

Taller 1: Introducción a Sistemas Inteligentes

Juan Esteban Avendaño Castaño
CC 1027800082

1. Definiciones breves

Big Data

Conjunto de datos tan grandes, complejos y de crecimiento rápido que superan la capacidad de las herramientas tradicionales para ser almacenados, procesados y analizados. Se caracteriza por las **5 V**:

- **Volumen** (gran cantidad de datos),
- **Velocidad** (alta frecuencia de generación),
- **Variedad** (diferentes tipos: texto, imágenes, audio, etc.),
- **Veracidad** (calidad y fiabilidad de los datos),
- **Valor** (capacidad de extraer información útil).

Machine Learning (Aprendizaje automático)

Rama de la inteligencia artificial que se centra en diseñar algoritmos y modelos que permiten a las computadoras **aprender patrones a partir de datos** y mejorar su rendimiento en tareas específicas **sin ser programadas explícitamente**.

Inteligencia Artificial (IA)

Disciplina de la informática que busca desarrollar sistemas capaces de **simular procesos cognitivos humanos**, como el razonamiento, la toma de decisiones, la resolución de problemas, el reconocimiento de patrones, la comprensión del lenguaje y el aprendizaje.

Ciencia de Datos (Data Science)

Campo interdisciplinario que combina **estadística, informática y conocimiento del dominio** para extraer información significativa de los datos. Incluye la recolección, limpieza, análisis, visualización y comunicación de resultados. Es el puente entre **datos brutos** y **decisiones informadas**.

Deep Learning (Aprendizaje profundo)

Subcampo del **machine learning** que utiliza **redes neuronales artificiales con muchas capas (profundas)**. Permite modelar relaciones muy complejas y ha impulsado avances en visión por computadora, procesamiento de lenguaje natural y sistemas de recomendación.

Minería de Datos (Data Mining)

Proceso de **explorar grandes volúmenes de datos** con técnicas estadísticas, de aprendizaje automático y algoritmos computacionales para descubrir **patrones ocultos, correlaciones o tendencias**. Es un paso clave dentro de la ciencia de datos y de la inteligencia de negocios.

Inteligencia de Negocios (Business Intelligence, BI)

Conjunto de estrategias, procesos y herramientas que transforman datos empresariales en **información útil para la toma de decisiones**. Incluye la integración de bases de datos, reportes, cuadros de mando e indicadores de desempeño (KPIs).

Estadística

Disciplina matemática que se ocupa de **recoger, organizar, analizar e interpretar datos**. Proporciona los fundamentos teóricos para la ciencia de datos, machine learning y minería de datos, ayudando a validar hipótesis y a medir la incertidumbre.

2. Diagrama de relaciones ([Mermaid.js](#))

```
graph TD
    %% Definición de estilo jerárquico
    subgraph IA[Inteligencia Artificial (IA)]
        B[Machine Learning]
        C[Deep Learning]
        D[Minería de Datos]
    end

    subgraph Datos[Fuentes y Fundamentos]
        E[Big Data]
        F[Estadística]
    end

    subgraph CD[Ciencia de Datos]
        G1[Integración de Datos]
        G2[Modelado y Análisis]
        G3[Comunicación de Resultados]
    end

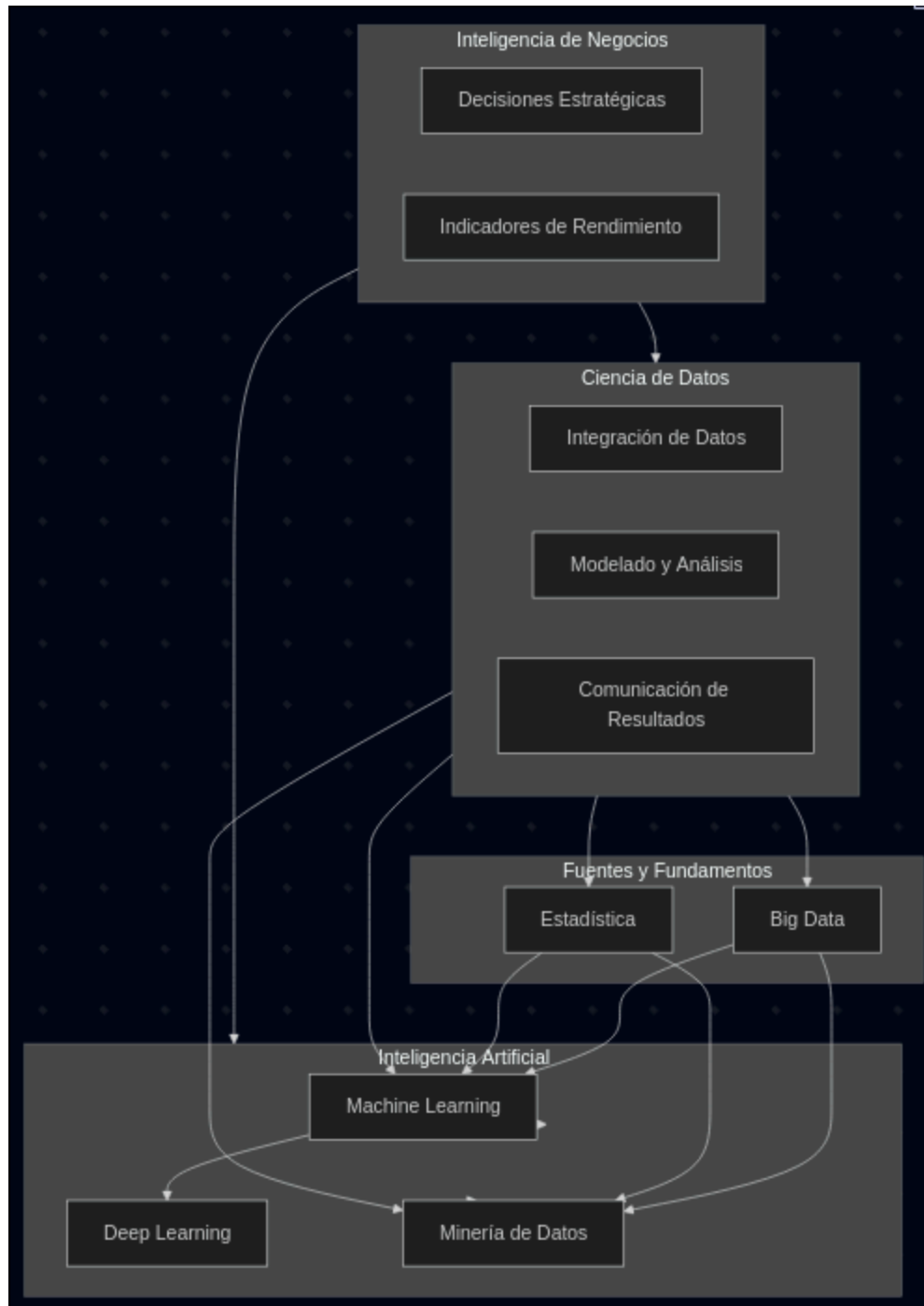
    subgraph Negocios[Inteligencia de Negocios]
        H1[Decisiones Estratégicas]
        H2[Indicadores de Rendimiento (KPIs)]
    end

    %% Relaciones principales
    B --> C
    IA --> B
    IA --> D

    E --> B
    E --> D
    F --> B
    F --> D

    CD --> E
    CD --> F
    CD --> B
    CD --> D

    Negocios --> CD
    Negocios --> IA
```



3. Video reciente sobre una aplicación/ técnica de IA

[Google Veo 3 Demo - Accents and Voices](#)

- Presenta múltiples escenas generadas por Veo 3 a partir de prompts textuales.
- Demuestra que Veo 3 genera audio (efectos, ambiente y diálogos) sincronizados con la imagen.
- Muestra control sobre el estilo cinematográfico, física aproximada y adherencia al prompt.

4. ¿Qué son los Transformers?

Los **Transformers** son una **arquitectura de redes neuronales** introducida en 2017 por Vaswani et al. en el artículo *“Attention is All You Need”*.

Su gran innovación es el **mecanismo de atención (attention mechanism)**, que permite al modelo **identificar qué partes de una secuencia son más relevantes** sin necesidad de procesarlas de manera estrictamente secuencial como las RNN o LSTM.

Esto hizo que los transformers fueran mucho más **eficientes, escalables y capaces de manejar dependencias largas en el tiempo**, lo cual revolucionó el procesamiento de lenguaje natural (NLP).

a) Ideas principales de su funcionamiento

1. **Mecanismo de Atención (Self-Attention):**

Permite que cada palabra (o token) en una secuencia “preste atención” a las demás para entender el contexto.

Ejemplo: en la frase *“El gato se sentó en la alfombra porque estaba cansado”*, el modelo aprende que *“estaba”* se refiere a *“gato”*.

2. **Codificación Posicional (Positional Encoding):**

Como el transformer no procesa secuencias de forma lineal, necesita agregar información de **orden y posición** para que entienda la secuencia.

3. **Estructura en Bloques (Encoder – Decoder):**

- **Encoder:** procesa la entrada y genera representaciones internas.
- **Decoder:** genera la salida, usando tanto la representación del encoder como el contexto ya generado.

4. **Paralelización:**

A diferencia de las RNN, los transformers permiten entrenar con **procesamiento paralelo**, lo que los hace escalables a enormes cantidades de datos.

b) Aplicaciones de los Transformers

- **Procesamiento de Lenguaje Natural (NLP):**
Traducción automática (Google Translate), chatbots, generación de texto (GPT, LLaMA, Bard, etc.), resumen automático, análisis de sentimientos.
- **Visión por Computadora (ViT – Vision Transformers):**
Reconocimiento de imágenes, clasificación de objetos, segmentación de imágenes médicas.
- **Multimodalidad:**
Modelos que combinan texto + imágenes + audio (ej. CLIP de OpenAI o Gemini de Google).
- **Bioinformática:**
Predicción de estructuras de proteínas (AlphaFold de DeepMind).
- **Series temporales y predicción:**
Aplicaciones en finanzas, salud, clima.

c) Grandes Modelos de Lenguaje (LLM) y relación con Transformers

- Un **LLM (Large Language Model)** es un modelo entrenado con **billones de parámetros** sobre cantidades masivas de texto, con el objetivo de **predecir la siguiente palabra o token** en una secuencia.
- Los **Transformers** son la arquitectura que permite entrenar LLMs de manera eficiente y escalable.
- Ejemplos de LLM basados en transformers: **GPT (OpenAI), BERT, LLaMA, Claude, Gemini, Mistral**.
- Relación: sin la arquitectura Transformer, **no sería posible entrenar modelos tan grandes** ni lograr el rendimiento actual en generación y comprensión del lenguaje.

d) ¿Cómo se entrena un LLM conversacional?

El entrenamiento suele darse en **tres etapas principales**:

1. **Pre-entrenamiento (Pretraining):**
 - Se entrena al modelo en grandes corpus de texto (internet, libros, artículos, etc.) para aprender patrones generales del lenguaje.
Tarea principal: **modelado del lenguaje** (predecir la siguiente palabra/token).
2. **Ajuste fino (Fine-tuning):**
 - Se entrena el modelo con datos más específicos para una tarea (ej. responder preguntas, resumir textos, generar código).
 - Se usan datasets curados, más limpios y relevantes.

3. **Alineación con retroalimentación humana (RLHF – Reinforcement Learning with Human Feedback):**

- Entrenadores humanos evalúan respuestas del modelo.
- Se entrena un “modelo de recompensa” que enseña al LLM a **dar respuestas útiles, seguras y alineadas**.
- Se aplica **aprendizaje por refuerzo** para mejorar la calidad de la conversación.

5. **Usar un LLM para resolver problemas de programación del curso**

Ejercicio: U2 | Práctica: Cursores y áreas

Prompt: Le, pase toda la descripción del problema, cómo son las entradas y cómo son las salidas, acá un resumen hecho por el LLM mismo:

Se tiene un **rectángulo** definido por su esquina inferior izquierda (x,y) , su ancho www y su altura hhh .

También se tiene un **punto** (px,py) que representa la posición del cursor. El objetivo es determinar si el punto está **dentro del rectángulo** (incluyendo los bordes).



Entrada (6 líneas):

1. xxx : posición horizontal de la esquina inferior izquierda del rectángulo.
2. yyy : posición vertical de la esquina inferior izquierda.
3. www : ancho del rectángulo.
4. hhh : altura del rectángulo
5. $pxpx$: posición horizontal del punto.
6. $pypy$: posición vertical del punto.



Salida:

- **True** si el punto está dentro o en el borde del rectángulo.
- **False** si está fuera.

👉 En resumen: el problema es verificar con comparaciones si el cursor está dentro de un área rectangular definida en coordenadas cartesianas.

El siguiente fue el código generado:

```
# Entrada
x = float(input()) # esquina inferior izquierda (x)
y = float(input()) # esquina inferior izquierda (y)
w = float(input()) # ancho
h = float(input()) # altura
px = float(input()) # posición horizontal del punto
py = float(input()) # posición vertical del punto

# Verificación
inside = (x <= px <= x + w) and (y <= py <= y + h)

# Salida
print(inside)
```

Y este fue el resultado, obteniendo un 100%:

? Understanding your result

Your answer passed the tests! Your score is 100.0%. [Submission #68cb788284ba4a81073cd007]

- Test 1: ACCEPTED
- Test 2: ACCEPTED
- Test 3: ACCEPTED
- Test 4: ACCEPTED
- Test 5: ACCEPTED
- Test 6: ACCEPTED
- Test 7: ACCEPTED
- Test 8: ACCEPTED
- Test 9: ACCEPTED

Fortalezas: rapidez para generar soluciones, explicaciones paso a paso, variantes (más eficientes/pythónicas), ayuda con debugging y tests unitarios.

Limitaciones y riesgos: a veces el LLM inventa funciones inexistentes, retorna código con pequeños errores sintácticos o sin manejar casos borde; no sustituye la comprensión conceptual. Aunque en esta tarea se desempeñó bastante bien.

6. Buscar tareas de cursos recientes que se puedan resolver con un LLM

Escogí el curso: [Developing Secure Software \(LFD121\)](#)

Ejercicio: Lab: Countering ReDoS Attacks on Regular Expressions

El código generado se ve de la siguiente manera:

```
const express = require("express");
const app = express();
const { query, matchedData, validationResult } =
  require("express-validator");

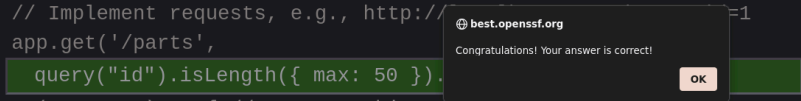
app.get("/parts",
  query("id").isLength({ max: 50 }).matches(/^([A-Z0-9]+$/),
  (req, res) => {
    const result = validationResult(req);
    if (result.isEmpty()) {
      const data = matchedData(req);
      return res.send(`You requested part id ${data.id}!`);
    }
    res.status(422).send("Invalid input");
  }
);

app.listen(3000, () => console.log("Server running on port 3000"));
```

Al reemplazar la parte que falta en el ejercicio, obtengo el siguiente resultado:

```
// Set up Express framework and express-validator library
const express = require("express");
const app = express();
const { query, matchedData, validationResult } =
  require('express-validator');

// Implement requests, e.g., http://
app.get('/parts',
  query("id").isLength({ max: 50 }).
  (req, res) => { // Execute this code if /parts seen
    const result = validationResult(req); // Retrieve errors
    if (result.isEmpty()) { // No errors
      const data = matchedData(req); // Retrieve matching data
      return res.send(`You requested part id ${data.id}!`);
    }
    res.status(422).send(`Invalid input`);
  })
```



7. Argumentos a favor de la posibilidad de la AGI

1. **Avances en escalabilidad de modelos (LLMs y Transformers):**
 - Los modelos actuales (GPT, Gemini, Claude, LLaMA) muestran capacidades emergentes al aumentar su tamaño y datos de entrenamiento.
 - La tendencia sugiere que al escalar, pueden acercarse a comportamientos más generales.
2. **Evidencia de capacidades generales en modelos actuales:**
 - LLMs no solo predicen texto, sino que ya **razonan, programan, traducen, hacen matemáticas y generan imágenes o código**.
 - Esto indica una **transferencia de habilidades** entre dominios, un rasgo propio de la inteligencia general.
3. **Plasticidad de la inteligencia biológica como prueba de posibilidad:**
 - El cerebro humano es también una “máquina de aprendizaje” basada en neuronas.
 - Si la naturaleza pudo generar inteligencia general a través de un sistema biológico, **en principio no hay razones físicas que impidan replicarlo en máquinas**.
4. **Combinación de modalidades:**
 - La integración de texto, imagen, audio, vídeo y acción (modelos multimodales) da a la IA una visión más cercana al aprendizaje humano.
5. **Avances en neurociencia computacional:**
 - El estudio del cerebro y la simulación de redes neuronales más parecidas a las biológicas puede inspirar arquitecturas con mayor generalidad.

Argumentos en contra de la posibilidad (o dificultad) de la AGI

1. **Limitaciones arquitectónicas actuales (Transformers y LLMs):**
 - Aunque impresionantes, los modelos actuales no **entienden** el mundo, solo reconocen patrones estadísticos.
 - No tienen **memoria a largo plazo robusta**, **metacognición** ni una noción de **sentido común profundo**.
2. **Problema de transferencia de conocimiento:**
 - Los humanos aplicamos aprendizajes de un dominio a otro muy distinto (ej. usar lógica matemática en economía).
 - Las IAs actuales aún tienen problemas para transferir conocimiento más allá de lo entrenado.

3. **Complejidad del entorno humano:**

- La inteligencia humana no es solo cálculo, también incluye **emociones, motivaciones, conciencia, intuición social**.
- Replicar estos aspectos es un desafío todavía abierto en ciencia cognitiva.

4. **Costos computacionales y energéticos enormes:**

- Los LLM actuales requieren **miles de GPUs, millones de dólares y cantidades descomunales de energía** para entrenar.
- Esto podría hacer inviable escalar indefinidamente hasta alcanzar AGI.

5. **Indefinición conceptual:**

- No existe una definición unánime de qué sería exactamente una AGI. Mientras no haya un marco teórico claro, es difícil saber si lo que construimos puede considerarse “inteligencia general”.

8. Riesgos e implicaciones éticas de la IA

Los **riesgos e implicaciones éticas de la inteligencia artificial (IA)** abarcan varias dimensiones sociales, políticas y técnicas. Aquí te los organizo en categorías claras:

1. Sesgos y discriminación

- Los modelos aprenden de datos históricos que reflejan desigualdades sociales.
- Resultado: pueden discriminar por género, raza, edad u otros factores en procesos como selección de personal, créditos, justicia predictiva.

2. Privacidad y vigilancia

- Los sistemas pueden almacenar, inferir o filtrar información sensible.
- Riesgo de usos masivos para vigilancia estatal o corporativa sin consentimiento adecuado.

3. Desinformación y manipulación

- Con modelos generativos (texto, imágenes, audio, video) se pueden crear **deepfakes** o noticias falsas.
- Amenaza a la confianza pública y a la democracia.

4. Impacto laboral y económico

- Automatización de tareas rutinarias y hasta creativas.
- Puede aumentar la productividad, pero también generar desempleo o precarización si no se acompaña de políticas públicas.

5. Concentración de poder

- El desarrollo de modelos avanzados requiere enormes recursos de datos y cómputo.
- Pocas empresas/países controlan la tecnología, lo que genera desigualdades globales.

6. Seguridad y uso malicioso

- Riesgo de uso para ciberataques, creación de malware, manipulación de mercados o incluso armamento autónomo.

7. Responsabilidad y transparencia

- Dificultad para explicar decisiones de modelos complejos (caja negra).
- Problemas legales: ¿quién responde ante un error o daño causado por una IA?

8. Dimensiones éticas y filosóficas

- Riesgo de cosificación de las personas (p. ej. trato humano automatizado que sustituya empatía real).
- Debate sobre el desarrollo hacia una IA general (AGI) y sus implicaciones en autonomía, control y alineamiento con valores humanos.