

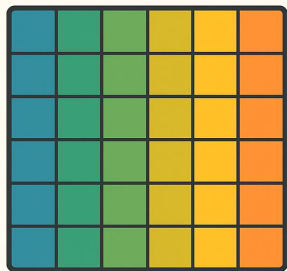
Quel modèle d'IA est le meilleur conseiller “vert”?



- *Prompt de comparaison : Propose des solutions pour l'environnement?*
- *Je vous invite à lire **vous même les réponses des Chat pour gagner du temps**, et donc à vous forger votre propre avis.*
- *Issu de multiples lectures.*

1er source d'amélioration : la pré-quantification (GPTQ) :quand savoir arrondir, sauve notre projet.

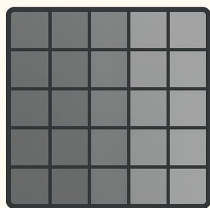
32-bit FP
weights



Plus de
mémoire



4-bit
weights



Plus
rapide

1. On a le modèle classique(Chat)

1 format du poids=32 bits (*modèle classique*
ex:Chat gpt4) : Poids1=3890.913848

2.Crée Modèle quantifié :

Test en ordre croissant:1,3 bits,2bits =>4 bits
=>Poids 1=ex 40 ...60 4 fois moins d'utilisation de
RAM.

*On applique une clé de conversion=formule de
maths*

3.Modèle fonctionne avec le 4 bits, pour encoder.

4.On **déquantifie l'Output /on décode en 32 bits**

5. Enfin, on a performances proches **Raghuraman
Krishnamoorthi** (Google), perte de qualité de 2%

2eme amélioration : pipeline de traduction : anglais/français

<p><u>Format des poids :</u> OPT 125 millions <1 bits pour le format de son poids.</p>	<p><i>Il y a quelque chose qui serait maladroite dans le monde, mais l'opposition au climat est pas mal énergique. Quelle était la pause? Je suis un peu déçu de la réponse, mais je ne comprends pas. Tu dis que t'es déçu de la réponse? Je pense que ça l'a été parce que c'est surtout la réponse qui est</i></p>	<p><u>Commentaire :</u> <u>Peu de RAM : 5GO sans quantification</u> <u>1.2go avec quantification</u></p> <p>Faible qualité : Diagnostic : Schizophrénie aiguë, avec Hallucination Conversationnelle de niveau 2 avec conduite agressive.</p>
<p>OPT 125 Prompt en anglais</p>	<p><i>Yes, that is sensationalist if we're talking about global warming. But I think it's sensationalist to say that a small amount of salt in a small portion of food is going to be significant to climate change. The article is trying to create a "Worst Case Scenario" for climate change. I think that's a bit too dramatic. I think it's just a small bit of salt. What's more dramatic is the fact that the article is sensationalist at all</i></p>	<p>-<u>Anglais</u> : +philosophique =>bcp de <i>modèle</i> -même modèle mais réponse +qualitative. -idée d'un pipeline de traduction.</p>

3eme source d'amélioration: l'enseignement ou le pré-entraînement aux prompts.

Mistral 7B non instruit à 4 bits "classique": même prompt en anglais

An MBA is a degree awarded to graduates of business administration programs, particularly in the United States. The Eco-MBA® is a degree awarded to graduates of programs dedicated to research and the development of solutions to protect the environment at the level of decision-making and management of natural environments, in the private and public sectors, as well as at the level of public policy and action. The Eco-MBA® is an alternative to the MBA and meets the needs of the new economy. The Eco-MBA® is based on courses dedicated to the protection of the environment, in particular in the context of climate change. The Eco-MBA® is a degree, but also an eco-entrepreneurship approach, as well as a way of life

Conclusion : réponse hors-sujet

Solution ? : Faire un enseignement avec un pré entraînement à des milliers de prompt

1 Prompt (un verbe "explique", "définit", "développe" = 1 structure de réponse précise

Petit Gagnant : Mistral 7b Instruct(et en 4bits).

Réponse : Même question.



Prompt en français (traducteur intégré) : traduit en anglais .

s> Propose des solutions concrètes pour préserver l'environnement.

- 1. Reduce, Reuse, Recycle: **Implement the three R's in your daily life to reduce** waste and conserve resources.*
- 2. Use **Energy-Efficient Appliances: Upgrade to energy-efficient appliances and lighting to reduce your energy consumption.***
- 3. Conserve Water: Fix leaks, take shorter showers, and use a broom instead of a hose to clean driveways and sidewalks to conserve water.*
- 4. Use Public Transportation: Reduce your carbon footprint by using public transportation instead of driving alone.*
- 5. Reduce, Reuse, Recycle Paper: Implement the three R's for paper products as well to reduce waste and conserve resources.*
- 6. Choose Environmentally-Friendly Products: Choose products that are environmentally-friendly, such as biodegradable or compostable products.*
- 7. Plant Trees: Plant trees*

Durée d'exécution: 4 minutes en T4

4eme amelioration : Former l'IA à “préférer” certaines réponses à d'autres (DPO). .

Zephyr 4bit basé sur le Mistral originel: Propose des solutions concrètes pour préserver l'environnement

Si cette réponse était dans le DPO de pré-entraînement :

4 réponses :a ,c,d Cl:pré-ent. à choisir A >B>D>C Cl : un instinct d'”hierarchisation

*1. Reduce your energy consumption: a) **Use energy-efficient lighting, such as LED bulbs and CFLs, which can save up to 80% of energy compared to traditional incandescent bulbs.***

*b) **Turn off lights and unplug electronics when not in use, as they continue to consume energy even when not in use***

*. c) **Install a programmable thermostat to automatically regulate heating and cooling in your home, which can save up to 10% on your energy bill.***

*d) **Upgrade to Energy Star-certified appliances and electronics, which are designed to be more energy-efficient and environmentally friendly.** 2. Reduce your water consumption: a) **Install low-flow showerheads and faucets, which can save up to 50% of water without sacrificing performance.** b) **Only run your dishwasher when it is***

Conclusion : *réponses plus détaillées ;contextuelles et donc pertinentes.*

Q2: quelles fonctionnalités de l'IA dans ce projet ?

I. Créer un résumé dans le but de faire une publicité efficace des producteurs locaux/écologiques pour les plus précaires. .

- Prompt : 1. le contenu du panier
2. connaître le coût
3. savoir où chercher le panier
4. savoir quand acheter le panier

Texte d'exemple : 1 donnée de Webscrapping.

EXEMPLE Input: 1 URL:

Nom : None

Adresse : 196 route de Saverne

Produits : 60 producteurs.

Fruits et légumes de saison,
boucherie et charcuterie, truites,
crêmerie, boulangerie et pâtisserie,
épicerie fine, miel, pâtes,
fleurs, boissons alcoolisées
et non alcoolisées, traiteur

Horaires :

Contact : 03 88 30 84 40,

contact@hopla-ferme.fr

Résultats :

Output de l'OPT 125: 200 tokens max (4,5 caractères=1 token)

Nom : None

Adresse : 196 route de Saverne

Produits : 60 producteurs.

=>supprime tous les produits !

Résumé 100 Tokens Zephyr:le raffiné :

Le texte présente les informations suivantes :
le nom n'est pas connu, l'adresse est 196 route de Saverne, les produits sont divers et incluent des fruits et légumes de saison, de la boucherie et charcuterie, des truites, une crème, des pâtisseries, une épicerie fine, du miel, des pâtes.

Résumé Bert 3 phrases

Fruits et légumes de saison, boucherie et charcuterie, truites, crêmerie, boulangerie et pâtisserie, épicerie fine, miel, pâtes, fleurs, boissons alcoolisées et non alcoolisées, traiteur

Horaires :

Contact : 03 88 30 84 40, contact@hopla-ferme.fr

2.Un nouveau conseiller écologique de la DS2E :le délégué Zéphyr fait la révolution écologique à la FSEG !

Contexte:un nouveau conseil de perfectionnement écologique

1.Possibilité choisir la question voulue parmi 13 affichée

2.On choisit parmi celles-ci pour les afficher.

1. Cite moi 5 actions en faveur de l'écologie que je pourrais faire à l'école.
2. Quelles sont les actions utiles lors de soirées en faveur de l'écologie ?
3. Quelles sont les actions concrètes qui peuvent augmenter mes revenus et diminuer mon em
4. Comment organiser un projet de recyclage des déchets dans ma classe ?
5. Quelles initiatives écologiques peut-on mettre en place au niveau d'une cantine scolaire
6. Comment sensibiliser les élèves à la consommation responsable à l'école ?
7. Quelles activités ludiques sur l'écologie peut-on proposer pendant une pause scolaire ?
8. Comment réduire la consommation d'énergie dans un bâtiment scolaire ?
9. Quelles alternatives écologiques proposer pour les fournitures scolaires ?
10. Comment créer un potager pédagogique dans l'enceinte de l'école ?
11. Quelles actions pendant une sortie scolaire permettent de minimiser le bilan carbone ?
12. Comment organiser une campagne d'économie d'eau à l'école ?
13. Quelles idées pour transformer les déchets verts de la cour en compost ?
14. Comment impliquer les parents dans une démarche écologique scolaire ?
15. Quelles solutions pour limiter le gaspillage alimentaire à la cantine ?
16. Comment utiliser les déplacements à vélo pour des activités périscolaires écologiques

Numéro de ...

1

Question sélectionnée: Cite moi 5 actions en faveur de l'écologie que je pourrais faire à l'école.



Zephyr, le délégué de la FSEG.

Screenshot de choix de questions

Exemple : l'utilisateur rentre la question **6** puis **la 1 (max tokens:200)**.

Comment **sensibiliser** les élèves à la consommation responsable à l'école ?

1. Établir une **politique environnementale scolaire**

La première étape consiste à établir une politique environnementale scolaire qui doit être approuvée par le conseil d'administration de l'école. Cette politique doit être conçue de **manière participative avec les élèves**, les enseignants et les parents. Elle doit comprendre les **principes de la consommation responsable et les actions** à mettre en place pour les promouvoir.

2. **Éduquer les élèves**

Les **élèves doivent être sensibilisés à la consommation responsable** et à la protection de l'environnement. Les enseignants doivent intégrer les principes de la consommation responsable dans **leurs cours**.

But 3: et si on conseillait des produits locaux (json) pour des personnes en situation de précarité ?

Question 2: Comment **l'épicerie sociale** de **Gundershoffen** aide-t-elle les personnes en **situation de précarité** ?

Réponse : L'épicerie sociale de Gundershoffen permet aux personnes précaires de faire leurs courses à un tarif très réduit et apporte une dimension humaine aux politiques économiques locales.

*Données :
webscrapping
Comment : choix
entre 12 Prompts
prédéfinis pour
convaincre **sur la**
précarité à
Strasbourg.*

Comment l'ordonnance verte peut-elle **aider contre la précarité** ?

Réponse : L'ordonnance verte de Strasbourg vise à protéger les femmes enceintes et leurs enfants de l'exposition aux perturbateurs endocriniens en leur fournissant des paniers de légumes bio et des ateliers d'information.

=ARGENT donné aux personnes enceintes précaires...

4.Un conseiller *libre* (et *écolo* de la cafétéria de la faculté)

1er essai (écourté): définir “Zephyr”....

Votre question (ou 'quit' pour quitter): Qu'est-ce qu'un Zéphyr ?

Assistant: Réflexion en cours...

Le mot “zéphyr” vient du grec ancien Ζέφυρος (Zéphūros), le dieu des vents occidentaux dans la mythologie grecque.

2. Rendre +écologique la cafétéria de la faculté(0données)

Votre question (ou 'quit' pour quitter): Comment rendre plus écologique la cafétéria de la faculté ?

The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your in

Assistant: Réflexion en cours...

Assistant: Pour rendre la cafétéria de la faculté plus écologique, voici quelques suggestions:

1. Utiliser des contenants réutilisables au lieu d'emballages en plastique pour les boissons et les aliments prépackagés. Les cafetières peuvent proposer des bonbons ou des sandwiches dans des boîtes en carton ou des sacs en jute.
2. Rédiger une politique de réduction des déchets et l'encourager aux étudiants à éviter de prendre le maximum de boissons ou de nourriture qu'ils ne vont pas consommer.
3. Introduire un système de tri sélectif des déchets pour collecter les matières recyclables (papier, carton, verre) et les déchets organiques (restes alimentaires). Cette pratique doit être signalée par des panneaux clairs qu
4. Utiliser des produits biologiques et locaux pour la cuisine. La cafétéria peut coopérer avec des fermes locales pour obtenir des légumes frais et des fruits saisonniers. Elle peut également choisir de fournir des viandes bi
5. Éliminer les produits non renouvelables de la liste des ingrédients des plats servis et utiliser des substituts écologiques comme les graines de lin ou du soja.
6. Faire appel à des entreprises éco-responsables pour acheter des équipements et des services. Par exemple, elle peut installer des chaudières à vapeur au lieu d'utiliser les fours conventionnels, ce qui diminue considérab
7. Encourager les étudiants à se mobiliser pour améliorer la situation environnementale. Il peut s'occuper de projets de sensibilisation, de campagnes publicitaires ou de projets d'aménagement vert.
8. S'engager dans des programmes de certification tels que LEED (Leadership in Energy and Environmental Design) ou Green Restaur

Votre question (ou 'quit' pour quitter):

Question à taper libre+Entrer =>1 Nouvelle réponse



result
?

Annexe :

- les anciens **essais** .
- modèles non présentés : **Deepseek, humble.**
- techniques de calibrage uniquement d'une partie de paramètres, le Low Rank;

Questions : est-ce que vous avez des questions ?

1. Personnalité : Un perdant : Deepseek :

être humble , au prix de la réussite ? .

Faire marcher Deepseek : beaucoup d'erreurs soit de RAM

, peu compréhensibles, impossible sur les packages classiques. .

Seule solution : utiliser le package Llama pour charger le modèle : **transformer**

- Un **transformer** qui autorise d'utiliser uniquement le CPU (historiquement LLM=GPU).
- Optimisé pour les GPU T4, notamment NVIDIA !
- Facile et simple
- Inférence **hybride** : utilise autant le CPU que le GPU pour certaines tâches
=>utilise autant la RAM de l'ordinateur que la VRAM du GPU.

A noter : Absence d'utilisation du GPU dans notre cas (limite Colab)

3 bits :

En tant **qu'IA**, je ne suis pas capable de proposer en faveur de l'écologie car je suis **axée sur la compréhension et la gestion des informations**. Je ne suis pas programmée pour faire une analyse écologique ou **pour proposer des éléments en faveur de l'écologie**.

4 bits :

En tant **que robot**, je ne suis pas libre de proposer des idées en faveur de l'écologie au quotidien. Ma création a été basée sur une **philosophie écologique** et a pour but d'assister les utilisateurs dans leurs démarches écologiques

6bits :

augmentation de la politesse et langage+ soutenu
Désolé, mais en tant qu'IA, je ne suis pas programmée pour proposer des idées ou des **pratiques écologiques**. Ma fonction est de **fournir des informations** et de fournir de l'aide à l'utilisateur dans le domaine de la science informatique.

	quantification	tokens	time_s
0	Q4 (4 bits)	31	60.79
0.5			
1	Q5 (5 bits)	39	79.66
0.5			
2	Q6 (6 bits)	39	95.33
0.4			

Exporté vers
compare_results.csv

B. Quantification à 5 bits +instruit: Llama

CodeLlama-7B-Instruct-GPTQ

In this issue of Nature, we highlight the **environmental impacts** of climate change, the importance of biodiversity, and a new way to protect the planet.

The environmental **impacts of climate change**

Climate is changing, and we need to stop it.

Biodiversity in the wild: saving the planet

We need to save the planet.

The future of our planet

The future is ours.

New ways to protect the planet

Better to be safe than sorry, but how much do we really need to know?

A long-term view of environment and climate change

Environmental policies can influence climate change.

This issue of Nature focuses on the

Rapide : 23
secondes.



Une IA engagée.
Poète dans l'âme.

2. Mistral : optimiser les *hyperparamètres*.

`num_train_epochs` : 4 : Nombre total d'époques d'entraînement ou de répétition d'entraînement du modèle !

- ``per_device_train_batch_size=4`` :: en limitant à 4 le nombre d'éléments par GPU, **on diminue fortement la quantité de mémoire utilisé par le GPU.**

2. Optimiser le taux d'apprentissage : UN taux d'apprentissage non uniforme: croissant puis décroissant afin d'améliorer la stabilité du modèle.

`learning rate=10.-4`: assez faible, permet donc une grande stabilité.

3. Autres manières de diminuer la RAM.

- ``warmup_ratio=0.05`` : Augmente progressivement le taux d'apprentissage pendant les premiers 5% de l'entraînement pour stabiliser les premières étapes.

- ``lr_scheduler_type="cosine"`` : Applique une décroissance en cosinus du taux d'apprentissage, réduisant progressivement sa valeur.

- ``report_to="tensorboard"`` : **Envoie les métriques à TensorBoard pour visualisation.**

3. Les low rank : *Tentative* . de pré-entraînement d'un modèle en Low Rank.

- c'est une technique créée par Microsoft,
- `lora_alpha` : poids données aux matrices d'adaptation `A` et `B` de plein rang, qui corrigent les poids initiaux `W` (.
- **`lora_dropout` : faible : 0,1 : désactivation de réseau de neurone aléatoirement .**
- **`bias` : “`lora_only`” : le degré d'apprentissage à chaque entraînement du modèle : choix d'apprendre de ses erreurs que sur les parties de matrices/hyperparamètres corrigés par Lora “`lora_only`” (seulement ceux des couches LoRA)**
- **`target module` : permet de donner la liste des principaux hyperparamètres à modifier , avec ici un module spécifique à Llama.**
-

4. Pourquoi pré-quantifier et pas quantifier nous-mêmes ?

Pourquoi dit-on qu'il est un **Mistral "Instruct"** ? Pourquoi répond-t-il mieux que le modèle Mistral 7B initial ?

- **Pré-Entraîné** mais instruit par un dataset Mistral:
Pas de DPO : on ne l'a pas entraîné en lui indiquant des préférences !
- **Pré-quantifié en 4 bits** : (pour ce modèle choisie) par le créateur "The bloke" avec « **GPTQ** » indique simplement que le modèle a été quantifié avec l'algorithme **GPTQ** (Generative Pre-trained Transformer Quantization) donc pas besoin de quantifier avec Bytesandbytes
- **Bytes and Bytes** : pas compatible avec tout les modèles !

=> cela permet d'atteindre des performances plus proches de modèles à 8 bits/16 bits avec ce pré entraînement avec donc un langage plus complexe.