



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Performances de l'intelligence artificielle générative dans la correction de RSCA

Hadrien PRELAUD – Maître de conférences associé de MG à l'UPEC

Anas TAHA – Professeur associé de MG à l'UPEC



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

**3 AU 5 DÉCEMBRE 2025**

# Conflits/Liens d'intérêt

## Conflits d'intérêt

Pas de financement pour ce travail. Les auteurs déclarent n'avoir aucun conflit d'intérêt.

## Liens d'intérêt

Hadrien PRELAUD

- Maître de conférences associé de médecine générale à l'UPEC
- Médecin généraliste à la MSPU de Sucy-en-Brie
- Secrétaire général de la CPTS Sucy/Noisieu

Anas TAHA

- Professeur des universités associé de médecine générale à l'UPEC
- Médecin généraliste et gérant à la MSPU de Sucy-en-Brie
- Président de la CPTS de Sucy/Noisieu
- Coordinateur de l'ouvrage RSCA



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# De l'intérêt d'évaluer les LLM en pédagogie médicale

Usage croissant des LLM par les enseignants et les étudiants.

Applications diverses en pédagogie médicale

RSCA = outil d'évaluation partagé du DES-MG dans tous les DMG

Les LLM pourraient accompagner les étudiants et les enseignants

Pas de données sur l'évaluation des RSCA par les LLM



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

**3 AU 5 DÉCEMBRE 2025**

# Objectif

*Évaluer les performances de GPT-4o et Mistral-7B dans la correction de RSCA en les comparant à des experts de MG selon la grille nationale.*





# Méthode

Étude mixte convergente (quantitatif + qualitatif de manière synchrone)

Vingt-et-un RSCA de l'ouvrage national

Comparaison GPT-4o, Mistral-7B avec un groupe de 5 MG experts en pédagogie

Utilisation invite standardisée construite par itération et simulant une correction

Notation de 13 à 14 critères selon le RSCA : Pertinent, Améliorable ou Insuffisant



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Méthode

## Comparaison :

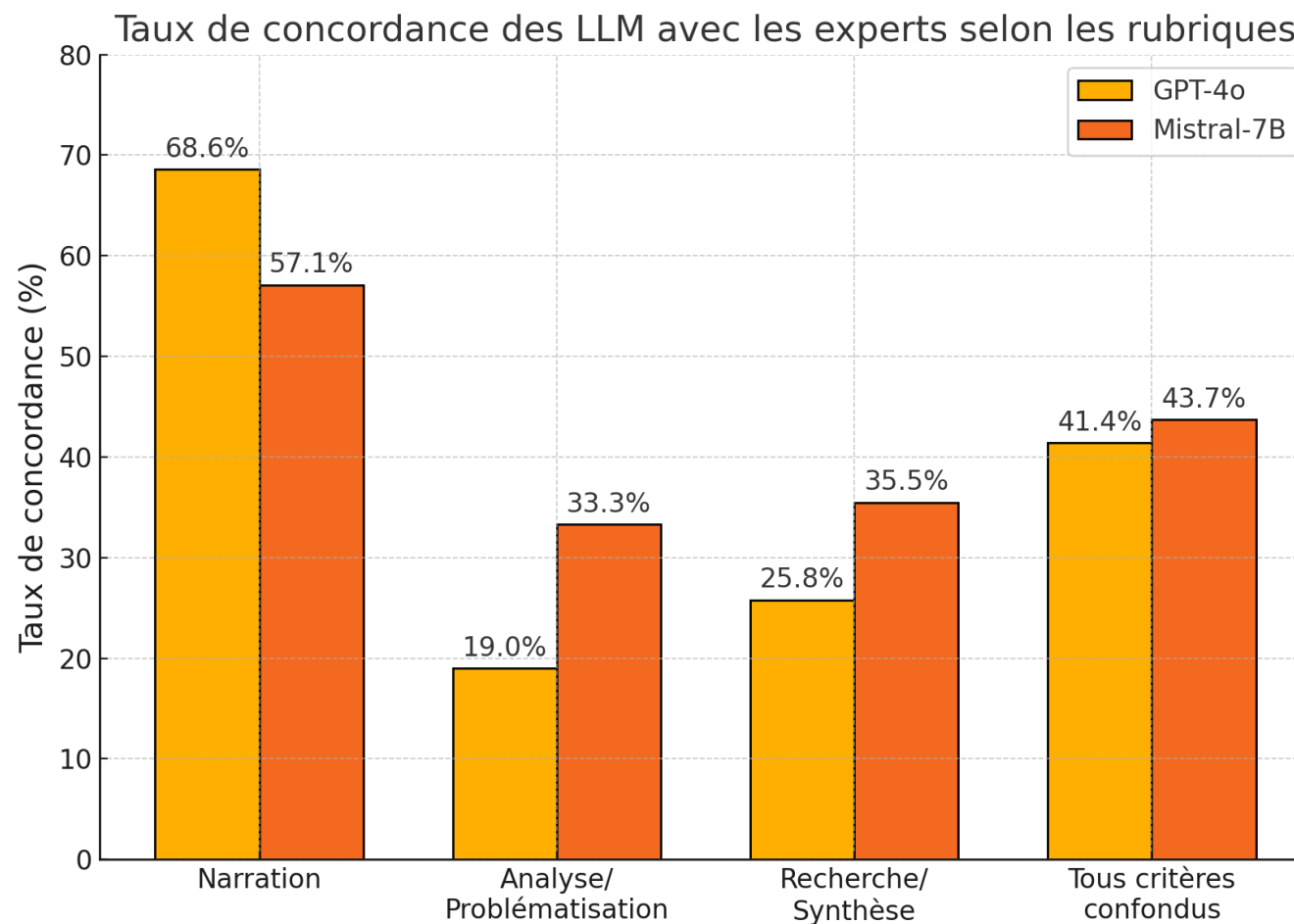
Experts vs GPT-4o

Experts vs Mistral-7B

**Versant quantitatif** : description des taux de concordance (%) et analyse Kappa

**Versant qualitatif** : analyse thématique comparative des commentaires

# Comparaison des notations LLM/experts





MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Comparaison en fonction de la notation des experts

| Mention des experts | GPT-4o | Mistral-7B |
|---------------------|--------|------------|
| P (n = 33)          | 84,8 % | 69,7 %     |
| A (n = 36)          | 22,2 % | 41,7 %     |
| I (n = 18)          | 0 %    | 0 %        |





MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Comparaison LLM/experts

## Rubrique / Comparaison

### GPT-4o ↔ Experts

### Mistral-7B ↔ Experts

## Récit / Narration

● 0,29

● 0,14

## Analyse / Problématisation

● -0,21

● -0,01

## Recherche / Synthèse

● -0,02

● 0,11



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Illustration qualitative

| Critère                        | LLM (GPT-4o, Mistral-7B)          | Experts humains MG                        |
|--------------------------------|-----------------------------------|---|
| Structuration du retour        | Rigueur formelle, suivi de grille | Adaptée au contexte                       |
| Analyse réflexive              | Absente                           | Présente, posture réflexive personnalisée |
| Prise en compte des implicites | Non                               | Oui, implicites intégrés                  |
| Evaluation globale             | Faible                            | Relie les parties entre elles             |
| Critique des sources           | Non                               | Oui, analyse de pertinence                |
| Posture pédagogique            | Standardisée, binaire             | Individualisée, engageante                |



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Des limites techniques et méthodologiques

- Les LLM structurent les corrections mais restent limités :
  - Concordance faible avec les experts
  - Absence d'évaluation critique, posture indulgente
  - Limites techniques et conceptuelles
- Les LLM testés ont probablement été peu exposés à des données de soins premiers dans leur **entraînement initial**
- Limites des LLM aussi liée à **l'invite de commande utilisée**



# Perspectives

Usage hybride sous supervision.

Entraîner LLM sur corpus de soins premiers.

Utilisation aide à l'écriture et à la recherche

Réflexion à avoir sur l'impact cognitif des usages des LLM





MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Forces/limites

## Forces

- Sujet innovant
- Méthode mixte
- Comparaison avec 5 experts MG utilisant la grille nationale
- Applicabilité pour la formation des internes

## Limites

- Échantillon modeste (21 RSCA)
- Pas de consensus inter-experts préalable
- Evaluation LLM via prompt standardisé uniquement
- Impact sur apprentissage des internes non évalué
- Dépendance aux versions des LLM



MEETT Centre de Conventions  
& Congrès de

**Toulouse**

3 AU 5 DÉCEMBRE 2025

# Conclusion

- Merci de votre attention
- Coordonnées de contact
  - Hadrien PRELAUD : [hadrien.prelaud@u-pec.fr](mailto:hadrien.prelaud@u-pec.fr)
  - Anas TAHA : [anas.taha@u-pec.fr](mailto:anas.taha@u-pec.fr)
- Avez-vous des questions ?

