**Tarea 3: “Motores de Inferencia”**

**Angel Bernardo Márquez Valdivia**

**22110348**

**6E**

**Sistemas Expertos**



**01/03/2025**

**Motores de Inferencia**

Un motor de inferencia es un componente esencial en los sistemas de inteligencia artificial y en los sistemas expertos. Su función principal es procesar información basada en reglas, datos o modelos entrenados para tomar decisiones, generar predicciones o realizar deducciones.

En términos generales, los motores de inferencia pueden dividirse en dos grandes categorías:

* **Motores de inferencia en Sistemas Expertos:** Se basan en reglas lógicas predefinidas y bases de conocimiento para tomar decisiones.
* **Motores de inferencia en Aprendizaje Automático e IA:** Utilizan modelos entrenados en grandes volúmenes de datos para realizar inferencias probabilísticas.

**Motores de Inferencia de la actualidad**

* **Novita AI LLM Inference Engine**

Novita AI ha desarrollado un motor de inferencia para modelos de lenguaje de gran tamaño (LLM) que se distingue por su alto rendimiento y rentabilidad. Es capaz de procesar hasta 130 tokens por segundo con el modelo Llama-2–70B-Chat y 180 tokens por segundo con el modelo Llama-2–13B-Chat.

* **OpenAI o3 y o3-mini**

OpenAI ha introducido su modelo de inferencia de nueva generación, conocido como o3, que incluye dos versiones: o3 y o3-mini. Estos modelos han demostrado avances significativos hacia la Inteligencia Artificial General (AGI), obteniendo una puntuación del 87.5% en el test de referencia ARC-AGI, superando la media humana. Destacan en tareas matemáticas y de programación, y han sido entrenados con técnicas de "alineación deliberativa" para garantizar su seguridad.

* **DeepSeek R1**

DeepSeek, una empresa china de inteligencia artificial ha desarrollado el modelo de razonamiento DeepSeek R1. Este modelo ha superado en descargas a ChatGPT en China y Estados Unidos, alcanzando una precisión del 97% en la resolución de problemas matemáticos y superando al 96% de los participantes humanos en pruebas de programación. Su desarrollo ha sido notablemente más económico, con un costo aproximado de seis millones de dólares, lo que lo posiciona como una alternativa competitiva en el mercado de la IA.

**Aplicaciones de los motores de inferencia**

Los motores de inferencia tienen una amplia gama de aplicaciones en diversos sectores:

* **Sector de la salud:** Permiten a los especialistas comparar historiales clínicos con datos actuales para detectar patrones y anomalías rápidamente, facilitando la identificación temprana de riesgos para la salud.
* **Sector financiero:** Identifican errores o anomalías en tiempo real, detectando posibles fraudes de manera rápida y anticipada, lo que optimiza los recursos y protege la privacidad de los clientes.
* **Sector automotor:** En vehículos autónomos, los motores de inferencia procesan datos en tiempo real de sensores para detectar objetos, peatones y señales de tráfico, permitiendo una conducción segura.

**Métodos de Inferencia**

Los métodos de inferencia son técnicas utilizadas en motores de inferencia para deducir información a partir de datos, reglas o modelos matemáticos. En el ámbito de la inteligencia artificial (IA) y los sistemas expertos, estos métodos pueden clasificarse en deductivos, inductivos y probabilísticos.

1. **Métodos de Inferencia en Sistemas Expertos**

Los sistemas expertos utilizan reglas lógicas y bases de conocimiento para tomar decisiones. Sus métodos de inferencia principales incluyen:

1. **Encadenamiento hacia adelante (Forward Chaining)**

* Parte de hechos iniciales y aplica reglas para obtener nuevas conclusiones.
* Se usa en sistemas de diagnóstico y toma de decisiones.

**Ejemplo:** Un sistema médico que determina una enfermedad a partir de síntomas.

1. **Encadenamiento hacia atrás (Backward Chaining)**

* Parte de una hipótesis y busca pruebas para confirmarla.
* Se usa en sistemas de planificación y resolución de problemas.

Ejemplo: Un sistema de IA que verifica si un paciente tiene una enfermedad específica buscando síntomas confirmatorios.

1. **Inferencia basada en reglas (Rule-Based Inference)**

* Utiliza reglas lógicas "Si-Entonces" para inferir conclusiones.
* Se encuentra en motores de reglas como Drools y CLIPS.

1. **Métodos de Inferencia en Aprendizaje Automático y Redes Neuronales**

En el Machine Learning y Deep Learning, la inferencia se realiza sobre modelos entrenados con grandes volúmenes de datos.

1. **Inferencia Bayesiana**

* Se basa en la teoría de probabilidad y el Teorema de Bayes.
* Se usa en modelos como Redes Bayesianas y PyMC.

**Ejemplo:** Predicción de diagnóstico médico basado en síntomas con probabilidades.

1. **Inferencia Probabilística**

* Se usa en Redes Neuronales Probabilísticas y modelos de Markov.
* Ayuda en procesamiento de lenguaje natural (NLP) y reconocimiento de imágenes.

**Ejemplo:** Algoritmos de detección de fraudes en bancos.

1. **Inferencia basada en Aprendizaje Profundo (Deep Learning Inference)**

* Usa redes neuronales preentrenadas en grandes conjuntos de datos.
* Se acelera con hardware como TensorRT, ONNX Runtime y TFLite.

**Ejemplo:** Reconocimiento facial en dispositivos móviles.

1. **Inferencia Lógica en Redes Neuronales**

* Integra lógica simbólica con aprendizaje profundo.
* Se usa en modelos de razonamiento matemático y sistemas híbridos de IA.