

Taller 1: Introducción a Sistemas Inteligentes.

- Kevin Daniel Genez Valencia, C.C: 1128269097, kgenezv@eafit.edu.co

1. Utilice un modelo grande de lenguaje (LLM) — tal como ChatGPT, Bard o Bing— para obtener una descripción breve de los siguientes términos:

- **Big data:**

Definición corta: Volúmenes masivos de datos con alta variabilidad y velocidad de crecimiento, que exceden la capacidad de las herramientas de procesamiento tradicionales.

Explicación ampliada: El Big Data no se define solo por el tamaño de los datos, sino por las "tres V": Volumen, la cantidad de datos generados; Velocidad, la rapidez con la que se generan y procesan; y Variedad, los distintos formatos de datos (estructurados, semi-estructurados y no estructurados). En ciencia de datos, su análisis requiere de tecnologías y algoritmos avanzados para extraer valor, conocimiento y patrones de comportamiento.

Ejemplo práctico: El análisis de miles de millones de interacciones de usuarios en una red social (likes, comentarios, shares) en tiempo real para personalizar el feed de noticias y mostrar publicidad relevante.

- **Machine learning:**

Definición corta: Rama de la inteligencia artificial que permite a los sistemas aprender de los datos sin ser programados explícitamente.

Explicación ampliada: Se centra en el desarrollo de algoritmos y modelos que mejoran su desempeño en tareas específicas a medida que procesan más información. En ciencia de datos, es clave para construir predicciones y encontrar patrones ocultos en grandes volúmenes de datos. Se divide en categorías principales como aprendizaje supervisado (con datos etiquetados), no supervisado (con datos no etiquetados) y por refuerzo.

Ejemplo práctico: Un modelo de machine learning que predice si un cliente cancelará su suscripción a un servicio de streaming en base a su historial de uso (horas de visualización, tipo de contenido, etc.).

- **Inteligencia artificial:**

Definición corta: Campo de la informática que busca crear sistemas capaces de imitar la inteligencia humana para realizar tareas como el aprendizaje, la resolución de problemas y la percepción.

Explicación ampliada: Es la disciplina que se enfoca en el desarrollo de algoritmos y programas informáticos que exhiben comportamientos inteligentes. A diferencia del simple procesamiento de datos, la inteligencia artificial busca que las máquinas razonen, planifiquen y se adapten a nuevas situaciones. El Machine Learning y el Deep Learning son subcampos de la IA que se utilizan en ciencia de datos para construir modelos predictivos y de clasificación, entre otros.

Ejemplo práctico: Un asistente de voz como Siri o Google Assistant, que utiliza IA para entender el lenguaje natural, procesar comandos de voz y responder de manera coherente a las preguntas del usuario.

- **Ciencia de datos (data science):**

Definición corta: Campo interdisciplinario que utiliza métodos científicos, procesos, algoritmos y sistemas para extraer conocimiento y percepciones de datos estructurados y no estructurados.

Explicación ampliada: La ciencia de datos integra conocimientos de estadística, informática y áreas de negocio para analizar grandes volúmenes de datos. Su objetivo principal es transformar datos brutos en información útil y accionable para la toma de decisiones. Se diferencia de la simple analítica de datos en que se enfoca en la formulación de preguntas de negocio, la recolección de datos, la limpieza, el modelado y la comunicación de resultados.

Ejemplo práctico: Un equipo de ciencia de datos en una empresa de e-commerce analiza el historial de compras y navegación de los usuarios para crear un sistema de recomendación que sugiera productos relevantes, aumentando así las ventas.

- **Deep learning:**

Definición corta: Subcampo del Machine Learning que utiliza redes neuronales artificiales con múltiples capas para modelar abstracciones de alto nivel en los datos.

Explicación ampliada: A diferencia de los modelos de Machine Learning tradicionales, el Deep Learning se caracteriza por su arquitectura de red neuronal "profunda", con numerosas capas ocultas que permiten a los modelos aprender características complejas y representaciones jerárquicas de los datos de forma automática. Es especialmente eficaz en el procesamiento de datos no estructurados como imágenes, audio y texto, y ha impulsado avances significativos en el reconocimiento de voz y la visión por computadora.

Ejemplo práctico: El modelo de reconocimiento facial de un teléfono móvil, que utiliza una red neuronal profunda para identificar y verificar a un usuario a partir de una fotografía de su rostro.

- **Minería de datos (data mining):**

Definición corta: Proceso de descubrir patrones, tendencias y conocimientos valiosos en grandes conjuntos de datos utilizando métodos de Machine Learning y estadística.

Explicación ampliada: La minería de datos es un paso crucial en el ciclo de vida de la ciencia de datos. Su objetivo es explorar los datos desde diferentes perspectivas para encontrar información que no es visible a simple vista. Se diferencia de la analítica tradicional en que busca patrones y relaciones de manera proactiva en lugar de solo probar hipótesis predefinidas. Las técnicas de minería de datos incluyen la clasificación, la regresión, la agrupación (clustering) y la asociación.

Ejemplo práctico: Un supermercado analiza las transacciones de sus clientes y descubre que quienes compran pañales también tienden a comprar cerveza. Esta asociación de productos le permite colocar ambos artículos cerca uno del otro para aumentar las ventas.

- **Inteligencia de negocios:**

Definición corta: Conjunto de estrategias, tecnologías y aplicaciones para analizar datos de una empresa y presentarlos en informes y paneles, facilitando la toma de decisiones.

Explicación ampliada: La inteligencia de negocios (BI) se enfoca en la descripción de lo que ya ha sucedido en una organización. A diferencia de la ciencia de datos, que busca predecir el futuro, la BI utiliza analítica descriptiva para medir el rendimiento pasado y actual de un negocio. Su objetivo es convertir datos brutos en información fácilmente comprensible para que los líderes y equipos tomen decisiones más rápidas e informadas, respondiendo a preguntas como "¿cuántas ventas tuvimos el mes pasado?" o "¿qué producto es el más rentable?".

Ejemplo práctico: El departamento de ventas de una compañía de software utiliza un panel de BI para visualizar en tiempo real las ventas por región, el desempeño de cada vendedor y el estado de los ingresos, permitiendo identificar rápidamente áreas de éxito y oportunidades de mejora.

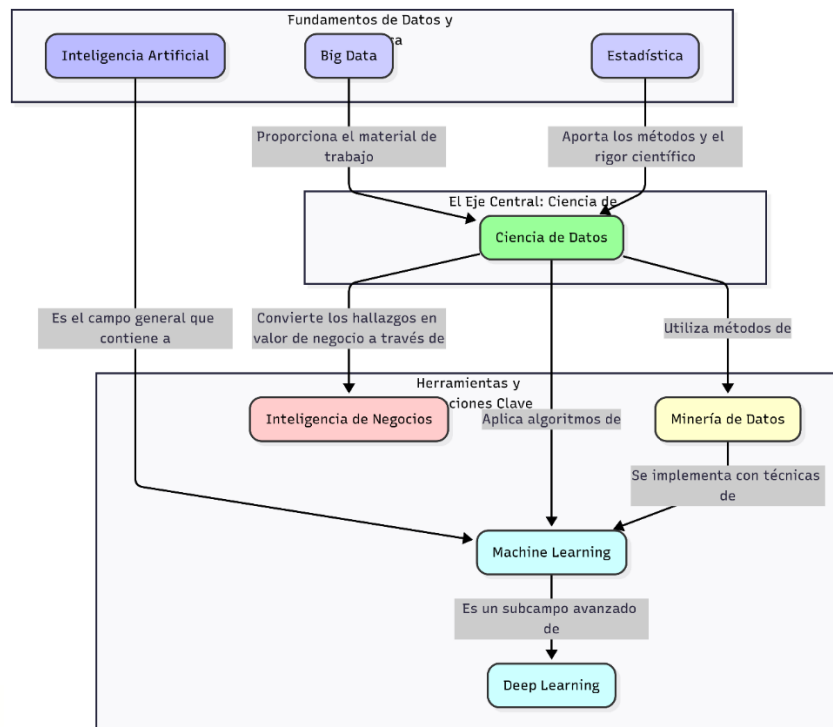
- **Estadística:**

Definición corta: La disciplina que se ocupa de la recopilación, organización, análisis, interpretación y presentación de datos para extraer conclusiones y tomar decisiones.

Explicación ampliada: La estadística es una de las bases matemáticas de la ciencia de datos. Proporciona las herramientas y el marco conceptual para entender y describir los datos (estadística descriptiva) y para hacer inferencias y predicciones sobre poblaciones más grandes a partir de muestras (estadística inferencial). Es fundamental para la fase de exploración de datos, la construcción de modelos, la evaluación de su rendimiento y la interpretación de los resultados de forma rigurosa.

Ejemplo práctico: Un científico de datos aplica estadística descriptiva para calcular la media y la desviación estándar de los ingresos de los clientes, y utiliza estadística inferencial para realizar una prueba de hipótesis que determine si una nueva estrategia de marketing ha generado un aumento "estadísticamente significativo" en las ventas.

2. Utilice un LLM para crear un diagrama que muestre las relaciones entre los anteriores conceptos. Pídale que genere el diagrama usando la sintaxis de mermaid.js y visualícelo usando <https://mermaid.live/>. Opcionalmente puede usar un plugin. Discuta el resultado, ¿tiene sentido? Pruebe dándole instrucciones al LLM para que lo mejore.



El gráfico que generamos presenta un mapa que explica la lógica de la disciplina.

La Ciencia de Datos como Eje Central: El diagrama posiciona la Ciencia de Datos (CD) como el corazón del proceso. Esto es crucial porque la CD no es una herramienta, sino una metodología. Su rol es el de un integrador: toma los recursos y los conocimientos de otras áreas para formular preguntas, diseñar experimentos, construir modelos y, finalmente, comunicar resultados. No existiría la ciencia de datos sin los cimientos de la Estadística, el material del Big Data y el marco de la Inteligencia Artificial.

Pilares Fundamentales del Conocimiento:

Big Data: Es el punto de partida. La flecha de Big Data hacia Ciencia de Datos es la más fundamental. No se puede hacer análisis sin datos masivos, que son los que ofrecen la oportunidad de encontrar patrones y valor.

Estadística: La flecha de Estadística a Ciencia de Datos es la del rigor. La estadística no solo provee los métodos para analizar los datos, sino que también ofrece la base para interpretar los resultados de los modelos. Es lo que nos permite saber si un hallazgo es significativo o simplemente una casualidad.

Inteligencia Artificial: La IA es el campo paraguas, la visión a largo plazo. De ella se desprende el Machine Learning, que es la parte de la IA que se enfoca en que los sistemas aprendan de los datos. Esta relación jerárquica es la razón por la que a menudo se confunden los términos, pero el diagrama lo aclara al mostrar que el ML es una pieza dentro de un concepto mucho más amplio.

El Proceso de Análisis y Modelado

Minería de Datos (Data Mining): La flecha que va de Ciencia de Datos a Minería de Datos ilustra que esta última es una fase, una técnica de exploración dentro del proceso. La minería de datos es la búsqueda proactiva de patrones, y se implementa usando algoritmos de Machine Learning.

Machine Learning (ML): El ML se nutre directamente de la IA y de la Minería de Datos. Es la herramienta de modelado por excelencia en la ciencia de datos. Su relación con el Deep Learning (DL) es de especialización; el DL es una forma avanzada de ML que usa redes neuronales profundas para problemas más complejos como el reconocimiento de imágenes o voz. El diagrama lo representa correctamente como un subcampo de ML.

La Salida y el Valor de Negocio

Inteligencia de Negocios (BI): La flecha que va de Ciencia de Datos a Inteligencia de Negocios es la que cierra el ciclo. Muestra que el propósito final del análisis no es solo generar modelos complejos, sino traducir los hallazgos en información digerible y útil para la toma de decisiones. Mientras la CD se enfoca en el "porqué" y el "qué pasará", la BI se enfoca en el "qué pasó" y el "dónde estamos", utilizando los resultados de la CD para enriquecer sus paneles y reportes.

En conjunto, el diagrama narra una historia. Comienza con los cimientos (IA, Big Data, Estadística), pasa por el proceso de análisis y modelado (Ciencia de Datos, Minería de Datos, Machine Learning) y culmina con la aplicación práctica y el impacto en el negocio (Inteligencia de Negocios). Es un modelo coherente y preciso del ecosistema de la ciencia de datos.

3. Busque un video reciente sobre alguna aplicación o técnica de inteligencia artificial que usted considere muy interesante. Véalo y haga una síntesis. Escoja un segmento de unos 30 segundos para mostrar al resto de la clase.

- **Video:** NeuralNine, Fine-Tuning Local LLMs with Unsloth & Ollama (https://www.youtube.com/watch?v=W_xh6qNSfAQ&t=17s)
- **Síntesis:** El video es un tutorial que explica cómo hacer *fine-tuning* (ajuste fino) de modelos de lenguaje grandes (LLMs) de manera local utilizando la biblioteca de Python **Unsloth** [00:06].

Puntos Clave:

- **Unsloth:** Esta herramienta acelera significativamente el proceso de ajuste, reduciendo el consumo de VRAM y el tiempo de entrenamiento mediante técnicas como 4-bit QLoRA [00:18].
- **Requisitos:** Para realizar el proceso de forma local, se necesita una GPU con soporte para CUDA. El video menciona una NVIDIA GeForce RTX 3060Ti, pero también sugiere el uso de un cuaderno de Google Colab con una GPU T4 para quienes no tengan el hardware [03:03].
- **Flujo del Proceso:** El tutorial detalla los pasos, que incluyen seleccionar un modelo base, preparar un conjunto de datos personalizado (en formato JSON), escribir el código para cargar y ajustar el modelo, y finalmente, exportar el modelo ajustado a un formato (GGUF) compatible con **Ollama**, una herramienta para ejecutar modelos localmente [00:45].

En resumen, el video muestra un método práctico para enseñar a un LLM a realizar una tarea específica, como extraer y formatear información en un JSON, destacando la facilidad y eficiencia que ofrece la biblioteca Unsloth para este tipo de proyectos.

4. Use un LLM para determinar qué son los transformers. Busque información al respecto sobre los siguientes aspectos:

a. Cuáles son las ideas principales de su funcionamiento:

Las ideas principales del funcionamiento de los Transformers se centran en un mecanismo revolucionario que les permite procesar secuencias de datos de manera simultánea, a diferencia de modelos anteriores que lo hacían de forma secuencial.

Mecanismo de Atención (Self-Attention): Este es el concepto más crucial y el corazón de la arquitectura Transformer. El mecanismo de auto-atención permite que el modelo pese la importancia de cada palabra en una secuencia de entrada al procesar una sola palabra. En lugar de procesar "él" y "ella" basándose solo en las palabras adyacentes, el modelo puede determinar la relevancia de cada una en relación con todas las demás palabras de la frase. Esto le permite capturar dependencias a largo plazo y entender el contexto completo de una oración, sin importar la distancia entre las palabras.

Codificadores y Decodificadores: La arquitectura tradicional de un Transformer está compuesta por dos partes principales:

- El codificador procesa la secuencia de entrada (por ejemplo, una frase en un idioma). Su función es generar una representación contextual de cada palabra, enriquecida por la información de la atención.
- El decodificador utiliza esta representación del codificador para generar la secuencia de salida (por ejemplo, la traducción a otro idioma), construyéndola palabra por palabra.

Procesamiento No Secuencial y Codificación Posicional: A diferencia de las redes neuronales recurrentes (RNN) que procesan las palabras una tras otra, los Transformers procesan todas las palabras de la secuencia de forma paralela. Esto los hace mucho más rápidos y eficientes, especialmente con secuencias largas. Sin embargo, al no procesar en orden, el modelo perdería la información de la posición de cada palabra. Para solucionar esto, se utiliza la codificación posicional, que es una técnica para inyectar información sobre la posición de cada palabra en la secuencia, garantizando que el modelo entienda el orden y la estructura sintáctica de la oración.

En esencia, el Transformer toma una frase, le añade la información de la posición de cada palabra y la procesa en paralelo. Usando la atención, se asegura de que cada palabra se relacione con todas las demás para entender su contexto completo, logrando así un nivel superior de comprensión.

b. Cuáles son las aplicaciones.

El poder de los Transformers radica en su capacidad para manejar secuencias de datos de manera eficiente, lo que ha revolucionado múltiples campos más allá del procesamiento de lenguaje natural (NLP). Sus aplicaciones más destacadas incluyen:

1. Procesamiento de Lenguaje Natural (NLP): Este fue el dominio original para el que se diseñaron los Transformers y donde han tenido el mayor impacto.

- Traducción automática: Modelos como los utilizados en Google Translate o DeepL usan Transformers para traducir texto de un idioma a otro de forma fluida y contextualmente precisa.
- Generación de texto: Son la base de los grandes modelos de lenguaje (LLM) como GPT-4, que pueden escribir artículos, poemas, código y responder preguntas de manera coherente.
- Resumen de texto: Pueden analizar documentos largos y generar resúmenes concisos, ya sea extrayendo frases clave o creando un nuevo texto que sintetice las ideas principales.
- Análisis de sentimiento: Se utilizan para clasificar el tono de un texto (positivo, negativo, neutro), lo que es crucial para analizar reseñas de clientes o comentarios en redes sociales.

2. Visión por Computadora: Aunque originalmente eran para texto, los Transformers han demostrado ser muy eficaces para tareas visuales. Tratan las imágenes como secuencias de "parches" o partes, analizando sus relaciones de la misma forma que lo hacen con las palabras.

- Clasificación de imágenes: Pueden identificar el contenido de una imagen, como reconocer un perro o un gato.

- Generación de imágenes: Son el motor de modelos como DALL-E y Midjourney, que crean imágenes realistas o artísticas a partir de descripciones de texto.
3. Procesamiento de Audio: Su arquitectura secuencial también se aplica al análisis de ondas sonoras.
- Reconocimiento de voz: Transcriben el habla a texto con alta precisión, como los sistemas utilizados en asistentes virtuales o servicios de transcripción.
4. Otras Aplicaciones Científicas: La capacidad de los Transformers para encontrar relaciones en secuencias de datos se ha extendido a la ciencia.
- Biología: En la bioinformática, se utilizan para predecir la estructura de proteínas (como en el famoso modelo AlphaFold de Google DeepMind) o para analizar secuencias de ADN.
 - Descubrimiento de fármacos: Ayudan a predecir las propiedades de las moléculas y a acelerar la búsqueda de nuevos compuestos.

c. Qué son los grandes modelos de lenguaje y qué relación tienen con los transformers.

¿Qué son los Grandes Modelos de Lenguaje (LLM)?

Los Grandes Modelos de Lenguaje (LLM) son un tipo de inteligencia artificial diseñada para comprender, generar y manipular el lenguaje humano. Se llaman "grandes" porque están entrenados con cantidades masivas de datos de texto y código (billones de palabras) y tienen una enorme cantidad de parámetros (cientos de miles de millones). Esta escala les permite aprender patrones complejos y realizar una amplia gama de tareas lingüísticas, desde responder preguntas hasta escribir artículos, resumir documentos o traducir idiomas con una fluidez notable.

La Relación con los Transformers

La relación es directa y fundamental: la arquitectura Transformer es el esqueleto tecnológico de casi todos los LLM modernos.

El Transformer no es un LLM en sí mismo, sino el mecanismo que hizo posible su existencia. Antes de los Transformers, los modelos de lenguaje usaban arquitecturas que procesaban el texto de forma secuencial (palabra por palabra), lo que era ineficiente y limitaba la capacidad del modelo para capturar relaciones a largo plazo en textos extensos.

La innovación clave del Transformer, el mecanismo de auto-atención, resolvió este problema. Permite que el modelo analice todas las palabras de una frase simultáneamente, lo que facilita el entrenamiento con enormes volúmenes de datos y le da al LLM la capacidad de:

- Comprender el contexto global: Relacionar una palabra con otras que se encuentran a gran distancia en el texto.
- Aprender a escala: Su arquitectura paralela permite entrenar modelos con miles de millones de parámetros en grandes supercomputadoras, algo imposible con arquitecturas anteriores.

En pocas palabras, los Transformers son la arquitectura subyacente que le da a los LLMs su poder, eficiencia y capacidad para manejar la complejidad del lenguaje. Son la tecnología que convirtió los modelos de lenguaje de una curiosidad académica en la fuerza dominante de la IA generativa que vemos hoy.

d. Cómo se entrena un LLM conversacional.

El entrenamiento de un LLM conversacional es un proceso en etapas que va desde un modelo genérico hasta una IA capaz de interactuar de forma útil y segura. Este proceso se divide en tres fases principales.

1. Pre-entrenamiento (Base Training)

Esta es la etapa inicial y más intensiva. El modelo es entrenado con una cantidad masiva de datos no estructurados de internet, incluyendo libros, artículos, código y sitios web.

Objetivo: El modelo aprende la gramática, la sintaxis, el conocimiento general y las relaciones entre las palabras.

Método: El modelo predice la siguiente palabra en una oración, rellenando huecos o completando fragmentos de texto.

El resultado es un modelo base que es muy bueno generando texto, pero no ha sido "instruido" para ser un asistente conversacional. Aún no sabe cómo seguir instrucciones ni mantener una conversación coherente y útil.

2. Ajuste Fino Supervisado (Supervised Fine-Tuning - SFT)

En esta etapa, se le enseña al modelo a ser un asistente. Se utiliza un conjunto de datos mucho más pequeño, pero de muy alta calidad, que ha sido curado por humanos.

Objetivo: El modelo aprende a responder de manera útil, seguir instrucciones y adoptar una "personalidad" de asistente.

Método: Se entrena el modelo con pares de datos de "instrucción/respuesta" o "pregunta/respuesta" generados por humanos. Por ejemplo, "Describe la fotosíntesis" y una respuesta bien escrita sobre la fotosíntesis.

Esto es lo que transforma un modelo de predicción de texto en un modelo que puede seguir comandos como "resume este texto" o "escribe un correo electrónico".

3. Aprendizaje por Refuerzo a partir de Retroalimentación Humana (Reinforcement Learning from Human Feedback - RLHF)

Esta es la etapa final y crítica para la seguridad y la alineación.

Objetivo: El modelo aprende a alinear sus respuestas con los valores y preferencias humanas. Se le enseña a ser útil, honesto y no generar contenido dañino.

Método: Recopilación de preferencias: El modelo genera varias respuestas para una pregunta, y los evaluadores humanos las clasifican de la mejor a la peor.

Modelo de recompensa: Se entrena un modelo separado, llamado "modelo de recompensa," para que aprenda a predecir qué tipo de respuestas preferirán los humanos.

Ajuste: El LLM principal es entrenado con un algoritmo de aprendizaje por refuerzo, usando el modelo de recompensa como guía. El modelo recibe una "recompensa" por generar respuestas que el modelo de recompensa califica como buenas, ajustando sus parámetros para producir resultados cada vez mejores y más seguros.

Este proceso es lo que hace que un LLM sea realmente conversacional y confiable, ya que corrige sutilmente su comportamiento para que se ajuste a las expectativas y la ética humana.

5. Utilice un LLM para resolver algunos de los problemas de programación en el curso: <https://uecp.edunext.io/courses/course-v1:uecp+mlds0a+001/about>. Discuta los resultados.
6. Busque tareas de cursos que haya tomado recientemente que se puedan resolver con un LLM, pruebe el LLM y discuta los resultados.
- **Curso:** Análisis de datos con Python
 - **Tarea:** Realizar un análisis de datos con el archivo "Satisfacción Estudiantil.csv"
 - **Prompt utilizado:** teniendo en cuenta el archivo csv que te comparto, construye el código de python que me permita hacer un análisis de los datos con las siguientes características, el objetivo es hacer un análisis detallado y grafico de la información:
 - Exploración y análisis de los datos
 - Análisis de la satisfacción promedio por programa académico
 - Calcular la satisfacción promedio para cada estudiante
 - Agrupar por programa y calcular el promedio de la satisfacción
 - Crear un gráfico de barras para visualizar la satisfacción promedio
 - Visualizar la distribución de calificaciones
 - Histograma para Satisfacción general, Histograma para Infraestructura
 - Histograma para Calidad docente e Histograma para Apoyo institucional
 - **Resultado:**

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

# --- 1. Cargar y Explorar los datos ---
try:
    df = pd.read_csv('satisfaccion_estudiantil.csv', delimiter=';')
    print("Datos cargados correctamente. Aquí están las primeras 5 filas:")
    print(df.head())
    print("\nInformación general del DataFrame:")
    df.info()
except FileNotFoundError:
    print("Error: El archivo 'satisfaccion_estudiantil.csv' no se encontró. Asegúrate de que esté en el mismo directorio.")
    exit()

# Asegurar que las columnas de calificación sean numéricas.
columnas_calificaciones = ['Satisfaccion_general', 'Infraestructura',
                           'Calidad_docente', 'Apoyo_institucional']
for col in columnas_calificaciones:
    df[col] = pd.to_numeric(df[col], errors='coerce')

# --- 2. Análisis de la satisfacción promedio por programa académico ---
print("\n--- Análisis de Satisfacción Promedio por Programa Académico ---")
# Agrupar por 'Programa' y calcular el promedio de las calificaciones.
```

```
satisfaccion_por_programa =  
df.groupby('Programa')[columnas_calificaciones].mean()  
print(satisfaccion_por_programa)
```

```
# Crear un gráfico de barras para visualizar la satisfacción promedio general por  
programa.
```

```
plt.figure(figsize=(12, 6))  
satisfaccion_por_programa['Satisfaccion_general'].sort_values().plot(kind='bar',  
color='skyblue')  
plt.title('Satisfacción Promedio General por Programa Académico')  
plt.xlabel('Programa Académico')  
plt.ylabel('Satisfacción Promedio (1-5)')  
plt.xticks(rotation=45, ha='right')  
plt.grid(axis='y', linestyle='--', alpha=0.7)  
plt.tight_layout()  
plt.show()
```

```
# --- 3. Calcular la satisfacción promedio para cada estudiante ---
```

```
print("\n--- Satisfacción Promedio de cada Estudiante ---")
```

```
# Calcular la media de las columnas de calificaciones para cada fila (estudiante).
```

```
df['Satisfaccion_promedio_estudiante'] = df[columnas_calificaciones].mean(axis=1)
```

```
print(df[['Id_Estudiante', 'Programa', 'Satisfaccion_promedio_estudiante']].head())
```

```
# --- 4. Visualizar la distribución de calificaciones con histogramas ---
```

```
print("\n--- Distribución de Calificaciones (Histogramas) ---")
```

```
# Configurar la figura para mostrar múltiples gráficos.
```

```
fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(15, 12))
```

```
axes = axes.flatten()
```

```
titulos = ['Satisfacción General', 'Infraestructura', 'Calidad Docente', 'Apoyo  
Institucional']
```

```
for i, col in enumerate(columnas_calificaciones):
```

```
    df[col].hist(bins=np.arange(0.5, 6.5, 1), ax=axes[i], edgecolor='black', alpha=0.7)
```

```
    axes[i].set_title(f'Histograma de {titulos[i]}')
```

```
    axes[i].set_xlabel('Calificación')
```

```
    axes[i].set_ylabel('Frecuencia')
```

```
    axes[i].set_xticks(range(1, 6))
```

```
plt.tight_layout()
```

```
plt.show()
```

```
print("\nAnálisis completado. Se han generado los gráficos y las tablas de  
resumen.")
```

- **Discusión del resultado:** a pesar de que el código generado por el LLM puede ser usado como una aproximación inicial al problema, realmente el código es muy

básico y los resultados pueden ser mejorados de manera drástica, las gráficas y tablas resúmenes que genera de la información, generan muchas inquietudes.

- **Conclusión:** el código requiere revisión y análisis detallado de los resultados con el fin de que sea mejor y los resultados mucho más claros.

7. Use el LLM para encontrar argumentos a favor y en contra de la posibilidad de una inteligencia artificial general.

Argumentos a favor de la Inteligencia Artificial General (AGI)

La AGI, definida como la capacidad de una máquina para entender, aprender y aplicar conocimiento para resolver cualquier tarea intelectual que un ser humano pueda realizar, es vista por muchos como una meta alcanzable e inevitable.

- **Avances tecnológicos exponenciales:** El progreso en la capacidad de computación y el acceso a enormes volúmenes de datos han sido exponenciales. Cada año, las GPUs y los centros de datos se vuelven más potentes y económicos, permitiendo entrenar modelos cada vez más grandes y complejos.
- **Modelos de Transformadores y auto-atención:** El desarrollo de arquitecturas como la de los Transformers, que utilizan el mecanismo de auto-atención, ha sido un punto de inflexión. Estos modelos pueden procesar y entender relaciones complejas en el lenguaje y otros tipos de datos a una escala sin precedentes, lo que se considera un paso fundamental hacia la comprensión multifacética que caracteriza a la AGI.
- **Aprendizaje multimodal:** Los modelos actuales ya están superando las barreras de un solo tipo de datos. El aprendizaje multimodal permite a una IA procesar y conectar información de texto, imágenes, video y audio al mismo tiempo, imitando la forma en que los humanos perciben el mundo. Esto acerca los modelos a una comprensión holística de la realidad.
- **Implicaciones positivas:** La AGI podría resolver problemas que hoy parecen insolubles. Podría acelerar la investigación científica, crear curas para enfermedades, diseñar soluciones sostenibles para el cambio climático y automatizar tareas repetitivas y peligrosas, liberando a la humanidad para dedicarse a la creatividad y la innovación.

Argumentos en contra de la Inteligencia Artificial General (AGI)

A pesar del optimismo, existen obstáculos significativos que plantean dudas sobre si la AGI es realmente alcanzable o si los avances actuales son suficientes.

- **Falta de comprensión causal y sentido común:** Los modelos de IA actuales son expertos en encontrar correlaciones en los datos, pero no comprenden la causalidad subyacente. No tienen sentido común ni conciencia de la realidad física, lo que limita su capacidad para razonar fuera de los datos con los que fueron entrenados. Un modelo puede "saber" que el fuego quema, pero no lo "entiende" como lo haría un ser humano.
- **El problema de la alineación y la ética:** Desarrollar una AGI no es solo un reto técnico, sino también ético. Alinear los objetivos de una AGI con los valores y la seguridad humana es un problema sin resolver. Una AGI con objetivos mal definidos o consecuencias no intencionadas podría actuar de formas impredecibles y catastróficas.
- **Limitaciones del enfoque actual:** El enfoque predominante de entrenar modelos con más datos y más parámetros puede no ser suficiente para alcanzar una verdadera inteligencia. Algunos investigadores argumentan que se necesita un avance conceptual fundamental, no

solo un escalado, para replicar la capacidad humana de aprender de la experiencia, adaptarse a nuevas situaciones y razonar de forma abstracta.

- **Costos de computación insostenibles:** Los modelos más grandes ya requieren cantidades ingentes de energía y recursos financieros, lo que los hace inaccesibles para la mayoría. Lograr la escala necesaria para una AGI podría tener un costo computacional y ambiental tan alto que la haría inviable.

Análisis comparativo

Las posturas a favor y en contra de la AGI no son mutuamente excluyentes, sino que reflejan una tensión entre el **progreso tecnológico** y los **desafíos conceptuales**.

- **Puntos en común:** Ambos lados reconocen el increíble progreso de la IA en los últimos años y el poder de las arquitecturas como los Transformers. También coinciden en que la escala, en términos de datos y computación, es un factor crítico.
- **Tensiones clave:** La principal tensión radica en si el enfoque actual de **escalar modelos** es suficiente para lograr una verdadera inteligencia. Los defensores de la AGI creen que la cantidad se convertirá en cualidad y que los modelos lo suficientemente grandes desarrollarán capacidades emergentes. Por otro lado, los críticos argumentan que los problemas de **comprensión causal, sentido común y alineación** son obstáculos cualitativos que no se resuelven solo con más datos o más potencia de cálculo, y que se necesita una nueva arquitectura o un entendimiento más profundo de la inteligencia misma para superarlos.

8. Use el LLM para indagar sobre los riesgos e implicaciones éticas de la inteligencia artificial.

Principales Riesgos de la Inteligencia Artificial

La rápida adopción de la IA a gran escala ha expuesto una serie de riesgos significativos que requieren una atención urgente.

- **Sesgos Algorítmicos:** Los sistemas de IA aprenden de los datos con los que son entrenados. Si estos datos reflejan prejuicios sociales, raciales o de género, la IA no solo los replica, sino que a menudo los amplifica. Esto puede llevar a decisiones discriminatorias en la contratación de personal, la aprobación de préstamos o la justicia penal.
- **Pérdida de Empleos y Desigualdad Económica:** La automatización impulsada por la IA puede desplazar a trabajadores en sectores como el transporte, la manufactura y los servicios. Esto podría agudizar la desigualdad, concentrando la riqueza y el poder en manos de quienes controlan la tecnología.
- **Vigilancia Masiva y Erosión de la Privacidad:** La IA es una herramienta poderosa para el análisis de datos masivos. Su uso en el reconocimiento facial, el seguimiento del comportamiento en línea y la recolección de datos personales puede facilitar la vigilancia a gran escala por parte de gobiernos y corporaciones, erosionando la privacidad y la libertad individual.

- **Dependencia y Vulnerabilidad Tecnológica:** La integración de la IA en infraestructuras críticas (redes eléctricas, sistemas de transporte, mercados financieros) crea una dependencia que, en caso de fallas o ataques cibernéticos, podría tener consecuencias catastróficas.
- **Autonomía de Sistemas y Usos Militares:** El desarrollo de armas autónomas letales (LAWS) plantea un dilema ético profundo: delegar la decisión de vida o muerte a un algoritmo. Esto no solo deshumaniza el conflicto, sino que también introduce el riesgo de una escalada militar incontrolable.

Implicaciones Éticas Clave

Estos riesgos se manifiestan en dilemas éticos que desafían los principios fundamentales de la sociedad.

- **Justicia y Equidad:** ¿Cómo garantizamos que los beneficios de la IA se distribuyan de manera equitativa y que sus decisiones no discriminen a grupos vulnerables? La equidad algorítmica se convierte en un imperativo ético.
- **Privacidad:** ¿Cómo protegemos los datos personales de los individuos en un mundo donde la IA se nutre de ellos para funcionar? El derecho a la privacidad debe ser redefinido en la era digital.
- **Responsabilidad y Rendición de Cuentas:** Cuando una IA toma una decisión errónea o causa un daño, ¿quién es el responsable? ¿El programador, la empresa que la implementó o la propia IA? La falta de un marco claro de responsabilidad crea una "caja negra" en la que es difícil asignar culpa o exigir reparación.
- **Transparencia:** Para confiar en la IA, debemos entender cómo funciona. Sin embargo, muchos modelos complejos son opacos, lo que dificulta auditar sus decisiones y detectar sesgos. La transparencia se vuelve esencial para la rendición de cuentas y la confianza pública.
- **Derechos Humanos:** El uso de la IA en la justicia, la policía o la vigilancia podría violar derechos fundamentales como la libertad de expresión, la presunción de inocencia y la protección contra la discriminación.

Estrategias y Marcos Éticos para la Mitigación

Para enfrentar estos desafíos, se están desarrollando diversas estrategias que buscan un uso responsable y ético de la IA.

- **Regulación y Gobernanza Internacional:** Los gobiernos y organismos internacionales deben establecer leyes y marcos regulatorios que guíen el desarrollo y la implementación de la IA. Esto incluye normas de privacidad (como el GDPR en Europa), prohibiciones de usos de alto riesgo (como la IA en armas autónomas) y estándares de transparencia.
- **Principios Éticos y Códigos de Conducta:** Organizaciones, empresas y profesionales pueden adoptar principios éticos (ej. no maleficencia, beneficencia, autonomía, justicia) que sirvan como una brújula moral para sus proyectos de IA. Esto promueve una cultura de responsabilidad desde el diseño.

- **Buenas Prácticas de Desarrollo:** Los equipos de IA deben incorporar la ética en el ciclo de vida del desarrollo. Esto incluye:
 - **Auditoría de Datos:** Evaluar y limpiar los conjuntos de datos para eliminar sesgos antes del entrenamiento.
 - **Evaluación de Impacto:** Realizar análisis para prever y mitigar los posibles daños de un sistema de IA antes de su despliegue.
 - **Diseño Centrado en el Humano:** Involucrar a las comunidades afectadas en el diseño de las soluciones para asegurar que se respeten sus necesidades y valores.

Resumen Crítico

Los riesgos de la IA no son meros efectos secundarios, sino implicaciones directas de su poder y su naturaleza. La **justicia** se ve amenazada por el sesgo algorítmico, la **privacidad** se erosiona por la vigilancia masiva, y la **responsabilidad** se diluye por la complejidad de los sistemas autónomos.

La adopción de la IA en la sociedad, tal como se plantea hoy, no puede ser un camino sin control. La tecnología avanza, pero los marcos éticos y regulatorios deben seguirle el ritmo. El verdadero desafío no es solo técnico, sino humano: se trata de cómo garantizamos que esta poderosa herramienta sirva para el bien común, promueva la equidad y respete los derechos fundamentales de todos. La mitigación de los riesgos no es opcional; es un requisito para un futuro en el que la inteligencia artificial coexista de manera segura y beneficiosa con la humanidad.

9. Enlace Repositorio GitHub:

<https://github.com/kdgenez/Taller-1.-Aprendizaje-de-Maquinas-Aplicado>