

Réunion 27 mai 2026

- Administratif et organisationnel
- Comparer deux versions du modèle
- Librairie pour modifier le modèle
- Considérations sur les perturbations calculées ce WE
- Pipeline fonctionnelle pour le challenge
- Slides pour le séminaire CA

Administratif et organisationnel

Évaluer et comparer le modèle et le modèle révisé avec MechIR

- si on ne peut récupérer les gold : générer les gold des questions générées avec un LLM (déconnecté d'internet)?

Modifications sur le modèle : librairie

`transformer lens` (utilisée en interne de MechIR) [doc](#) permet:

- modifier le comportement d'un ou plusieurs noeuds ciblés
- passer les données et récupérer (loss, logit)
- garder en mémoire la trace d'exécution d'une donnée (ou un dataset entier)
- ...

=> prise en main à prévoir

Perturbations du 1% des données considérés

- pas de perturbation remarquable
- des perturbations choisies sur la base du dataset total
- un dataset très clairsemé (50 questions, 30390 documents pour 38564 paires)
- recalcul des statistiques:
 - nouveaux mots : opération, armée, militaire, sécurité, contrôle, ...

=> Peut-il y avoir des phénomènes de compensation entre les noeuds sensibles à une perturbation ?

=> Doit-on accorder du sens à la présence d'une double bosse? Correspondent-elles aux paires (non) pertinentes? Quid de l'absence de double bosse dans les dernières expés?

Vérification du modèle

- fonctionne bien sur mirev
- 6 questions en moins de 5mn

Remarques:

- Peut refuser de répondre (Q1)
- Réponse en anglais
- Mentionne le contexte (Q1, Q5)

Challenge

- mail bien reçu ce matin
- calcul des coûts carbone à faire
- open-source ... git à faire
- rapport de participation
- Cohere : clé API d'essai : 10req/mn

Slides Séminaire CA

- "les bonnes questions"
- Que faut-il mettre dans un *protocole expérimental*? Quel niveau de précision est attendu? à l'écrit et à l'oral