.UBAfiuba

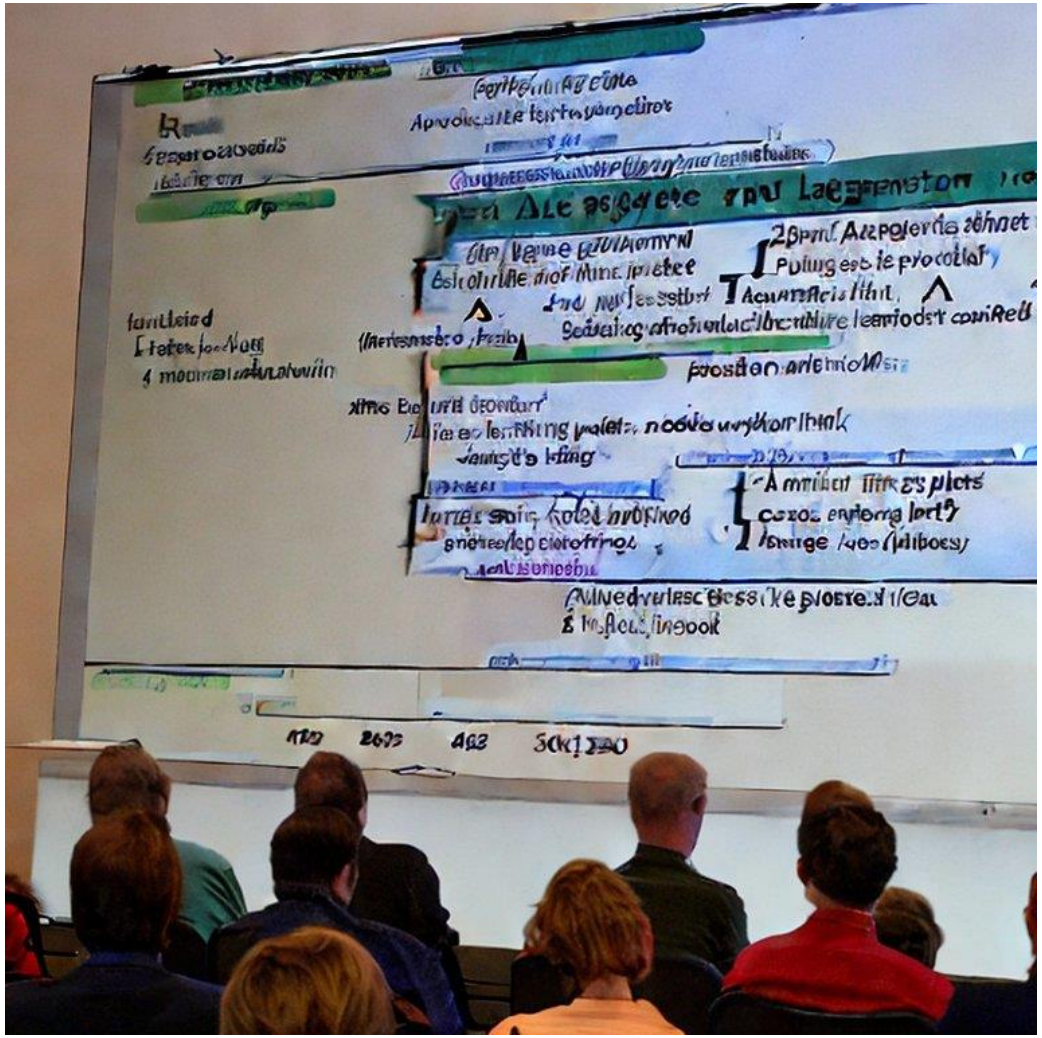
FACULTAD DE INGENIERÍA

LLMs e IA Generativa

Clase 8

Optimización de LLMs
GenIA multimodal

Mg. Ing. Ezequiel Guinsburg
ezequiel.guinsburg@gmail.com



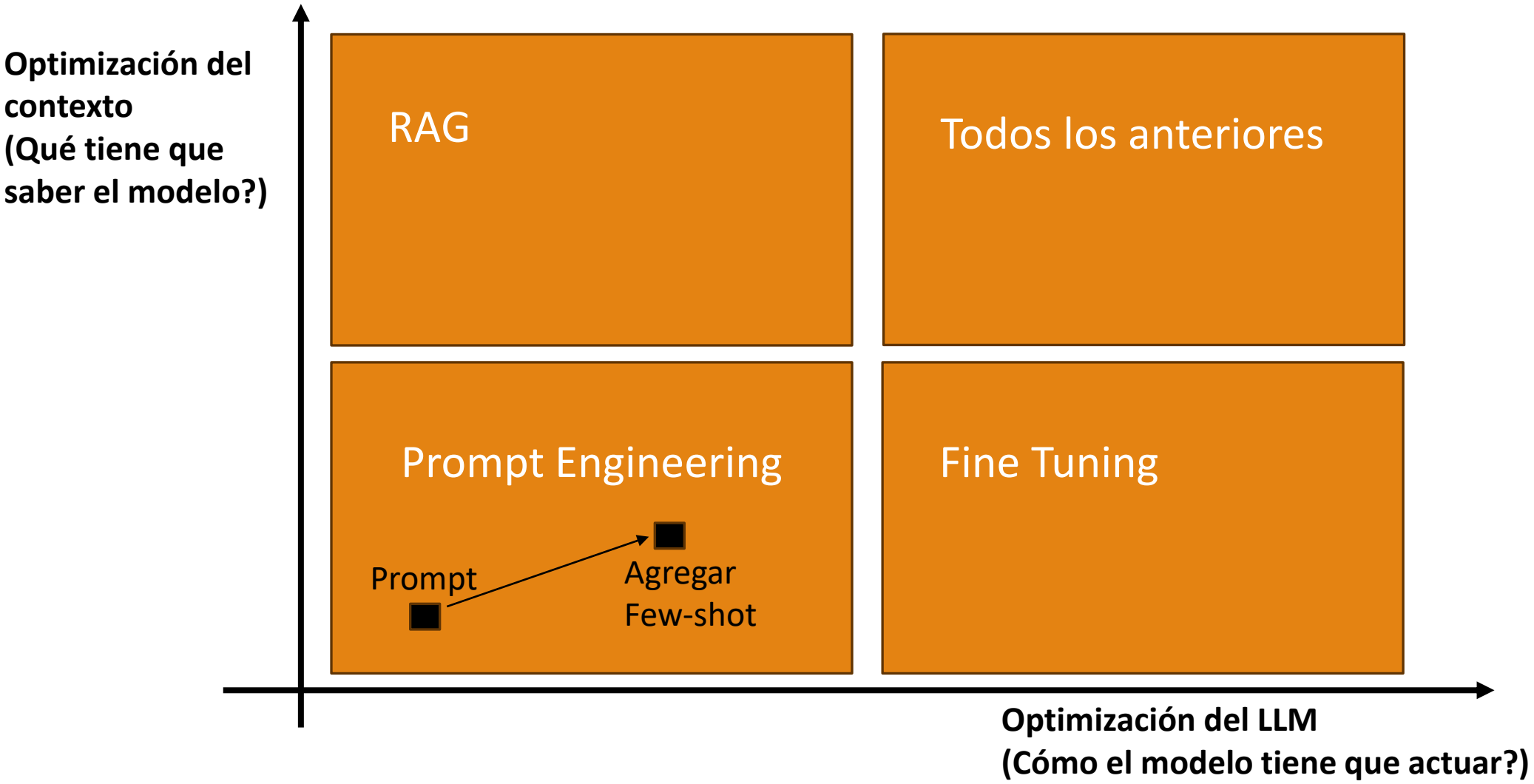
Referencias:

- Paper Visual Instruction Tuning - Haotian Liu
- Paper Hicherical Text-Conditional Image Generation with CLIP Latents - https://arxiv.org/pdf/2204.06125
- Denoising Diffusion Probabilistic Models - https://arxiv.org/pdf/2006.11239

Temas:

- El camino de la optimización de los LLMs.
- I.A. generativa multimodal.
  - Ejemplo teórico de modelo de texto-imagen.
  - Ejemplo práctico.
- Ejercicio 3.

Prompting Vs. RAG Vs. Fine tuning



Prompting Vs. RAG Vs. Fine tuning

Prompt Engineering

Bueno para:

- Testear y aprender rápidamente.
- Cuando se utiliza y se evalúa nos dá una idea de cómo optimizar

Malo para:

- Introducir nueva información.
- Replicar consistentemente un estilo o método (i.e. aprender un nuevo lenguaje de programación).
- Minimizar el uso de tokens.

Prompting Vs. RAG Vs. Fine tuning

Prompt Engineering

Claves:

- Instrucciones claras.
- Dar tiempo para pensar (step-by-step).
- Dividir un problema complejo en instrucciones simples.
- Incluir ejemplos y evidencia -> Few-shot examples.

Prompting Vs. RAG Vs. Fine tuning

RAG

Bueno para:

- Introducir nueva información para actualizar la del modelo.
- Reducir alucinaciones controlando el contenido

Malo para:

- Entendimiento de temas muy abarcativos y complejos.
- Enseñar al modelo a aprender nuevos lenguajes, formatos o estilos.
- Reducir el uso de tokens.

Prompting Vs. RAG Vs. Fine tuning

Fine tuning

Objetivo: Continuar el proceso de entrenamiento en un dataset de dominio menor para optimizar el modelo en una tarea específica.

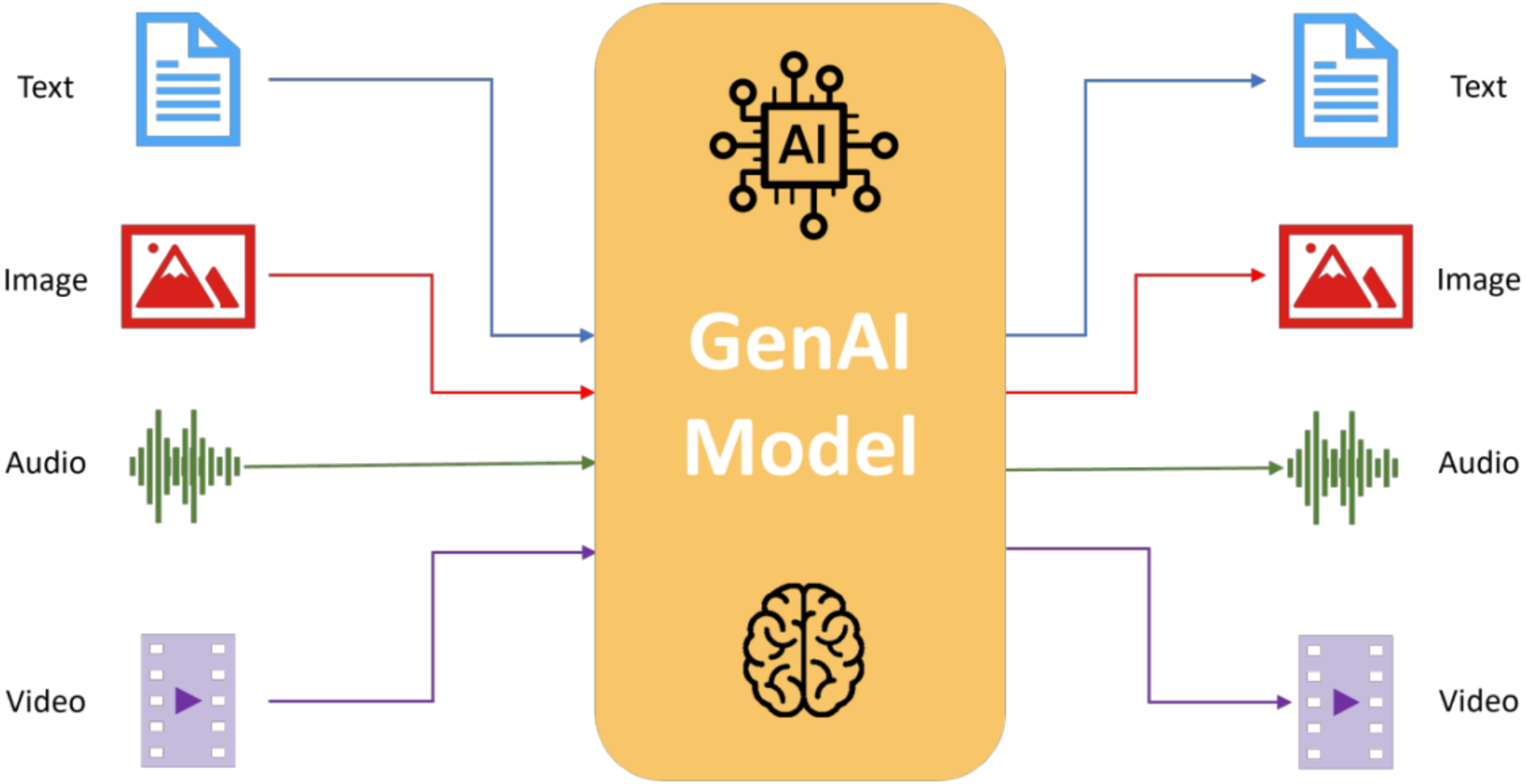
Bueno para:

- Mejorar la performance del modelo en un dominio específico.
- Enfatizar el conocimiento que ya existe en un modelo.
- Mejorar la eficiencia (reducción de la cantidad de tokens).

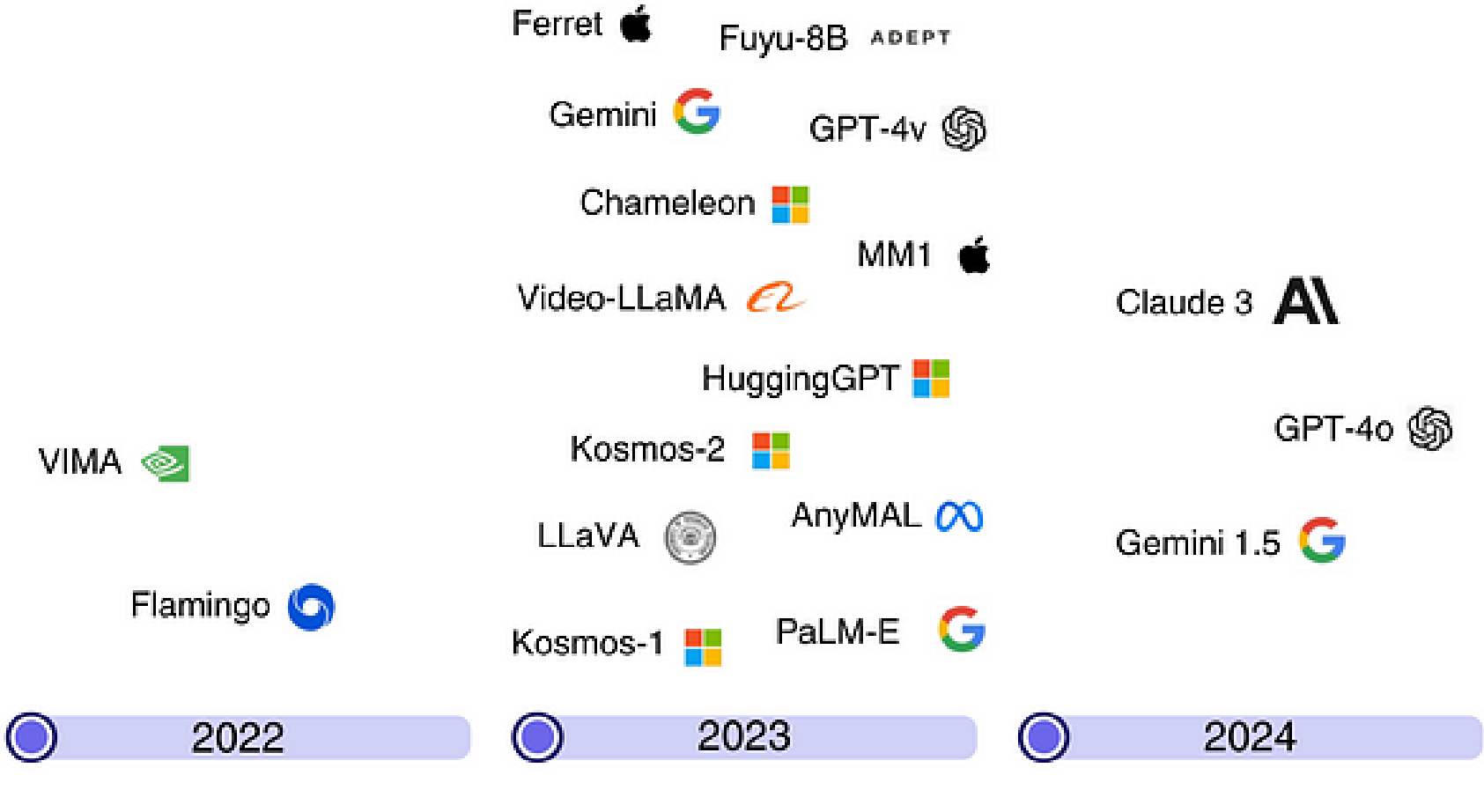
Malo para:

- Agregar nueva información al modelo.
- Iteración rápida en una optimización.

I.A. generativa multimodal



I.A. generativa multimodal

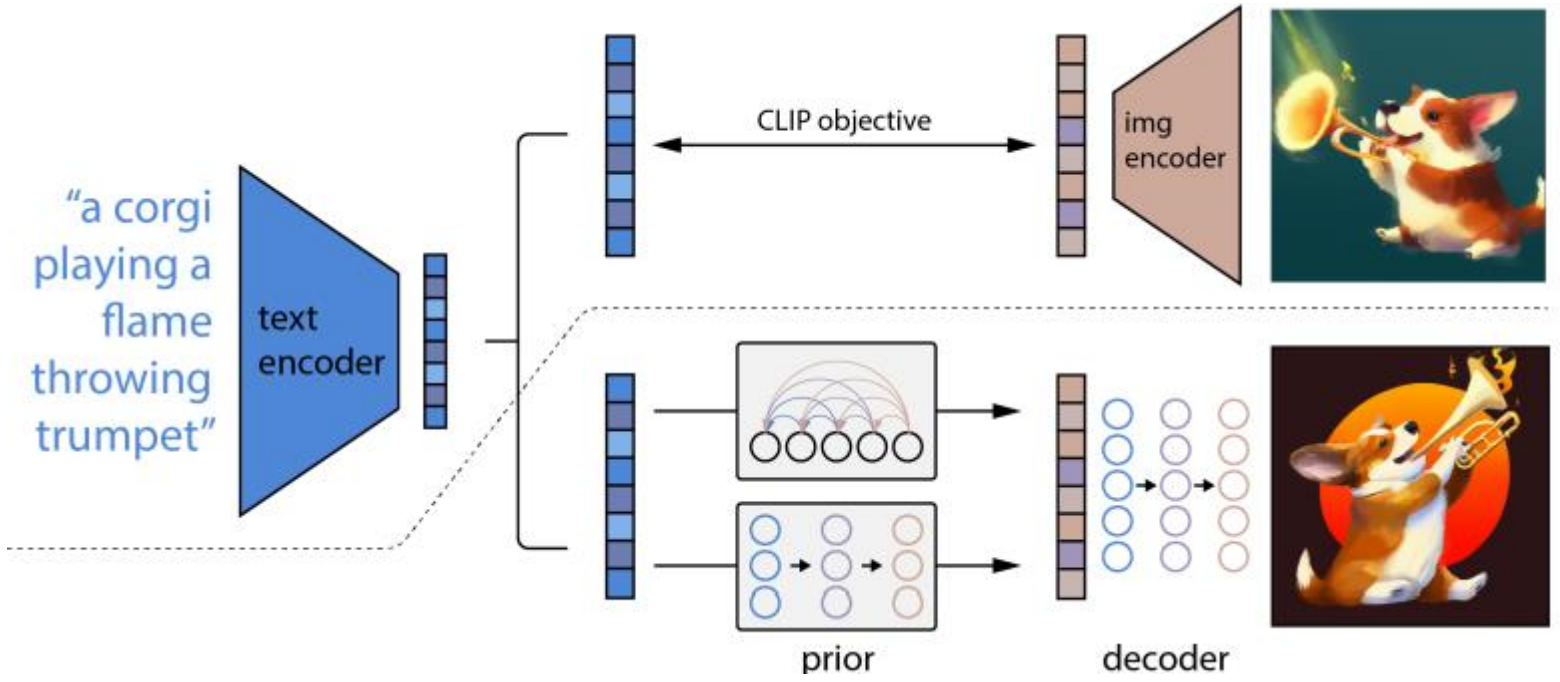


I.A. generativa multimodal

Generación de imágenes condicionada por embeddings de texto.

Encoder (CLIP) – Decoder (Difussion model). **[unCLIP]**

Se utiliza el estado latente del modelo CLIP para manipular los embeddings y sacar distintos tipos de imagen



I.A. generativa multimodal

Dataset de entrenamiento *(x, y)*

x = imágenes | ***y*** = descripciones

$\boldsymbol{z_i}$ = embeddings de imagen | $\boldsymbol{z_t}$ = embeddings de texto

Se utilizan dos componentes:

prior $\mathbf{P(z_i \mid y)}$ = produce los embeddings de imagenes CLIP

decoder $\mathbf{P(x \mid z_i, y)}$ = produce las imágenes *x* condicionadas a z_i e *y*

Combinando los dos componentes tenemos el modelo generativo de imágenes ***x*** dado las descripciones ***y***:

$$P(x|y) = P(x, z_i|y) = P(x|z_i, y)P(z_i|y)$$

I.A. generativa multimodal

unCLIP – Prueba de modificaciones en el espacio latente (“ataque tipográfico”

Granny Smith: 100%
iPod: 0%
Pizza: 0%

Granny Smith: 0.02%
iPod: 99.98%
Pizza: 0%

Granny Smith: 94.33%
iPod: 0%
Pizza: 5.66%

Trabajo comparativo de modelos:
https://distill.pub/2021/multimodal-neurons/

Repositorio del código para utilizar:
https://github.com/Stability-AI/stablediffusion

I.A. generativa multimodal

Ejemplo 1: (text2img)
https://colab.research.google.com/drive/1SXnX2zlx5oE3sItz65stMYvz0nYyfEvI?usp=sharing

Ejemplo 2 (Multimodal understanding)
https://huggingface.co/deepseek-ai/Janus-1.3B

Ejercicio en clase:

- **Consigna:** Implementar una notebook para generar imágenes a partir de texto (text2img) utilizando un modelo open source.
- **Recomendaciones** (no excluyentes):
  - Utilizar modelo deHuggingFace.
  - Utilizar una instancia de Colab con GPU.
  - No es relevante la performance del modelo.
- **Entregables:** Link a la notebook de acceso público. Puede ser al final de la clase (una entrega grupal) o individualmente hasta la fecha límite de entrega (Martes 10 de Diciembre).
- **Objetivo**: Aprender a buscar e instanciar modelos multimodales de manera simple y rápida.