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 sobrepasar 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

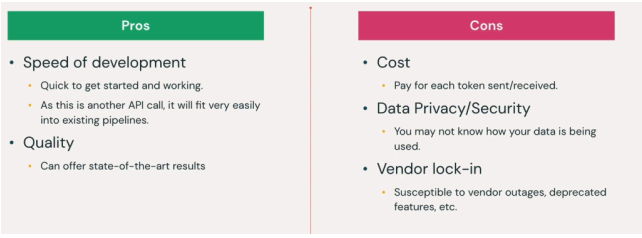
Foundation Model:
Model de ML entrenado y afinado para el entendimiento de lenguajes en especifico y generacion de tareas.

- 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.

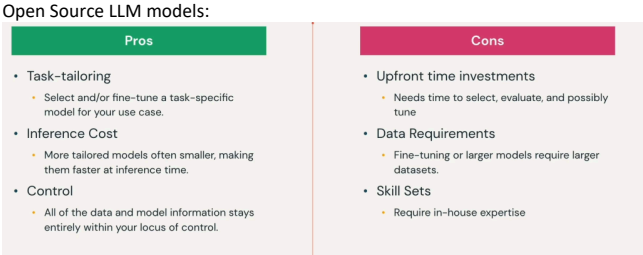
TUS DATOS te diferencian de la competencia, toma accion basada en ellos.

Factores a tener en cuenta a la hora de escoger un LLM:
- Privacidad: Como va a ser usada tu data. Data handilg, 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



Open Source LLM models:



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

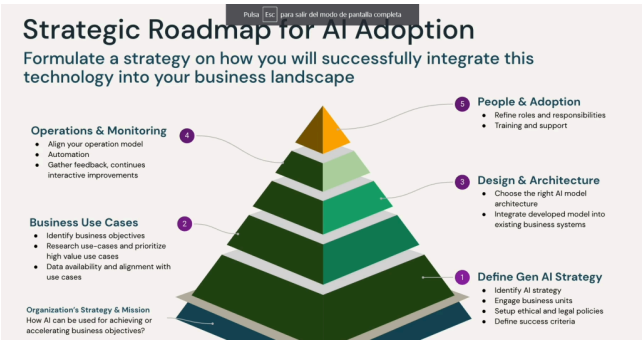
Como hacer fine-tuning?
- Empezar con foundational model
- Comenzar un entrenamiento supervisado usando conjuntos de datos pequeños etiquetados.
- Ejemplo: Si se necesita para responder preguntas, le damos un conjunto de datos de preguntas y respuestas.

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 integracion de estos modelos guardando el estado de la cadena en una base de datos vectorizada para la coordinacion de los modelos.

Montar aplicaciones basadas en LLM ciclo de vida:

- Prepara 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 opcion de "olvidar" datos personales.

- Definir los permisos que se necesitarian.

- Que datos van a ser usados para entrenar modelos o comprtir con terceros.

- El historico de interaccion del usuario se almacena? Es seguro el almacenamiento?

Mejores prácticas:

- Anonimizar y encriptar los datos con controles de acceso

- Defnir una politica de uso de datos.

- Establecer un gobierno de datos: controlar versiones, monitoring, auditing.

Seguridad:

- Data Leaks: os LLMs tienen la capacidad de memorizar y volver a reproducir datos de entrenamiento.

- Prompt Inyection: Manipular el comportamiento normal del modelo.

Consideraciones éticas:

Sesgo: Dos tipos de sesgo en los datos:

Sesgo humano: Percepciones sociales, estereotipos y factores historicos. 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 datso sesgados y las personas deciden / aprenden sobre estos datos sesgados. Estos nuevos datos son dados al modelo reforzando el sesgo presente en el mismo.

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

Puede resultar en la degradacion de la calidad de la informacion

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

Alucinacion no intrinseca: 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 compañias y tecnologias 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.