

# ☐☐ APOLLO: SGD-Like Memory, AdamW-Level Performance

## ☐☐ Integrantes:

- ☐☐ Henry Mantilla
- ☐☐ Sebastián Solano

## ☐☐ Material de apoyo:

- ☐☐ **Diapositivas:** [Ver presentaciones](#)
  - ☐☐ **Paper:** [APOLLO: SGD-like Memory, AdamW-level Performance](#)
  - ☐☐ **Código externo:** [Repositorio de código](#)
- 

## ☐☐ Objetivos

☐☐ En esta sección se definen los objetivos de la sesión:

☐ Destacar cómo la alta demanda de memoria en optimizadores como AdamW limita el entrenamiento de grandes modelos.

☐ Explicar brevemente que APOLLO utiliza actualizaciones estructuradas (a nivel de canal o tensor) y proyecciones aleatorias de bajo rango para reducir el consumo de memoria.

☐ Mostrar que APOLLO (y su variante Mini) logran resultados comparables o superiores a AdamW, pero con un coste de memoria similar al del SGD.

---

## ☐☐ Resultados Esperados

□□ Esta sección describe de manera general lo que se espera obtener al final de la sesión:

□□ Aprender a analizar las métricas de entrenamiento como perplexity para evaluar la eficiencia y el ahorro de memoria obtenido con APOLLO.

□□ Contrastar las diferencias clave entre APOLLO, AdamW y SGD, identificando las ventajas y limitaciones de cada uno.

□□ Los participantes podrán contrastar la complejidad y los costos de realizar una SVD exacta frente a la utilización de métodos de proyección de bajo rango, y argumentar por qué una aproximación estructurada es suficiente para entrenar LLMs.

---

## □□ Referencias

□□ Esta sección recopila enlaces a recursos relevantes sobre procesamiento de imágenes:

□□ [Teoría](#)

□□ [Optimizing LLMs for Speed and Memory](#)

---

Revision #4

Created 10 April 2025 18:54:05 by Sebastian Solano

Updated 12 April 2025 16:32:50 by Henry