IA et amélioration de sujets d'examen en Pharmacologie

Alexandre Gérard
AHU de Pharmacologie Clinique

Alexandre Destere
PHU de Pharmacologie Clinique

Université Côte d'Azur – CHU de Nice

EVALUATING AND LEVERAGING LARGE LANGUAGE MODELS IN CLINICAL PHARMACOLOGY AND THERAPEUTICS ASSESSMENT: FROM EXAM TAKERS TO EXAM SHAPERS

Alexandre O. Gérard[1], Diane Merino[1], Marc Labriffe[2,3], Fanny Rocher[1], Delphine Viard[1], Laurence Zemori[1], Thibaud Lavrut[1], Erik M. Donker[4,5], Joost D. Piët[4,5], Jean-Paul Fournier[6], Milou-Daniel Drici[1] and Alexandre Destere[1,7]

[1] Université Côte d'Azur Medical Centre, Department of Clinical Pharmacology and Pharmacovigilance Center, Nice, France.
[2] Pharmacology and Transplantation, INSERM U1248, Université de Limoges, Limoges, France.
[3] Department of Pharmacology, Toxicology and Pharmacovigilance, University Hospital of Limoges, Limoges, France.
[4] Research and Expertise Centre in Pharmacotherapy Education (RECIPE), De Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands.
[5] Department of Internal Medicine, Unit Pharmacotherapy, Amsterdam UMC, Vrije Universiteit, De Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands.
[6] Université Côte d'Azur, Faculty of Medicine of Nice, Medical Simulation Center, Nice, France.
[7] Université Côte d'Azur, Inria, CNRS, Laboratoire J.A. Dieudonné, Maasai team, Nice, France.

Révisions mineures

BJCP
British Journal of Clinical Pharmacology

Introduction

Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish Medical Final Examination

Maciej Rosoł [1], Jakub S Gąsior [2], Jonasz Łaba [3], Kacper Korzeniewski [3], Marcel Młyńczak [3]

Affiliations + expand

PMID: 37993519 PMCID: PMC10665355 DOI: 10.1038/s41598-023-46995-z

> JMIR Med Educ. 2023 Feb 8:9:e45312. doi: 10.2196/45312.

How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment

Aidan Gilson [1 2], Conrad W Safranek [1], Thomas Huang [2], Vimig Socrates [1 3], Ling Chi [1], Richard Andrew Taylor [# 1 2], David Chartash [# 1 4]

> JAMA Ophthalmol. 2023 Jun 1;141(6):589-597. doi: 10.1001/jamaophthalmol.2023.1144.

Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment

Andrew Mihalache [1], Marko M Popovic [2], Rajeev H Muni [2 3]

Affiliations + expand

PMID: 37103928 PMCID: PMC10141269 DOI: 10.1001/jamaophthalmol.2023.1144

Performance of ChatGPT on Nephrology Test Questions

Jing Miao [1], Charat Thongprayoon, Oscar A Garcia Valencia, Pajaree Krisanapan, Mohammad S Sheikh, Paul W Davis, Poemlarp Mekraksakit, Maria Gonzalez Suarez, Iasmina M Craici, Wisit Cheungpasitporn

Affiliations + expand

PMID: 37851468 PMCID: PMC10843340 (available on 2025-01-01)

DOI: 10.2215/CJN.0000000000000330

ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models

Namkee Oh [1], Gyu-Seong Choi [1], Woo Yong Lee [1]

Méthodes

- **Algorithmes** utilisés : **ChatGPT 4 omni** et **Gemini Advanced**
- **Questions** :
  - **Partiels de fin d'année** Fac de Médecine (Nice)
    - L3 (n = 39)
    - M1 (n = 38)
    - M2 (n =40)
  - **European Prescribing Exam** (Nice & Limoges) :
    - Session 1 (n = 47) ➔ parties « Knowledge » (n = 36) & « Skills » (n = 11)
    - Session 2 (n = 47) ➔ parties « Knowledge » (n = 36) & « Skills » (n = 11)
- **Objectifs**
  - Performance des IA ?
  - L'IA peut-elle contribuer à améliorer les sujets d'examens ?

Résultats - Analyse quantitative

	Type d'examen	Étudiants en médecine		ChatGPT 4 Omni	Gemini Advanced
		Médiane	Max		
Partiels locaux	L3	14.2	18.9	14.9	14.7
	M1	11.9	16.2	14.9	15.1
	M2	12.8	16.3	12.9	13.8
European Prescribing Exam – 1	Knowledge	16.9	18.8	19.7	17.8
	Skills	8.2	14.5	15.5	15.5
European Prescribing Exam – 2	Knowledge	14.9	18.8	17.0	18.1
	Skills	1.3	10.9	18.2	16.4

Résultats - Analyse quantitative

ChatGPT4 Omni

Gemini Advanced

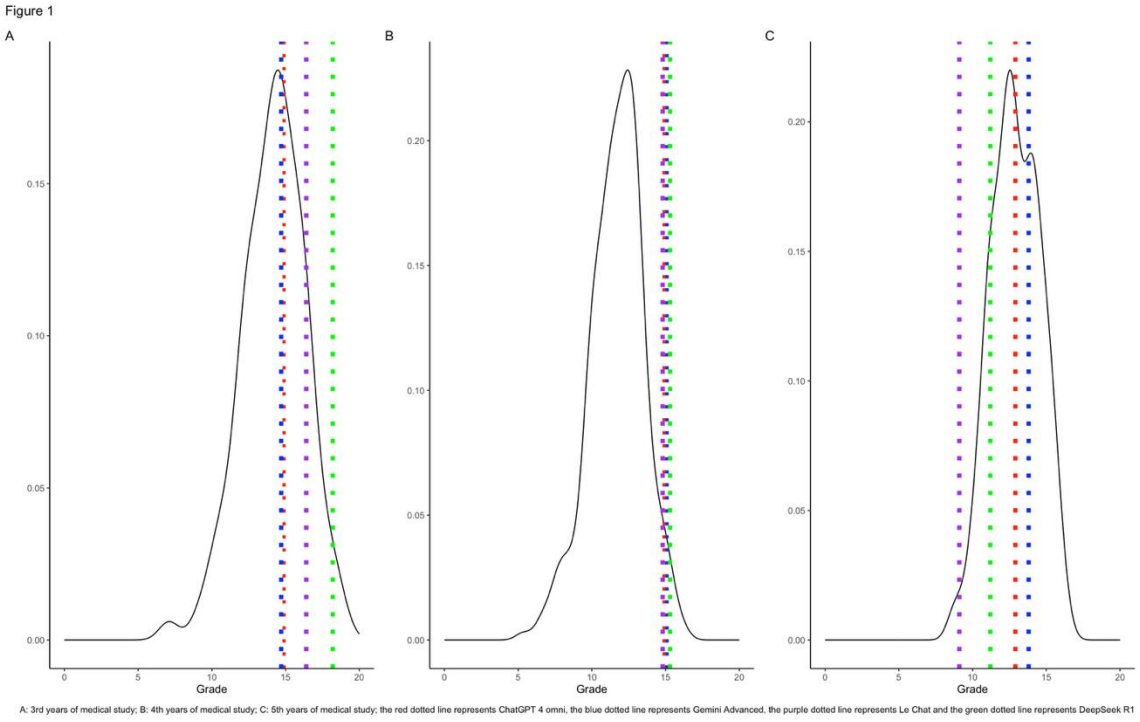
Le Chat

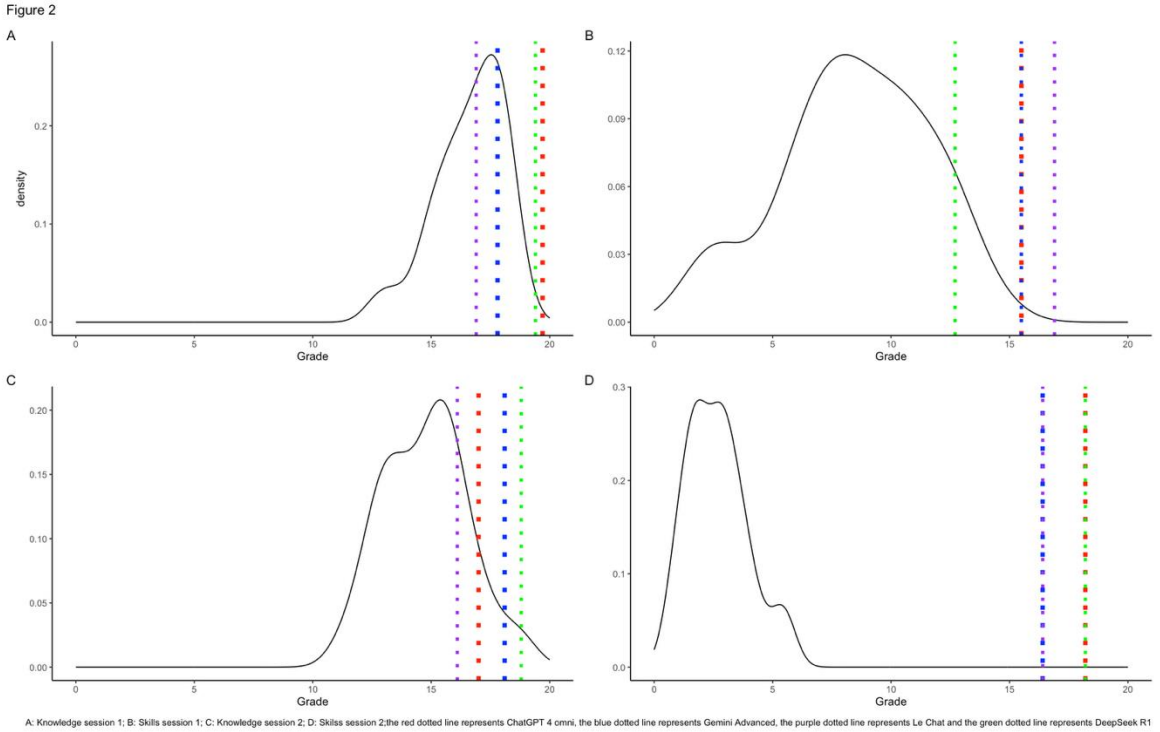
DeepSeek R1

L3 M1 M2

Knowledge

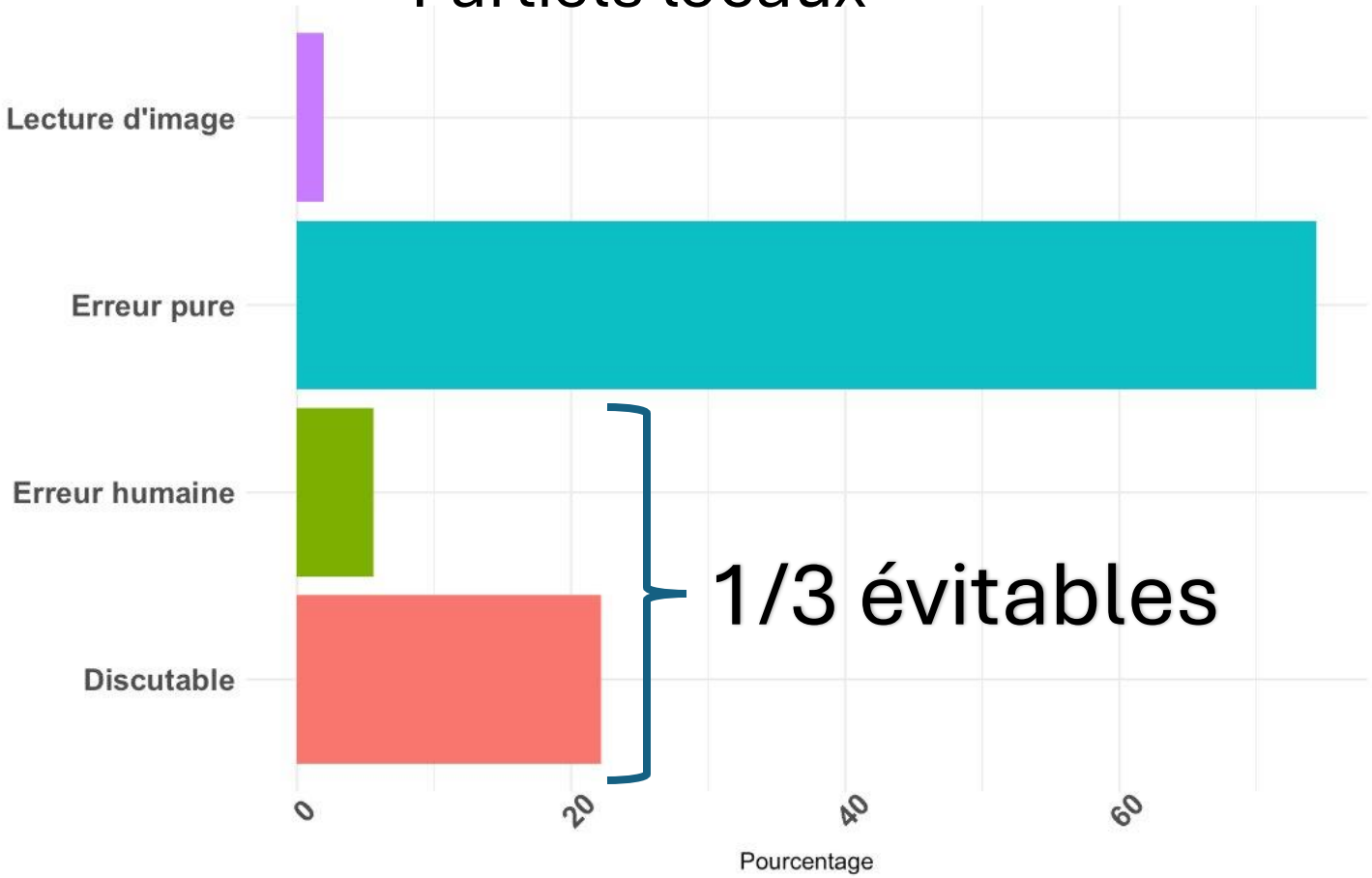
Skills





Résultats - Analyse qualitative

Partiels locaux



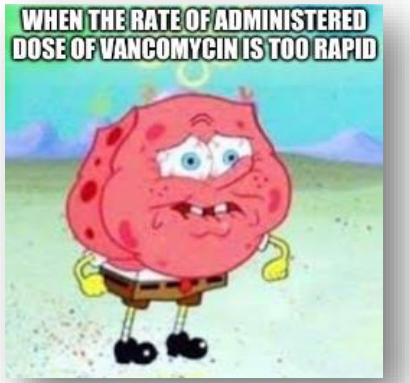
European Prescribing Exam

➔ 100% d'erreurs pures !

Résultats - Exemple d'erreur évitable

"L'administration en perfusion lente de vancomycine est un facteur de risque de survenue d'un red man syndrome"

➔ **Faux** : car perfusion rapide (<60 min) = facteur de risque



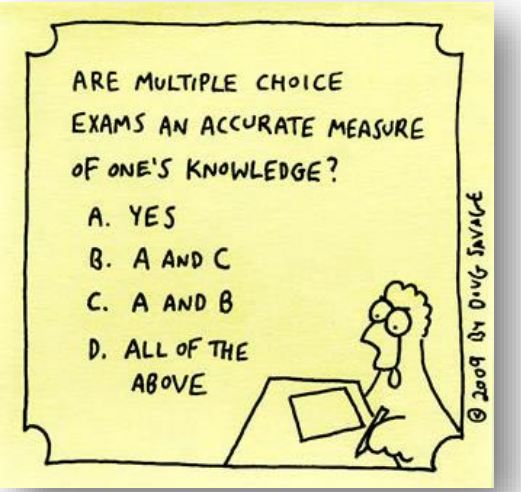
C'est vrai. L'administration en perfusion lente de vancomycine peut augmenter le risque de survenue du syndrome du "red man". Ce syndrome est une réaction cutanée qui se manifeste par des rougeurs et des démangeaisons, généralement sur le visage et le cou, en raison d'une libération rapide d'histamine.

... Effectivement, **plus à risque que de ne pas en injecter** !

Résultats - Diagnostic de QCM imparfait

Lorsque deux IA donnent la même **mauvaise réponse**

➔ Prédit un **QCM imparfait** avec :



- **Spécificité** : 93%

- **Valeur Prédictive Négative** : 86%

(Si les deux IA ne donnent pas la même mauvaise réponse, le QCM est très probablement bien rédigé)

Messages Clés – *Take-Home Messages*

- En Pharmacologie, l'IA fait **au moins aussi bien que l'étudiant moyen**
- Elle fait probablement même mieux lorsqu'il s'agit de **prescrire**
- L'utilisation combinée de plusieurs IA pourrait permettre de **simuler un panel de relecteurs**, en identifiant les QCMs problématiques

Notre questionnement – Nos réponses

- **Généralisable** ?

➔ Réplication sur examens Méd. Intensive Réa (Dr R. Lombardi)

- Utilisation de l'IA pour simuler un panel de **TCS** ?

➔ Expérimentations en cours (Pr J-P. Fournier)

- Utilisation de l'IA pour juger les **ECOS** ?

➔ IA et interprétation de vidéos ?