

Qu'est-ce que la "taille" d'un modèle ?

 Meta/Llama 3.3 70B

SEMI-OUVERT

70 MDS DE PARAMÈTRES

SORTIE 12/2024

Impact énergétique de la discussion

70 milliards param.
taille du modèle

x

260 tokens
taille du texte

=

 **1.98** Wh
énergie conso.

Ce qui correspond à :



1.98g
CO₂ émis



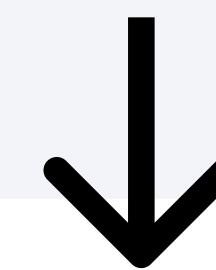
38min
ampoule LED



3min
vidéos en ligne

Quantité de **paramètres**

70B = 70 milliards de paramètres



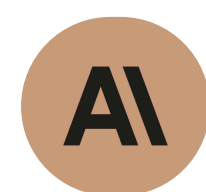
Meta/Llama 3.3 70B

Un paramètre est comme une "synapse" artificiel. Plus il y a de paramètres, plus (en théorie) un modèle a de capacité d'apprendre. Néanmoins, ça ne veut pas dire qu'un modèle plus grand sera toujours plus "intelligent".

Qu'est-ce que la "taille" d'un modèle ?



Les duels de l'IA
par compar:IA



Anthropic/Claude 3.5 Sonnet

→ **300** milliards de **paramètres** (env) → **XL**



Meta/Llama 3.3 70B

→ **70** milliards de **paramètres** → **L**



Mistral/Mistral Small

→ **24** milliards de **paramètres** → **M**



Google/Gemma 3 4b

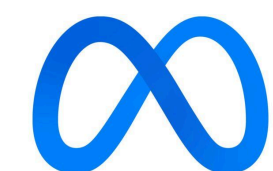
→ **4** milliards de **paramètres** → **XS**

La distillation

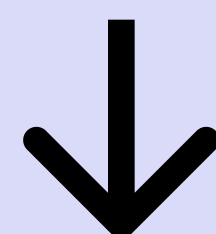


Les duels de l'IA
par compar:IA

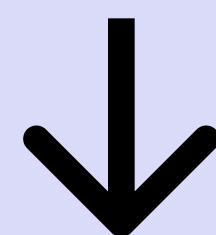
Famille Llama 3



Meta/Llama 3.1 405B



Meta/Llama 3.3 70B



Meta/Llama 3.3 8B

La distillation d'IA consiste à **entraîner un petit modèle en imitant un grand pour le rendre plus rapide et efficace.**

