uso
  - Chatbots de atención al cliente
  - Gestionan preguntas rutinarias



# ANÁLISIS DE DATOS Y PERSPECTIVAS

- Análisis de grandes cantidades de datos
  - Rápido y preciso
  - Extracción de información valiosa
- Comprensión del comportamiento de los clientes
  - Identificación de patrones
  - Innovación de nuevos productos
- Tendencias del mercado
  - Predicción de cambios
  - Adaptación a nuevas demandas



# SEGURIDAD Y PROTECCIÓN

- Detección de anomalías
  - La IA puede identificar patrones inusuales
  - Prevención de posibles amenazas
- Análisis de datos de sensores
  - Monitoreo de equipos industriales
  - Predicción de necesidades de mantenimiento



# INNOVACIÓN Y CREATIVIDAD

- Inspiración para Nuevas Ideas
  - La IA generativa puede inspirar nuevas ideas
  - Generación de obras de arte y composiciones musicales
- Innovación en Expresión Artística
  - Nuevas vías de expresión artística
  - Exploración de nuevas formas creativas





# DISPONIBILIDAD 24/7

- Funcionamiento Continuo
  - Los sistemas de IA operan sin interrupciones
  - Disponibilidad constante para los usuarios
- Asistentes Virtuales
  - Disponibles en cualquier momento
  - Brindan información de manera inmediata





# SESGOS Y DISCRIMINACIÓN

## ■ Sesgos

- Los sistemas de IA pueden heredar sesgos presentes en los datos de entrenamiento
- Sesgos de datos-Sesgos algoritmo-Sesgos de Interpretación- Sesgos de no Inclusion

## ■ Discriminación

La **discriminación** es el resultado del sesgo en la IA cuando este genera decisiones o acciones injustas contra ciertos grupos o individuos.

- Reclutamiento: Un sistema de IA que prefiere candidatos masculinos porque fue entrenado con datos históricos de contrataciones sesgadas.
- Justicia: Algoritmos de predicción del crimen que penalizan desproporcionadamente a ciertos grupos raciales.



# INVASIÓN DE LA PRIVACIDAD

- Recopilación de Datos Personales
  - Los sistemas de IA recopilan, usan y almacenan datos personales
  - Genera preocupaciones sobre la invasión de la privacidad
- Acceso No Autorizado
  - Posibilidad de acceso no autorizado a información confidencial
- Ejemplo de Dispositivo Inteligente
  - Dispositivo doméstico con IA y reconocimiento de voz
  - Graba y almacena conversaciones sin consentimiento explícito
  - Recopila información personal confidencial



# FALTA DE TRANSPARENCIA

- Falta de Transparencia en IA
  - Dificulta la comprensión de decisiones
  - Genera falta de responsabilidad
  - Posible desconfianza
- Ejemplo de Plataforma de Redes Sociales
  - Uso de algoritmos de IA para gestionar el feed
  - No se proporciona información sobre el funcionamiento de los algoritmos
  - Factores que influyen en la selección de contenido no son claros
  - Usuarios desconocen cómo se personalizan sus feeds
  - Genera falta de transparencia en las operaciones



---

# CONSECUENCIAS NO DESEADAS

- Complejidad de los Sistemas de IA
  - Puede producir consecuencias no deseadas e imprevistas
  - Potencial daño a personas o sociedad
- Escenario Hipotético: Moderación de Contenido
  - Sistema de IA en redes sociales para detectar y eliminar contenido inapropiado
  - Algoritmos de IA analizan el contenido
- Problemas del Sistema de IA
  - Marca y elimina por error publicaciones legítimas
  - Supresión de conversaciones significativas
  - Limitación de la libertad de expresión



# FALTA DE ACCESIBILIDAD

- Impacto de la falta de accesibilidad en IA
  - Soluciones de IA poco prácticas o inutilizables
  - Afecta especialmente a personas con discapacidades y deficiencias
- Ejemplo de un asistente de vídeo impulsado por IA
  - Limitaciones en la accesibilidad para personas con necesidades especiales



---

# DESINFORMACIÓN

- Desinformación y Desinformación
  - La IA generativa puede crear contenido inexacto o falso
  - Contribuye a la desinformación (inexactitudes no intencionadas)
  - Contribuye a la desinformación (contenido intencionalmente engañoso)
- Ejemplo de Desinformación
  - Investigación basada en un modelo de lenguaje amplio
  - Conjunto de datos del modelo contenía puntos obsoletos
  - Conclusiones inexactas sobre los hallazgos del estudio

---

# JUSTICIA

- Comprobación de Datos de Entrenamiento
  - Identificación de posibles sesgos
  - Garantía de representación equitativa en todos los grupos demográficos
- Seguimiento Periódico de Predicciones
  - Proceso de selección de candidatos equilibrado
- Compromiso con la Equidad
  - Promoción de la diversidad
  - Alineación con prácticas de contratación éticas
- Definición de Justicia en IA
  - Garantizar un trato justo y sin prejuicios ni discriminación

# FIABILIDAD Y ROBUSTEZ



- Importancia de la Robustez en IA
  - Garantiza el buen funcionamiento en situaciones difíciles
  - Permite manejar entradas inesperadas
- Escenario: Sistema de Detección de Fraude
  - Construcción de un sistema de detección de fraude financiero con IA
  - Diversificación de datos de entrenamiento
  - Pruebas del modelo para reducir errores
  - Implementación de diferentes modelos de IA
  - Adaptación a tácticas de fraude en evolución
- Aumento de la Robustez
  - Mejora la capacidad del sistema frente a amenazas emergentes





# SEGURIDAD Y PRIVACIDAD

- Importancia de la Seguridad y Privacidad
  - Proteger la privacidad de los datos
  - Garantizar la seguridad de los sistemas de IA
- Escenario: Asistentes Virtuales Seguros y Privados
  - Desarrollo de un asistente virtual para ayudar a las personas
- Medidas de Seguridad y Privacidad
  - Integración de cifrado de extremo a extremo para interacciones de voz confidenciales
  - Anonimización de datos de usuarios
  - Almacenamiento en servidores seguros con estrictos controles de acceso
- Beneficios
  - Mejora de la confidencialidad e integridad de los datos



# SOSTENIBILIDAD

- Impacto Ambiental de la IA
  - Evaluación del impacto ambiental y energía necesaria
  - Innovaciones para reducir el impacto ambiental
- Escenario: Sistema de Recomendación de Eficiencia Energética
  - Construcción de un motor de recomendaciones para comercio electrónico
- Optimización de Algoritmos
  - Minimización de demandas computacionales
  - Reducción del consumo de energía en un 30%
- Uso de Energía Renovable
  - Entrenamiento de modelos con fuentes de energía renovables
  - Reducción adicional del impacto ambiental
- Enfoque Sustentable



## IA CENTRADA EN EL SER HUMANO

- Consideración Ética de la IA
  - Evaluación del impacto en el empleo
  - Condiciones de trabajo humanas
  - Distribución equitativa de oportunidades
- Escenario de IA en Fabricación
  - Creación de soluciones de IA para empresas de fabricación
- Sistema de Producción Asistido por IA
  - Colaboración con trabajadores humanos
  - Optimización de tareas y reducción del esfuerzo físico
  - Sugerencias de capacitación en tiempo real
- Promoción de un Entorno de Trabajo Equilibrado
  - Combinación de tecnología y experiencia humana



# PRINCIPIOS DE IA RESPONSABLE



- Transparencia
  - Claridad en el funcionamiento y decisiones de la IA
- Responsabilidad
  - Asumir las consecuencias del uso de la IA
- Accesibilidad
  - Disponibilidad para todos los usuarios
- Confiabilidad
  - Funcionamiento seguro y predecible
- Equidad
  - Evitar sesgos y discriminación
- Respeto por los derechos humanos
  - Garantizar la protección de los derechos fundamentales



---

# PROMOCIÓN DE LA IA RESPONSABLE

- Creatividad y Participación
  - Necesidad de creatividad en el desarrollo de IA
  - Importancia de la participación de personas de diversos orígenes
- Desarrolladores y Organizaciones
  - Rol crucial de los desarrolladores en la creación de IA responsable
  - Contribución de las organizaciones para abordar desafíos
- Usuarios
  - Participación activa de los usuarios en el uso responsable de la IA
  - Colaboración para mitigar riesgos asociados a la IA



# Preguntas