

keep coding

# MLOPs y LLMOPs

Grado en Ingeniería de Materiales (UPM)  
Master en Ingeniería Computacional y Matemáticas (UOC)

---

Undergraduate Research Assistant - MIT



Data Scientist - NewCastle University



Sr Machine Learning Engineer - Tecnicas Reunidas



TECNICAS REUNIDAS

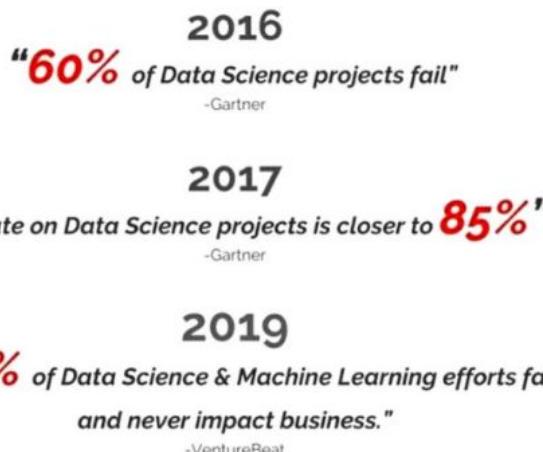
Sr Machine Learning Engineer - BASF



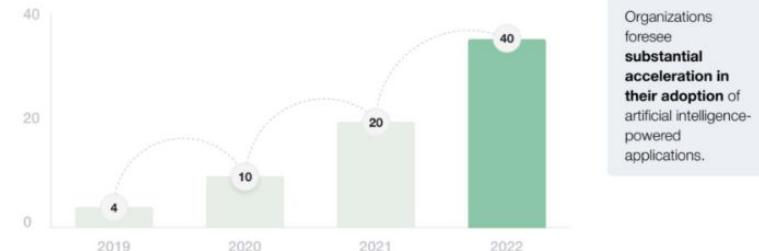
Lead Machine Learning Engineer - BASF

## La importancia del despliegue de algoritmos

- Generar valor
- Predicciones de manera automática
- Usuarios sin conocimiento pueden usar los modelos de ML.



**Figure 1 - Average Number of AI or ML Projects Deployed**  
Estimated Number of Projects Deployed (Mean)



- **MLOPs**
  - **Tipos de inferencia**
  - **Hardware para inferencia**
  - **Monitorización con MLFlow**
  - **Proveedores cloud**
  - **Google cloud (GCP)**
  - **FastAPI**
  - **LangSmith**
-

# La importancia del despliegue de algoritmos

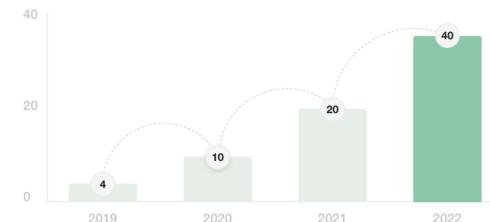
- Generar valor
- Predicciones de manera automática
- Usuarios sin conocimiento pueden usar los modelos de ML.

**2016**  
**"60%** of Data Science projects fail"  
-Gartner

**2017**  
"Fail rate on Data Science projects is closer to **85%**"  
-Gartner

**2019**  
"**87%** of Data Science & Machine Learning efforts fail  
and never impact business."  
-VentureBeat

**Figure 1 - Average Number of AI or ML Projects Deployed**  
Estimated Number of Projects Deployed (Mean)



# Modelos y entrenamiento

# Model Selection

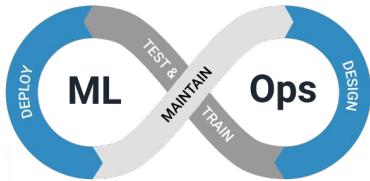
- Que no debemos hacer: No seguir las buzzword
- Que debemos hacer: Escoger el modelo más simple para nuestra solución, no el más llamativo.

**Be solution oriented.**

# Problemas de un ML Engineer en él day to day

- Querer probar el potencial de los modelos DL Sin mucha inversión.
- Es difícil conseguir un buen rendimiento sin inversión en tiempo y dinero en el etiquetado de dineros.

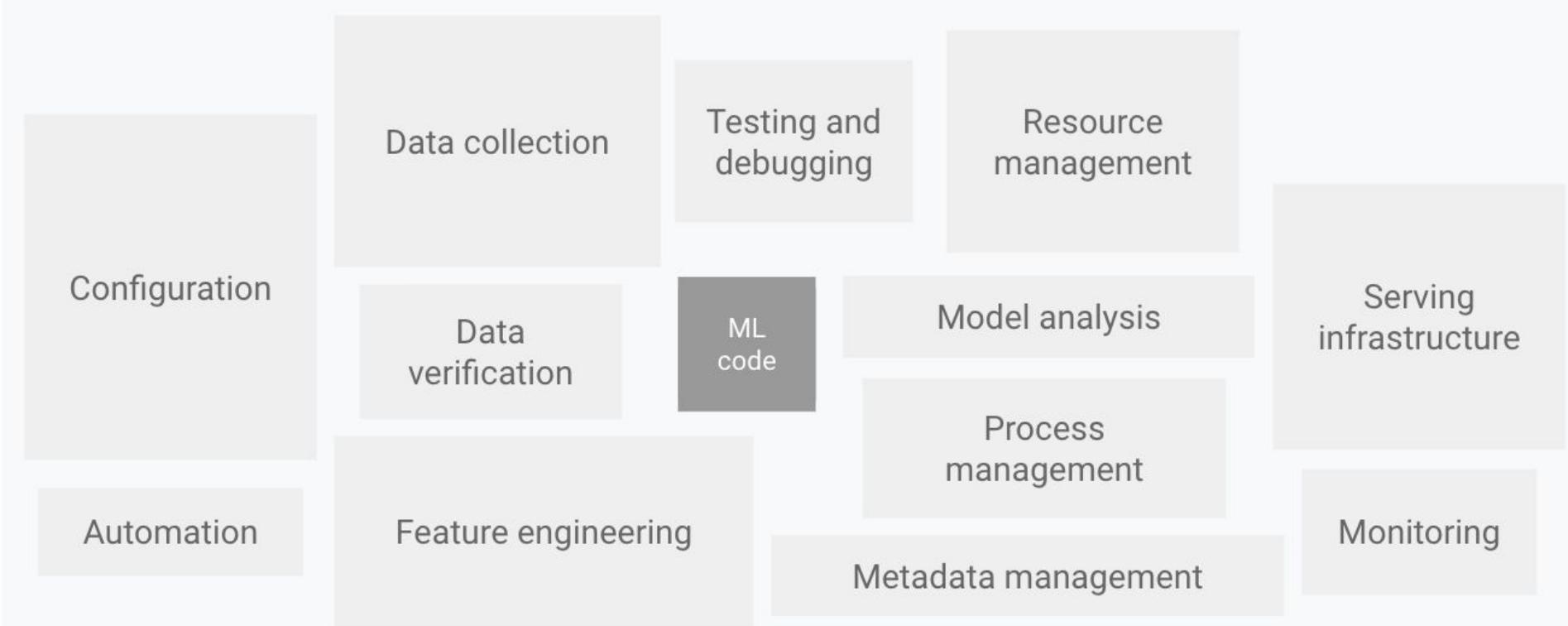
# MLOPs

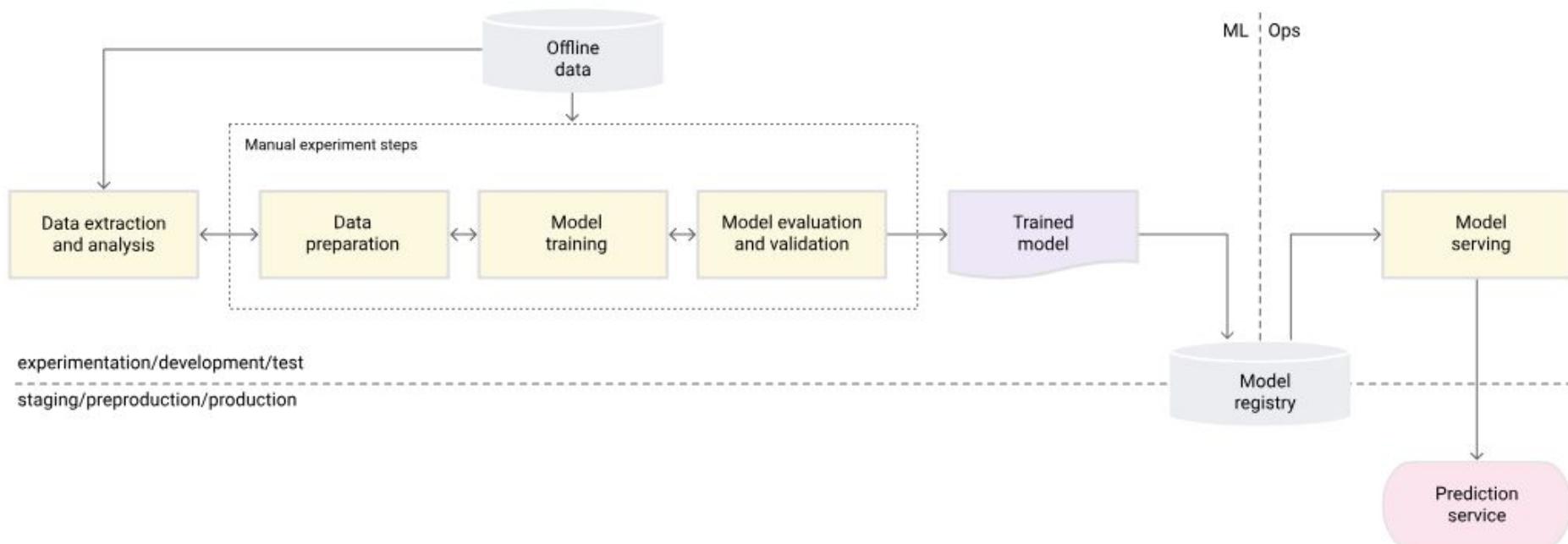


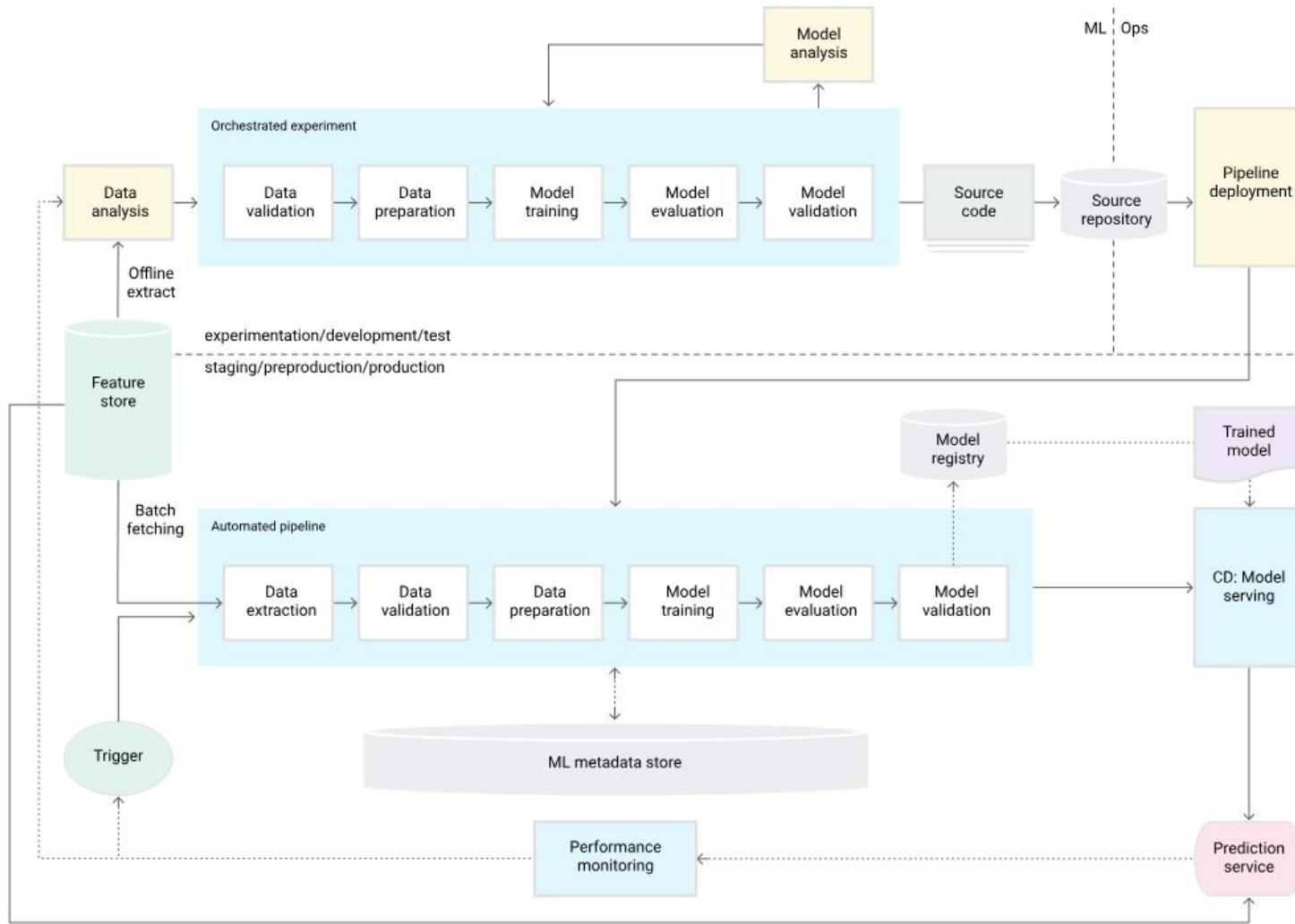
# MLOPs

## Machine Learning Operations

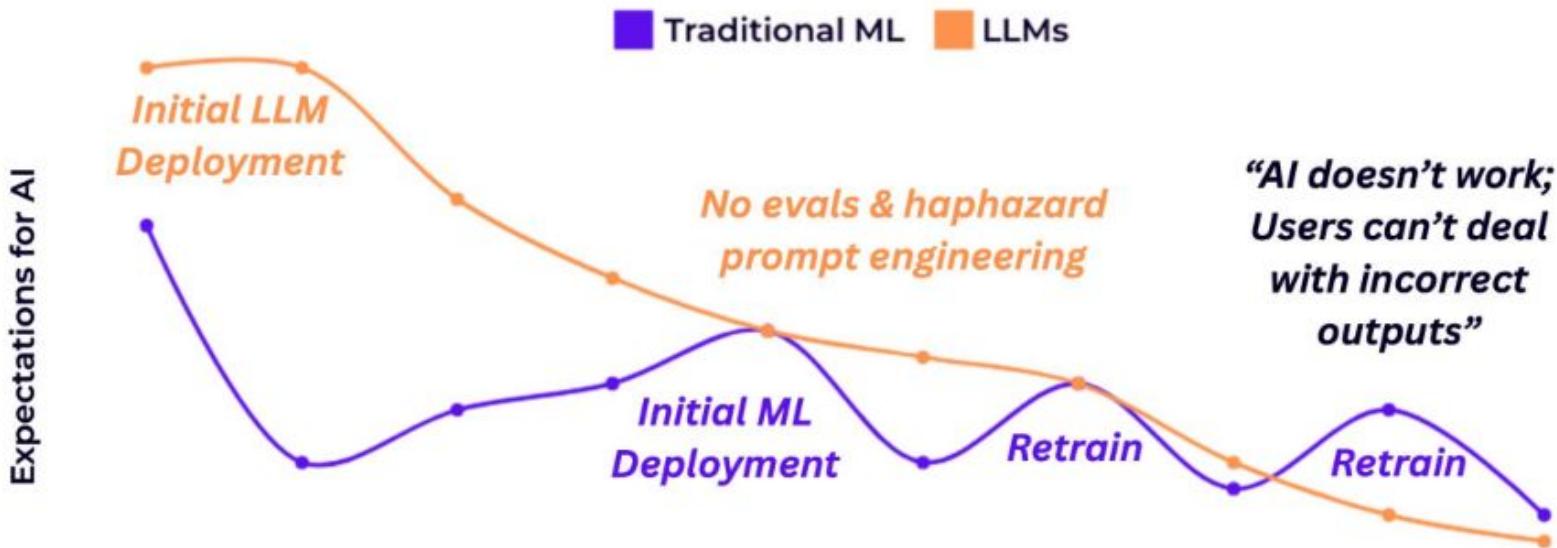
Unificar el Desarrollo (Dev) con las Operaciones (OPs) en sistemas de Machine Learning.







# Expectations While Building (Failed) AI Products



# Cuales son los objetivos del MLOps

- Automatización del despliegue:** Implementación automatizada y reproducible de los modelos en entornos de producción.
- Gestión del ciclo de vida:** Gestionar eficientemente el ciclo de vida completo de los modelos.
- Colaboración y comunicación:** Facilitar la colaboración efectiva entre equipos de desarrollo.
- Monitorización y mantenimiento continuo:** Establecer sistemas robustos de monitorización para evaluar el rendimiento de los modelos en tiempo real.



# Inferencia

# Tipos de Inferencia

- **Inferencia batch:** Realizan predicciones para un conjunto completo de datos de entrada de una sola vez. Es eficiente cuando se pueden procesar grandes cantidades de datos simultáneamente
- **Inferencia en tiempo real:** realizar predicciones a medida que llegan los nuevos datos, en lugar de esperar y procesarlos por lotes.

# Inferencia en Batch

Ventajas: Eficiencia computacional, optimización de recursos y facilidad de implementación.

Desventajas: Tiempo de respuesta, actualizaciones asincrónicas y menos adecuado para trabajar con datos en tiempo real.



# Inferencia en tiempo real

Ventajas: Respuestas inmediatas, adaptabilidad dinámica e interactividad

Desventajas: Requiere recursos inmediatos, menor eficiente para grandes volúmenes de datos, una planificación más compleja y un coste más elevado.



# Elección del tipo de inferencia

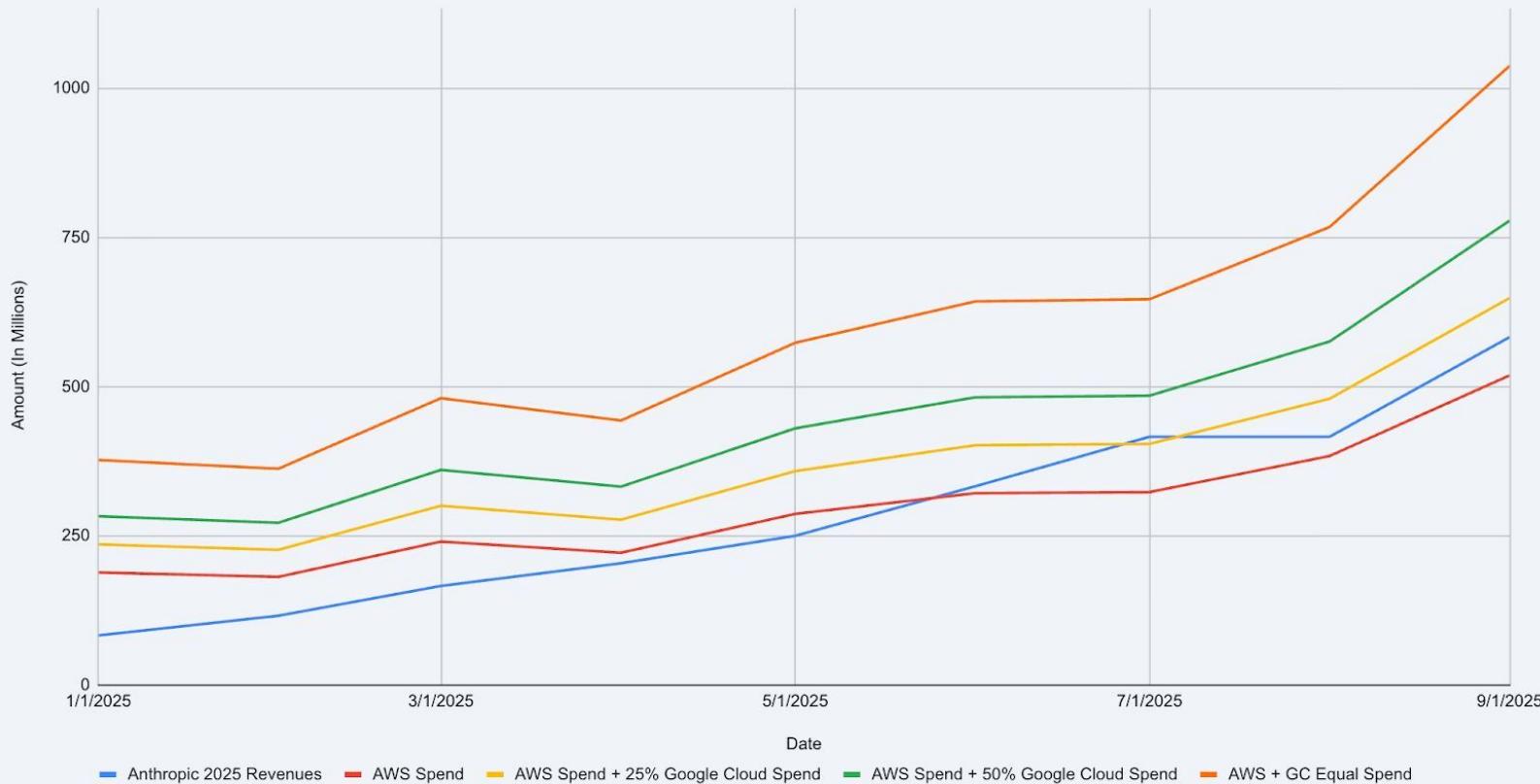
- Eficiencia Computacional
  - Tiempo de Respuesta
  - Escalabilidad
  - Costes Operativos
-

# Hardware

Unidades de procesamiento gráfico (GPUs): Las GPUs se utilizan mayoritariamente para ejecutar modelos de Deep Learning. Su uso tiene un coste muy elevado.

Unidades de procesamiento Central (CPUs): Mayor versatilidad que las GPUs pero menos eficiente a la hora de trabajar con modelos de Deep Learning. Pero cumple la mayoría de requisitos al trabajar con modelos kNN, SVM, XGBoost, etc.

## Anthropic's Revenue Versus Cloud Spend Through September 2025



# Proveedores cloud

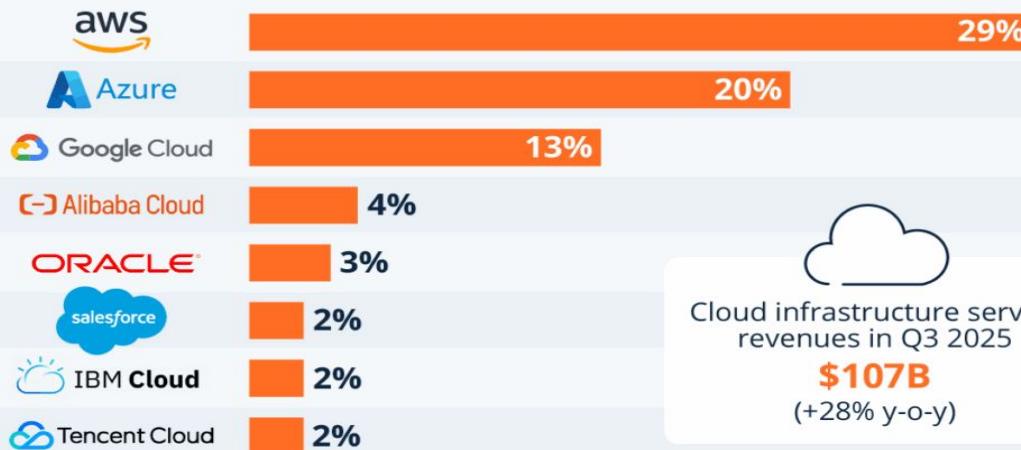


Google Cloud Platform



# AWS Stays Ahead as Cloud Market Accelerates

Worldwide market share of leading cloud infrastructure service providers in Q3 2025\*



Cloud infrastructure service  
revenues in Q3 2025

\$107B

(+28% y-o-y)

\* Includes platform as a service (PaaS) and infrastructure as a service (IaaS)  
as well as hosted private cloud services

Source: Synergy Research Group



# Pros y contras de usar servicios en cloud

Ventajas: Escalabilidad, pago por uso, acceso global, variedad de servicios, mantenimientos y actualizaciones automatizados, facilidad de implementación y facilidad en la recuperación de desastres.

Desventajas: Costos variables, dependencia del proveedor, latencia, limitaciones de personalización y conectividad a internet.



# FastAPI

# FastAPI



Es un framework de Python para la creación de APIs.

Sencillo, rápido y eficaz.

Permite integrar la inferencia de nuestros modelos de ML/DL en aplicaciones entre otras muchas cosas.

# Get, Put, Post, Delete

Get: Recuperamos datos.

Post: Envía datos para crear un nuevo recurso o entrada.

Put: Actualiza los datos existentes.

Delete: Elimina.

# keep coding



[www.keepcoding.io](http://www.keepcoding.io)



[cursos@keepcoding.io](mailto:cursos@keepcoding.io)



(+34) 916 33 1779