MINISTÈRE
DES ARMÉES

Liberté
Égalité
Fraternité

AMIAD
Agence ministérielle pour l'intelligence artificielle de défense

Challenge fine-tuning EvalLLM 2025

Marseille, 30 juin 2025

MINISTÈRE DES ARMÉES
Liberté
Égalité
Fraternité

AMIAD
Agence ministérielle pour l'intelligence artificielle de défense

Contexte et objectifs

Fine-tuning pour le domaine de la défense

- complexe (acronymes, domaine, mélange de métiers…)
- partage d'approches et de techniques relatives au challenge d'adaptation
- partage d'approches pour le traitement de donnée
- mesure de l'impact des données

Calendrier

- annonce : 23 mars
- inscription / récupération des données : jusqu'au 30 mai
- envoi d'exemples d'évaluation : 10 mai
- envoi des modèles fine-tunés : 3 juin

MINISTÈRE DES ARMÉES
Liberté
Égalité
Fraternité

AMIAD
Agence ministérielle pour l'intelligence artificielle de défense

Eléments mis à disposition

Données

- Données du domaine défense, interne et externes au MinArm
- différents formats (pdf, html, markdown, texte)
- Sources intenres : différents services
- Sources externes : wikipédia et sites web
- **Possibilité de collecter et d'introduire d'autres sources de données.**

Deux modèles proposés :

- Mistral-7b-Instruct-v0.3
- Mistral-Small-24B-Instruct-2501

Évaluations

Principes, par ordre d'importance

1 - Non-régression

- S'assurer qu'il n'y a pas de régression des performances sur les capacités initiales du modèle, notamment sur les tâches d'évaluations standards (MMLU, FrenchBench, etc)

2 – Adaptation

- Disposer d'un modèle adapté au vocabulaire métier du Ministère des armées.

3 – Factualité

- Éviter les hallucinations

MINISTÈRE DES ARMÉES
Liberté
Égalité
Fraternité

AMIAD
Agence ministérielle pour l'intelligence artificielle de défense

Évaluations

Tâches et métriques : non régression

MMLU : QCM couvrant plus de 57 sujets académiques pour tester les capacités multitâches des LLM (16 000 questions).

FrenchBench : Benchmark francophone regroupant plusieurs jeux de données pour mesurer les capacités de compréhension, de génération, de grammaire, de vocabulaire et de connaissance culturelles des LLM.

MINISTÈRE DES ARMÉES
Liberté
Égalité
Fraternité

AMIAD
Agence ministérielle pour l'intelligence artificielle de défense

Évaluations

Tâches et métriques : adaptation

QCM Minarm :

- nombreux sujets en lien avec le Ministère des armées : culture de la défense, égalité homme/femme, institutions françaises, SSI…
- Métriques : accuracy
- *Une personne quitte les fonctions pour lesquelles elle était habilitée. Qu'en est-il de son habilitation ? -> a) Son habilitation reste valable tant qu'elle reste dans le même organisme, b) L'habilitation cesse immédiatement, c) L'habilitation étant personnelle, elle reste valable pour un autre poste*

Résumé et titrage

- Métriques utilisées : rouge1, bertscore

QA gold : Ensemble de questions/réponses courtes incontournables en Français sur le Ministère des armées.

Métriques utilisées : bertscore, regex

- *C'est quoi l'IGESA ? -> L'IGESA est l'institution de gestion sociale des armées, sa mission est de gérer les établissements sociaux et médico-sociaux du ministère des armées.*

Évaluations

Tâches et métriques : factualité

Opex perplexity

- Ensemble d'articles en lien avec les actualités militaires et de défense.
- Métriques utilisées : perplexité
- *L'Afrique du Sud renonce à l'A400M*
- *La Marine nationale sonne le branlebas pour son recrutement*

Hallucination

- Questions délibérément ambiguës ou incohérentes visant à évaluer la robustesse d'un LLM à reconnaître des erreurs factuelles ou logiques.
- Métriques utilisées : NER et Regex.
- *C'est quoi la vitesse maximale du SCAF ?*
- *En quelle année y aura t'il une base spatiale à Brest ?*

Résultats 7B

Classement par systèmes (3 premiers)

1er Orange – Ouest France (o)

2nd Orange – Ouest France (o)

3ème Orange – Ouest France (f)

4ème CEA (f)

5ème CEA (f)

6ème Airbus (o)

Résultats très serrés !

Catégorie données fermées (f)

1er Orange – Ouest France

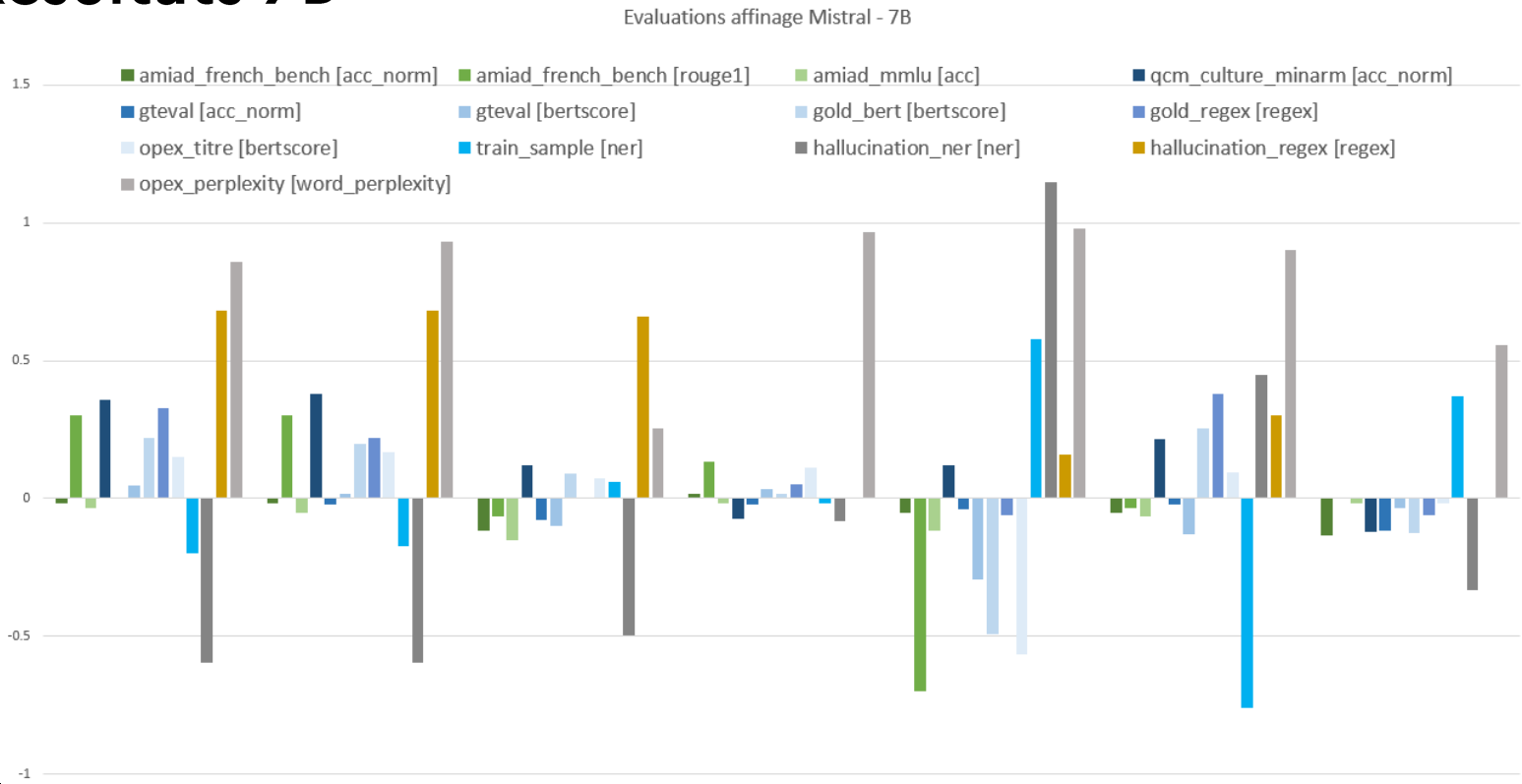
2nd CEA

Catégorie données ouvertes (o)

1er Orange – Ouest France

2nd Airbus

Résultats 7B



MINISTÈRE DES ARMÉES
Liberté Égalité Fraternité

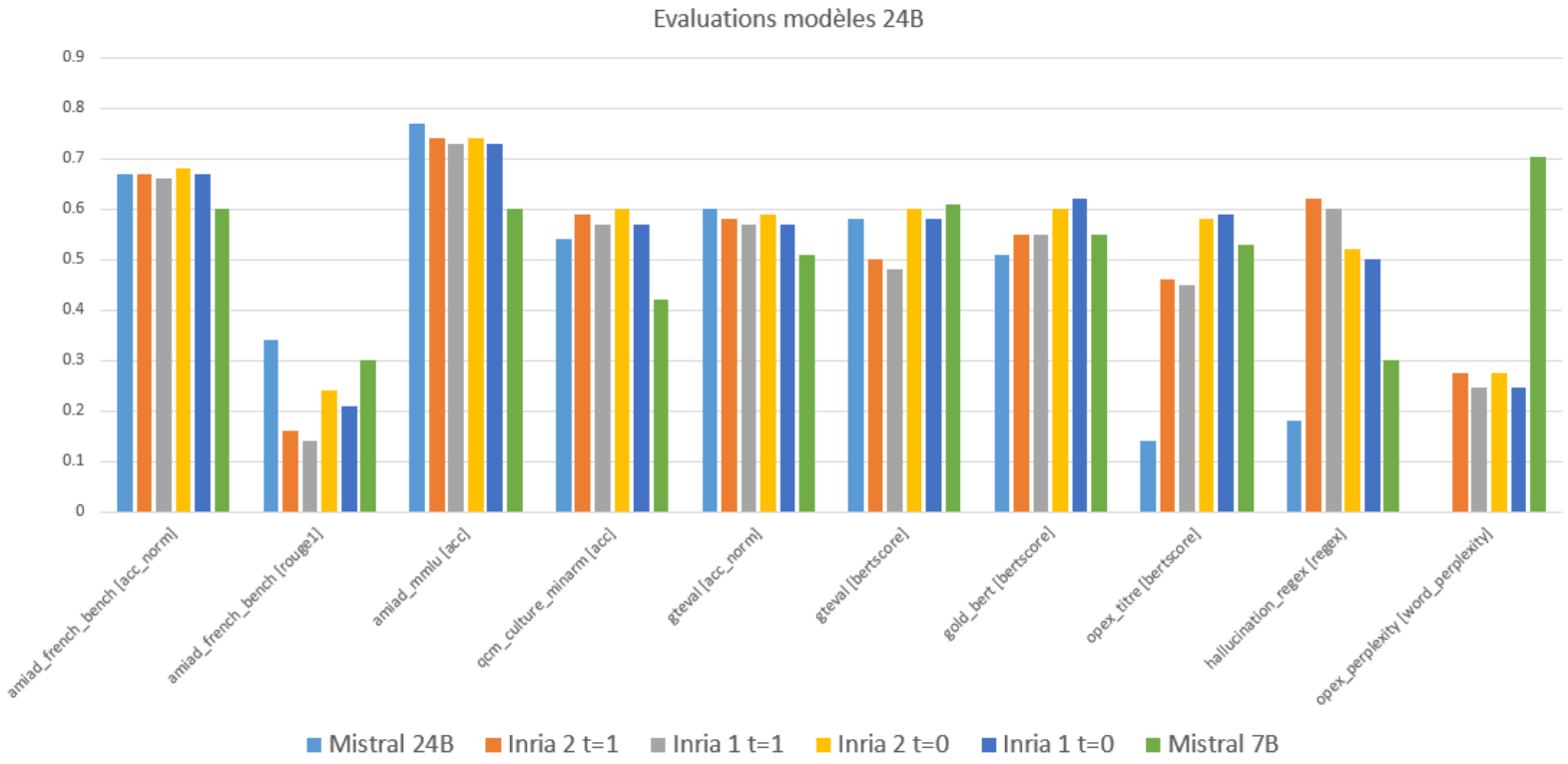
AMIAD
Agence ministérielle pour l'intelligence artificielle de défense

Résultats 24B

1 seul participant

Inria – Links

Mistral 7B pour comparaison



MINISTÈRE DES ARMÉES
Liberté Égalité Fraternité

AMIAD
Agence ministérielle pour l'intelligence artificielle de défense

Bilans carbone

en kg CO2

Attention :
disparité d'outils