

Generative AI Fundamentals

jueves, 20 de julio de 2023 22:33

Subconjunto de la inteligencia artificial que se enfoca en generar nuevo contenido.

IA: Crear sistemas capaces de imitar y superar la inteligencia humana.

Por que el auge de estos modelos?

- Grandes conjuntos de datos.
- Alto poder computacional.
- Avances en modelo de deeplearning.

LLMs (Large Language Models):

Procesamiento del lenguaje avanzado, basado en redes neuronales

Foundation Model:

Model de ML entrenado y afinado para el entendimiento de lenguajes en específico y generacion de tareas.

- Encoder: Toma un texto de entrada y lo convierte a tokens, luego los transforma en valores numericos,
- Entrenamiento: Se entrena un Transformer pre entrenado con los tokens.
- Human Feedback (opcional)
- Decoder: Convierte los tokens dados a palabras.

TUS DATOS te diferencian de la competencia, toma accion basada en ellos.

Factores a tener en cuenta a la hora de escoger un LLM:

- Privacidad: Como va a ser usada tu data. Data handlig, storage and deletion.
- Calidad: Entender como fue entrenado el modelo es importante. Accuracy and realiability of the models predictions.
- Costo: Requisitos de infra. Costos por adquirir el modelo. Costos por mantenimiento.
- Latencia: Tiempo de procesamiento y respuesta se ajusta a las necesidades del negocio.

LLMS as a service: LLMS de terceros

| Pros | Cons |
|---|---|
| <ul style="list-style-type: none"> • Speed of development <ul style="list-style-type: none"> • Quick to get started and working. • As this is another API call, it will fit very easily into existing pipelines. • Quality <ul style="list-style-type: none"> • Can offer state-of-the-art results | <ul style="list-style-type: none"> • Cost <ul style="list-style-type: none"> • Pay for each token sent/received. • Data Privacy/Security <ul style="list-style-type: none"> • You may not know how your data is being used. • Vendor lock-in <ul style="list-style-type: none"> • Susceptible to vendor outages, deprecated features, etc. |

Open Source LLM models:

| Pros | Cons |
|--|---|
| <ul style="list-style-type: none"> • Task-tailoring <ul style="list-style-type: none"> • Select and/or fine-tune a task-specific model for your use case. • Inference Cost <ul style="list-style-type: none"> • More tailored models often smaller, making them faster at inference time. • Control <ul style="list-style-type: none"> • All of the data and model information stays entirely within your locus of control. | <ul style="list-style-type: none"> • Upfront time investments <ul style="list-style-type: none"> • Needs time to select, evaluate, and possibly tune • Data Requirements <ul style="list-style-type: none"> • Fine-tuning or larger models require larger datasets. • Skill Sets <ul style="list-style-type: none"> • Require in-house expertise |

Fine-tuning: Entrenar mas un modelo pre entrenado para adaptarlo a unos datos y necesidades particulares.

Como hacer fine-tuning?

- Empezar con foundational model
- Comenzar un entrenamiento supervisado usando conjuntos de datos pequeños etiquetados.
- Ejemplo: Si se necesita para responder preguntas, le damos un conjunto de datos de preguntas y respuestas.

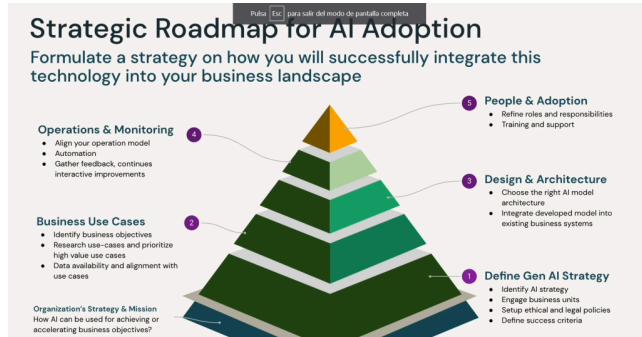
En algunos casos puede ser necesario usar múltiples LLMs en el flujo de trabajo por medio de cadenas (langchain)

Una herramienta para encadenar modelos permite la integración de estos modelos guardando el estado de la cadena en una base de datos vectorizada para la coordinación de los modelos.

Montar aplicaciones basadas en LLM ciclo de vida:

- Preparar tus datos para ML.
- Encontrar y ajustar los modelos
- Desplegar y servir la aplicación.

Buena práctica: Hacer todas estas dentro de la misma plataforma y no en tres sistemas distintos.



Desafíos y riesgos:

Legal:

En este momento los modelos no poseen una opción de "olvidar" datos personales.

- Definir los permisos que se necesitarían.
- Que datos van a ser usados para entrenar modelos o compartir con terceros.
- El historico de interacción del usuario se almacena? Es seguro el almacenamiento?

Mejores prácticas:

- Anonimizar y encriptar los datos con controles de acceso
- Definir una política de uso de datos.
- Establecer un gobierno de datos: controlar versiones, monitoring, auditing.

Seguridad:

- Data Leaks: los LLMs tienen la capacidad de memorizar y volver a reproducir datos de entrenamiento.
- Prompt Injection: Manipular el comportamiento normal del modelo.

Consideraciones éticas:

Sesgo: Dos tipos de sesgo en los datos:

Sesgo humano: Percepciones sociales, estereotipos y factores históricos. Conocimientos pre-concebidos, factores culturales, etc.

Sesgo humano anotado: Errores o limitaciones en el razonamiento y juicio humano.

Ciclo de refuerzo del sesgo:

Sesgo en los datos de entrada son aprendidos por el modelo, el modelo saca datos sesgados y las personas deciden / aprenden sobre estos datos sesgados. Estos nuevos datos son dados al modelo reforzando el sesgo presente en el mismo.

Alucinación: Cuando el modelo genera respuestas que suenan plausibles pero no precisas, inadecuadas o sin sentido debido a limitaciones en la comprensión.

Puede resultar en la degradación de la calidad de la información

Alucinación intrínseca: El modelo produce una salida que contradice directamente los datos dados en la entrada.

Alucinación no intrínseca: El modelo produce una salida que no puede ser confirmada dados los datos de entrada.

Como enfrentarse a estos problemas:

- Establecer revision de calidad de los datos dados por los modelos.
- Examinar la data y actualizarla de una forma mas frecuente para combatir el sesgo en los mismos.
- Indagar la fuente de datos. ¿Son coherentes y verdaderos los datos de entrada a mi modelo?
- Regular el uso de LLMs

Auditorias:

Governance audit: Ver que compañías y tecnologías hay detras de los modelos.

Model audit: Auditar los mismos modelo antes de su despliegue al publico.

Application Audit: Auditar en base al uso que hacen los usuarios del modelo.

How will AI Impact Society

Impact on the workforce

| Pro Arguments | Counter Arguments |
|---|---|
| <ul style="list-style-type: none">• Personalization: Enables personalized experiences in our life• Automation and Efficiency: AI will be used for repetitive tasks → Increased efficiency and higher productivity• Accessibility: GenAI making technology more inclusive and accessible by generating alternative formats, providing real-time translations, and assisting individuals with disabilities | <ul style="list-style-type: none">• Job Displacement: AI automation may lead to job losses or displacement of workers → economic inequalities and unemployment• Ethical Concerns: Entrench existing discrimination and biases.• Overreliance: The increased trust and reliance on AI systems may lead to unnoticed mistakes and loss of important skills• Privacy & Security: Privacy concerns, cyber threats and malicious attacks, AI being used for political goals |