

Manual de Estudio: AWS Certified AI Practitioner

Dominio 2.2: Capacidades y Limitaciones en el Ámbito Empresarial

Material de Preparación Detallado

1. Ventajas de la IA Generativa en la Empresa

A diferencia de los modelos de IA tradicionales, que suelen ser complejos y caros de construir, la IA generativa ofrece tres pilares fundamentales:

- **Adaptabilidad:** Una sola tecnología puede aplicarse a múltiples sectores y tareas.
- **Capacidad de respuesta:** Generación de contenido en tiempo real.
- **Simplicidad:** Permite crear aplicaciones útiles de manera más rápida y a un menor coste que los modelos personalizados desde cero.

El Test del Niño de 10 Años

Una regla práctica para evaluar si una tarea es apta para un LLM: Si un niño de 10 años puede seguir las instrucciones de la petición (prompt) y completar la tarea (ej. identificar si un correo es una reclamación), el modelo probablemente también podrá hacerlo.

2. Limitaciones y Riesgos Técnicos

Es vital entender qué NO puede hacer un modelo para construir sistemas éticos y seguros.

2.1. Problemas de Comportamiento y Alucinaciones

- **Falta de Memoria Nativa:** Los LLM no recuerdan conversaciones anteriores de forma automática (son *stateless*); cada petición es como preguntar a un "niño diferente".
- **Lenguaje Tóxico o Agresivo:** Los modelos pueden heredar tonos combativos presentes en sus datos de entrenamiento (internet).
- **Alucinaciones:** El modelo genera una respuesta con total seguridad que es incorrecta o engañosa. Se requiere contrastar la información con fuentes autorizadas.

2.2. Valores Humanos en el Refinamiento

Para mitigar estos riesgos, se utiliza el **refinamiento con comentarios humanos**, buscando alinear el modelo con tres valores fundamentales:

1. **Amabilidad (Helpfulness)**: El modelo debe ser útil.
2. **Honestidad (Honesty)**: El modelo no debe mentir.
3. **Inocencia/Seguridad (Harmlessness)**: El modelo no debe generar contenido dañino o ilegal.

3. Interpretabilidad del Modelo

Existe un equilibrio (*trade-off*) entre el **rendimiento** (qué tan buena es la predicción) y la **interpretabilidad** (por qué el modelo tomó esa decisión).

Tipo de Análisis	Análisis Intrínseco	Análisis Post hoc
Aplicación	Modelos de baja complejidad y relaciones simples.	Modelos complejos como redes neuronales.
Nivel de Vista	Global (comportamiento general).	Local (un punto de dato) o Global.
Inconveniente	Menor rendimiento al no capturar interacciones no lineales.	Son mecanismos independientes del modelo para interpretar salidas.

4. Métricas de Evaluación para GenAI

A diferencia del ML tradicional, que es **determinista** (la exactitud es fácil de medir), la IA generativa es **no determinista** y requiere métricas basadas en lenguaje.

- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation)**: Utilizada para evaluar la calidad de los **resúmenes** automáticos comparándolos con referencias humanas.
- **BLEU (Bilingual Evaluation Understudy)**: Utilizada para evaluar la calidad de las **traducciones** automáticas.

5. Selección del Modelo y Métricas de Negocio

La elección del modelo depende de la arquitectura (VAE, GAN o Autorregresivos) y de los requisitos de conformidad.

5.1. Factores de Éxito Empresarial

Para obtener valor real, la organización debe integrar los Modelos Fundacionales (FM) con sus sistemas existentes:

- **Integración**: Conexión con bases de datos, ERP y CRM.

- **Calidad de Salida:** Medida por relevancia, exactitud, coherencia e idoneidad.
- **Eficiencia:** Medida por la reducción de esfuerzos manuales y tasas de finalización de tareas.
- **ROI:** Evaluar el coste de infraestructura frente a las ganancias en productividad.

5.2. Métricas de Negocio Clave (KPIs)

- **CLTV (Customer Lifetime Value):** Maximizado mediante experiencias personalizadas generadas por IA.
- **ARPU:** Promedio de ingresos por usuario.
- **Rendimiento entre dominios:** Evaluar cómo el conocimiento se transfiere entre diferentes tareas empresariales.