

Mixtral-8x7B: An Open Source Sparse Mixture of Experts



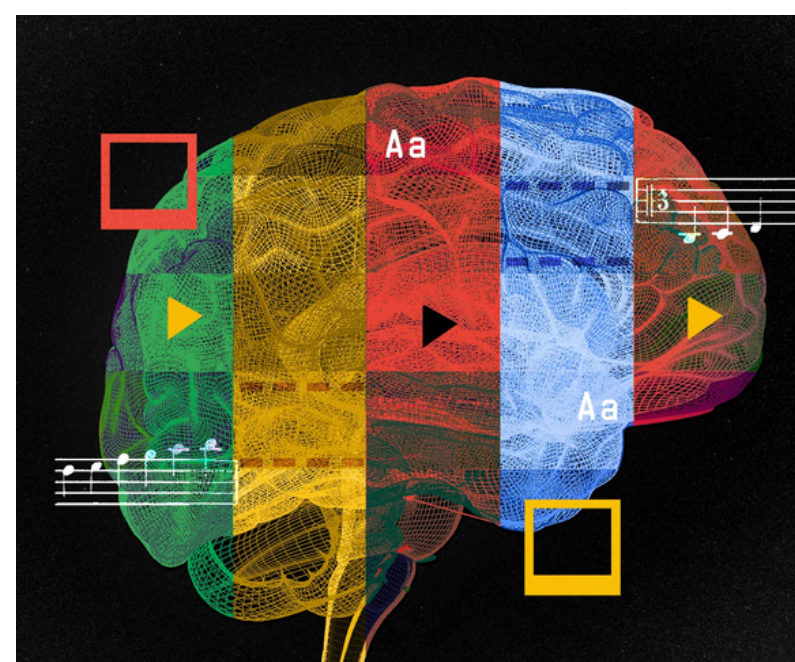
- Mn ba3d rumors li kherjo 3la l'architecture d'yal GPT-4, bli rah ghaliban a Mixture of Experts (MoE) fih 8 d'yal GPT-3-220B, 6 months mn ba3dha, Mistral AI releases **Mixtral-8x7B**, the strongest permissive open weights LLM l'ad sa3a.
- l'modèle fih 8 d'yal experts kula wahed fihom kiwzen around 7 billion params, pour chaque prédiction knsta3mlo ghi juj mn had experts, donc l'prédiction katqam 3lina b taman d'yal 14B params model o la qalitè d'la réponse 7sen mn Meta's 70B params LLaMA-2 model o mn OpenAI's GPT-3.5.

- <https://mistral.ai/news/mixtral-of-experts/>
- <https://huggingface.co/mistralai/Mixtral-8x7B-v0.1>

Gemini: Google's Remontada MMLLM

- Bach t3teq lmawqif o tghati 3la les critiques d'yal **Bard**, Google created **Google DeepMind** f avril 2023, labo wa7ed jem3at fih researchers mn juj d'akbar AI research labs, Google Brain o Deepmind. O db they announced **Gemini**, a **Multi Modal Large Language Model** (MMLLM).
- For context, a Large Language Model (LLM) kifahm ghi language (duh) wlkn kanqadro nzido lih ability bach yfham other types of inputs b7al images, to do so, kanlasqo lih a vision encoder (e.g. CLIP) o knkemlo training 3la image+text pairs bach nwaslo l **Feature Alignment**. Hadchi li tdar qbel f des projets b7al LLaVA.
- L'avantage ola l'moat" li ja bih Gemini huwa **Native Multimodality**, ola lqodra dyalo bach yfham o y'générer text, image, audio & video li machi ghi zadouha mn ba3d language pretraining, but rather dakhlouh fl'first stage dyl training, blackbox style.
- En même temps staghlat Google lforça bach daret ichhar l **TPUs (v5p)** fin t'entraîna had model.

- <https://deepmind.google/technologies/gemini>
- <https://blog.google/technology/ai/google-gemini-ai>



MLX: Deep Learning Framework for Apple Silicon



- Mo2akharan bdina knchoufo interest kbir f kifach tkhadem latest generative AI models (LLMs o Stable Diffusion) locally, o bdabt f les ordinateurs Mac li fihom des puces **M1/M2/M3 series**, 7it 3arfin bli had chips kikhadmo les logiciels d'design b sohoulà, donc kayn l compute power.
- Mouchkil huwa les solutions kola wahda o les avantages dyalha; **Pytorch MPS Backend** ki3tik flexibility bach tkhadem aya model mktoub b pytorch bsohoula wlkn tqil, o **GGML/GGUF** ki3tik maximum dl'hardware optimization mais s3ib tzid lih chi model ola train'ih.
- **Apple** fehmat o l'blan o released **mlx**, framework msayba bch tjma3 had jouj dl features, inspired mn **Pytorch**, interface familiar o optimized for Apple Silicon.
- Blan nadi 7it db Apple ghadi tqdar t synchroniser développement dl'hardware o software bwahed l'feedback loop, kifma daret Google b Tensorflow/JAX o TPU development.

- <https://github.com/ml-explore/mlx>

official partner

