27es JNI 2026 PARIS
et la région Île-de-France — Palais des Congrès

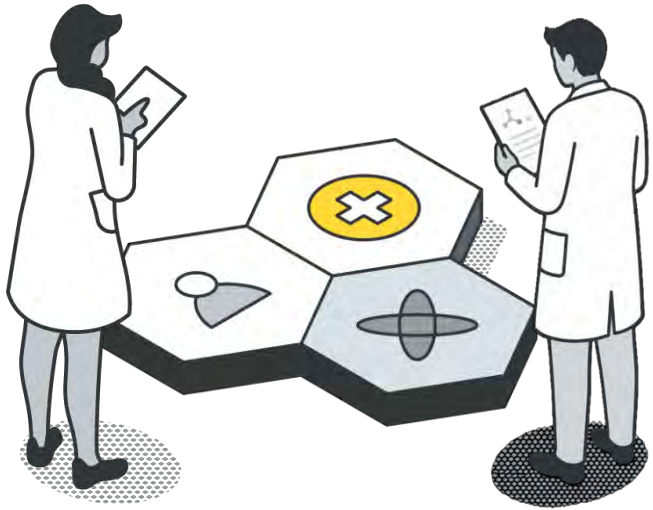
Journées Nationales d'infectiologie
du jeudi 18 juin au samedi 20 juin 2026
Journée Nationale de Formation des Paramédicaux en Infectiologie
Vendredi 19 juin 2026

& Infectiologie

Groupe Infectiologie Digit@le

Best Of — quelle preuve clinique en 2026 ?

Pr David Morquin — CHU de Montpellier



DE QUELLE IA PARLE-T-ON ?

IA PRÉDICTIVE

« Quelle est la bonne réponse ? »

Données structurées → une catégorie / un score.

Normal/pas normal.
Viral/Bactérien
Risque élevé/faible

IA GÉNÉRATIVE

« Quelle est la suite plausible ? »

Texte libre → texte libre (génère, ne récupère pas).

bandelette urinaire protéinurie hématurie nitrites négatifs leucocytes négatifs, FeNa 3.5%, natriurie 45 mmol/L, osmolarité urinaire 320 mOsm/kg, cylindres granuleux bruns, échographie rénale RAS pas d'obstacle reins taille normale, IRA intrinsèque post-hypoperfusion. Le diagnostic le plus probable est…

SYSTÈME AGENTIQUE

« Quelle est la prochaine action ? »

IA générative + outils : planifie, agit, vérifie.

Il faut identifier les étapes . .
Il convient d'interroger …

Interrogation référentiel

Réflexion en cours

DE LA PERFORMANCE D'UN SYSTEME D'IA À LA VALEUR CLINIQUE

Jauge de maturité pour chaque étude

4 états possibles pour chaque étape

- Atteint — preuve positive apportée
- Partiel — preuve amorcée mais incomplète
- Testé mais négatif ou insuffisant — l'étape est évaluée, la preuve manque
- Non évalué — l'étude ne dit rien à cette étape

1/8

BON USAGE DES ATB

JAMA Internal Medicine | Original Investigation | LESS IS MORE

Improving Empiric Antibiotic Selection for Patients Hospitalized With Skin and Soft Tissue Infection

The INSPIRE 3 Skin and Soft Tissue Randomized Clinical Trial

Gohil SK et al. — 2025

LE CONCEPT

- Patient hospitalisé → prescription d'ATB à large spectre
- **Calcul en temps réel du risque individuel de BMR**
- **Si risque < 10 % → alerte + proposition de désescalade**
- En parallèle : formation et restitution mensuelle aux prescripteurs

L'ÉTUDE

118 562 patients admis pour SSTI (hors réanimation) · 92 hôpitaux communautaires américains · essai randomisé en grappes

	Stewardship habituel (n = 57 837)	CPOE bundle (n = 60 725)
Jours d'ATB empirique à spectre étendu / 1000 j	511,5 → 488,7	**496,2 → 359,1**
Patients avec ≥ 1 ATB empirique à spectre étendu	57,0 % → 56,0 %	**55,4 % → 43,0 %**
Transfert en réanimation	3,0 %	3,0 %
Durée moyenne de séjour	6,5 jours	6,4 jours

RÉSULTATS CLÉS

−28 %

jours de traitement empirique anti-Pseudomonas / BGN multirésistants

−10 %

usage empirique de vancomycine (analyse post-hoc)

L'IA UTILISÉE ?

Modèle prédictif

exploitant les variables du dossier : démographie, expositions aux soins et aux antibiotiques, antécédents microbiologiques, comorbidités, biologie d'admission, prévalence locale des BMR.

Entraîné sur 195 040 patients (151 hôpitaux).

1/8

QUE NOUS APPREND CETTE ÉTUDE SUR L'IA ?

BON USAGE DES ATB

Ce n'est pas le modèle qui permet l'action, c'est l'architecture de choix : ***le bon signal, au bon moment, déclenche une action.***

L'ancrage dans la pratique et le workflow est parfait

1 · Modèle qualité & usage	2 · Validation externe multicentrique	3 · Essai d'impact prospectif	4 · Cadre réglementaire	5 · Intégration workflow & supervision	6 · Surveillance post-déploiement	Impact clinique
●	●	●	○	●	◕	●
Atteint modèle trivial classe >95% des patients « bas risque »	**Atteint** 92 hôpitaux, 15 États	**Atteint** cluster RCT	**Non évalué**	**Atteint** Prescription /DPI	**Partiel** audit algorithme	**Atteint** –28% ATB large spectre

Le succès est sociotechnique, pas purement algorithmique

La performance ne vient pas du modèle seul, mais du **système** qui l'entoure : intégration DPI · alerte · formation · feedback · gouvernance — *c'est l'ensemble qui change la prescription.*

SAME PLAYER SHOOT AGAIN

BON USAGE DES ATB

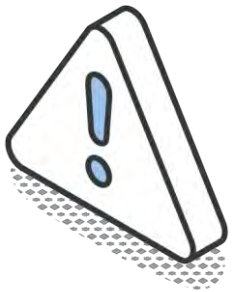
Réutilisation méthode, architecture, intégration (modèle différent)

JAMA Surgery | Original Investigation

Improving Empiric Antibiotic Selection for Patients Hospitalized With Abdominal Infection
The INSPIRE 4 Cluster Randomized Clinical Trial

Gohil SK et al. — 2025

−35 %
jours d'antibiothérapie empirique à large spectre (réduction relative)



Problème : scalable
ne veut pas dire universellement portable
(Système propriétaire MEDITECH)

3/8

BON USAGE DES ATB

Enhancing quality of antimicrobial prescribing through 'Ask Eolas' (language model): a user-testing and simulation evaluation

Waldock WJ et al. — 2026

npj | antimicrobials and resistance

LE CONCEPT

- Prescription antimicrobienne → besoin d'accéder rapidement aux recommandations locales. PDF ou applications statiques peuvent être difficiles à interroger.
- Question en langage naturel à un assistant IA → **réponse synthétique + justification clinique + liens vers les sources.**
- Le clinicien devait d'abord formuler sa propre décision avant de consulter l'IA
- Objectif : **améliorer la concordance aux recommandations**, réduire les erreurs et la charge cognitive

L'ÉTUDE

45 professionnels de santé · simulation monocentrique · 3 bras randomisés

Guidelines PDF vs Eolas App vs Ask Eolas · cas de complexité croissante

	Guidelines PDF (n = 15)	Eolas App (n = 15)	Ask Eolas (n = 15)
Prescriptions exactes	7/15 47 %	9/15 60 %	**15/15** **100 %**
Erreurs documentées	8	6	**0**
Confiance de prescription médiane	72%	68%	**94%**
Charge cognitive NASA-TLX	37	36	**14**

RÉSULTATS CLÉS

100 %

prescriptions exactes
avec Ask Eolas (15/15) vs 47% recherche classique

NNT 1,9

0

erreur documentée
vs 6 et 8 avec comparateurs

L'IA UTILISÉE ?

Modèle de langage pré-entraîné utilisé pour :

- comprendre une question clinique en langage naturel
- rechercher sémantiquement dans les recommandations locales
- récupérer les extraits pertinents
- produire une réponse courte, contextualisée et sourcée
- afficher les liens vers les recommandations hospitalières utilisées

3/8

BON USAGE DES ATB

QUE NOUS APPREND CETTE ÉTUDE SUR L'IA ?

Ask Eolas ne résout pas un problème de raisonnement médical, il résout un problème d'ergonomie documentaire → *c'est l'interrogeabilité du référentiel qui réduit les erreurs ...mais évalué uniquement en simulation*

1 · Modèle qualité & usage	2 · Validation externe multicentrique	3 · Essai d'impact	4 · Cadre réglementaire	5 · Intégration workflow & superv.	6 · Surveillance post-déploiement	Impact clinique
●	○	●	○	●	○	○
Atteint RAG + LLM	Non évalué	**Insuffisant** simulation, n=45	Non évalué	**Partiel** app déployée, testée simulation	Non évalué	Non évalué

L'atout rare : le passage à l'échelle est **potentiellement acquis**.
Eolas équipe ~85% des hôpitaux du NHS
Il manque encore l'intégration au DPI, données live, impact patient.

4/8

SEPSIS ?

Clinical Decision Support for Septic Shock in the Emergency Department: A Cluster Randomized Trial

Scott HF et al. — 2025

PEDIATRICS

LE CONCEPT

Enfant aux urgences avec suspicion de sepsis

- Prédire le risque de choc septique avant hypotension
- **Calcul automatique du risque à l'arrivée puis à H+2**
- Si haut risque → alerte Epic + actions suggérées
- Approche "assistance" : l'IA intervient après la suspicion clinique

L'ÉTUDE

1 331 participants · 4 urgences pédiatriques · essai randomisé stepped-wedge

	Soins usuels (n = 352)	Soins avec prédiction (n = 979)
Soins concordants dans l'heure	38,9%	39%
Choc septique	11,1%	12,7%
Délai administration antibiothérapie	41 min	47 min
Mortalité à J30	0,3%	0,7%
Bilan pour analyse de dysfonctions d'organe	16,8%	**31,9%**

RÉSULTATS CLÉS

Aucun effet positif vs témoin

0,3% des passages aux Urgences déclenchent une alerte

4/4 services pilotes maintiennent le dispositif au delà de l'étude

L'IA UTILISÉE ?

Modèles prédictifs intégrés dans le DPI, estimant le risque de choc septique dans les 24 h à partir des données du DPI.

Deux modèles :
à l'arrivée (10 variables)
puis **à 2 heures** (idem + 9 biologies),
seuil de sensibilité à 90%

Epic

QUE NOUS APPREND CETTE ETUDE SUR L'IA ?

SEPSIS ?

Une IA prédictive peut être performante, faisable, intégrée et bien acceptée, testée dans un essai randomisé

...et pourtant adoptée ?

...mais rester cliniquement neutre si elle intervient dans un système déjà performant ou si son signal ne déclenche pas une action réellement différenciante.

1 · Modèle	2 · Validation	3 · Essai	4 · Cadre	5 · Intégration	6 · Surveillance	Impact
qualité & usage	externe multicentrique	d'impact prospectif	réglementaire	workflow & supervision	post-déploiement	clinique
●	●	●	○	●	○	●
Atteint 2 modèles, seuil 90%	**Atteint** 4 sites pédiatriques	**Atteint** RCT stepped-wedge	**Non évalué**	**Atteint** alerte Epic + actions Poursuite au delà des pilotes	**Non évalué**	**Testé — négatif** aucun effet vs témoin

+ l'alerte peut modifier ce qu'elle prétend mesurer — soit en évitant le choc, soit en augmentant les investigations et l'étiquetage sepsis sans bénéfice démontré.

→ Démêler cet effet d'autoprophétie exige une méthodologie causale dédiée, insuffisante ici

5/8

SEPSIS ?

An artificial intelligence-powered learning health system to improve sepsis detection and quality of care: a before-and-after study

Despraz J et al. —2026

npj | digital medicine

LE CONCEPT

- Patient hospitalisé → données du DPI agrégées par fragments de 6 h
- HERACLES (Random Forest + LSTM) classe : pas de sepsis / possible / confirmé
- **Phénotypage rétrospectif post-sortie** → indicateurs qualité (pas d'alerte temps réel)
- **Intervention : démarche amélioration continue** (Parcours sepsis standardisé + registre + dashboards + réentraînement mensuel)

L'ÉTUDE

123 410 séjours · CHUV Lausanne · observationnelle · prospective · quasi-expérimentale, monocentrique · de type avant-après avec groupe contrôle non randomisé

	Avant → Après	OR (IC 95 %)
Codage ICD-10 sepsis (wards SLHS)	1,55 → 3,58 %	1,27 (1,23–1,32)
Mortalité hosp. (cas flaggés)	20,5 → 15,3 %	0,89 (0,83–0,96)
Mortalité 90 j (cas flaggés)	33,0 → 26,1 %	0,91 (0,86–0,97)
Wards contrôles (codage, mortalité)	—	NS

RÉSULTATS CLÉS

−26* %

mortalité hospitalière des cas de sepsis détectés (réduction relative)

−21 %

mortalité à 90 jours des cas de sepsis détectés

L'IA UTILISÉE ?

HERACLES : ensemble Random Forest + LSTM classant des fragments de 6 h en pas de sepsis / possible / confirmé. Phénotypage rétrospectif alimentant registre et dashboards qualité, avec validation experte (human-in-the-loop).

AUROC 0,88 · réentraînement mensuel continu (learning health system).

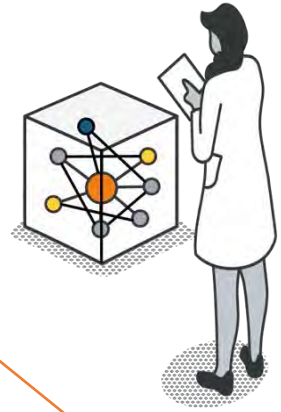
QUE NOUS APPREND CETTE ETUDE SUR L'IA ?

SEPSIS ?

Le Learning Health System déplace la valeur : le bénéfice ne vient pas du modèle, mais de la ***boucle institutionnelle apprenante***

= parcours standardisé + registre + dashboards + réentraînement continu + gouvernance qualité.

...mais à quel niveau de preuve ?



1 · Modèle qualité & usage	2 · Validation externe multicentrique	3 · Essai d'impact prospectif	4 · Cadre réglementaire	5 · Intégration workflow & supervision	6 · Surveillance post-déploiement	Impact clinique
●	●	●	○	●	●	●
Atteint RF + LSTM · AUROC 0,88	**Partiel** 2 sets internes, monocentrique	**Testé — insuffisant** avant-après, contrôles non comparables	**Non évalué**	**Atteint** parcours + registre + dashboards validation experte (human-in-the-loop)	**Atteint**	**Atteint — sous réserve** ↓ mortalité hosp. et 90 j (associations)

+ en améliorant détection et codage, le système modifie le dénominateur qu'il prétend mesurer (effet Will Rogers).

→

Design avant-après monocentrique avec contrôles non comparables : causalité non démontrable. La surveillance/amélioration continue est ici la fonction native du système.

6/8

SEPSIS ?

Development and prospective implementation of a large language model based system for early sepsis prediction

Shashikumar SP et al, 2025

npj | digital medicine

LE CONCEPT

- Patient aux urgences → COMPOSER (réseau de neurones) calcule un risque de sepsis chaque heure (modèle déjà publié)
- Score 0,50–0,75 = zone d'incertitude → on appelle un LLM (Mixtral 8×7B)
- Le LLM extrait des notes les signes cliniques (RAG, sortie JSON, vote majoritaire ×3)
- Un calcul bayésien classe sepsis vs 18 sepsis-mimics → décision d'alerte

L'ÉTUDE

2 500 passages aux urgences · 2 urgences UC San DIego Health · rétrospectif (1 746) + validation prospective (754), **mode silencieux**

	PPV	Fausses alertes /patient·h
COMPOSER seul	22,6 %	0,037
+ vraisemblance (SLT)	31,9 %	0,021
+ diagnostic différentiel (DDx)	52,9 %	0,0087
Validation prospective (Composer + DDx)	58,2 %	0,0086

RÉSULTATS CLÉS

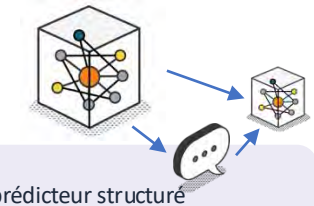
52,9 %

valeur prédictive positive
(vs 22,6 % pour COMPOSER seul)

÷ 4

fausses alertes par patient·heure (0,037 → 0,0087)

L'IA UTILISÉE ?



Architecture hybride : prédicteur structuré (COMPOSER) + LLM-extracteur ciblé sur la zone d'incertitude + raisonnement bayésien explicite. Le LLM seul est mauvais diagnosticien (F1 34,7 %) ; c'est l'architecture qui crée la valeur.

Mixtral 8×7B open-source, on-premise (HIPAA) · ~10 appels LLM/jour · explicable (JSON justifié).

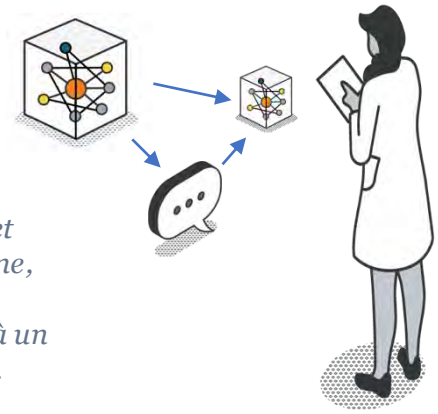
QUE NOUS APPREND CETTE ETUDE SUR L'IA ?

SEPSIS ?

Le LLM seul est un mauvais diagnosticien (F1 34,7 %). La valeur ne vient pas du LLM, mais de l'architecture

→

cantonner le LLM à une tâche pour laquelle il est fiable *(lire du texte libre et répondre oui/non sur la présence d'un signe, avec justification) et* ***sortir le raisonnement du LLM*** *pour le confier à un calculateur bayésien explicite et auditable.*



1 · Modèle qualité & usage	2 · Validation externe multicentrique	3 · Essai d'impact prospectif	4 · Cadre réglementaire	5 · Intégration workflow & supervision	6 · Surveillance post-déploiement	Impact clinique
●	●	●	○	●	○	○
Atteint	**Partiel**	**Testé — insuffisant**	**Non évalué**	**Atteint**	**Non évalué**	**Non évalué**
PPV 53 % · Fausses alertes/h ÷4	2 EDs, un seul système (UCSD)	prospectif mais mode silencieux		Capacité de traiter le DPI en temps réél		

→

Déploiement prospectif réel mais silencieux
La performance est démontrée, le bénéfice clinique reste à prouver.

7/8

DIAGNOSTIC ?

Integrating a host biomarker with a large language model for diagnosis of lower respiratory tract infection

Phan HV et al. — 2025

nature communications

LE CONCEPT

- Patient en réanimation, insuffisance respiratoire aiguë → infection respiratoire basse ?
- FABP4 : biomarqueur transcriptomique pulmonaire (RNA-seq sur aspiration trachéale)
- GPT-4 lit le dossier (1 note + 1 compte-rendu de radio) et juge IRB / pas IRB
- Régression logistique combinant les deux signaux → diagnostic d'IRB

L'ÉTUDE

145 patients de réanimation · UC SF · 2 cohortes observationnelles · dérivation (86) + validation (47) · cas indéterminés exclus

	AUC	Précision
Équipe médicale	—	72 %
FABP4 seul	0,84	—
GPT-4 seul	0,83	—
Combiné FABP4/GPT-4	0,93	84 %

RÉSULTATS CLÉS

0,93

AUC du classifieur combiné (dérivation) · 0,98 en validation

84 %

précision diagnostique vs 72 % pour l'équipe médicale

L'IA UTILISÉE ?

Multimodalité : biomarqueur biologique (FABP4) + GPT-4 lisant le dossier, combinés par régression logistique. Chaque signal seul ~0,83 ; ensemble 0,93. À données égales, GPT-4 fait jeu égal avec les médecins (84 %).

GPT-4 turbo commercial (instance HIPAA), prompt chain-of-thought, 3 sessions/patient (vote).

7/8

QUE NOUS APPREND CETTE ETUDE SUR L'IA ?

DIAGNOSTIC ?

Deux signaux faibles de natures opposées — un biomarqueur biologique et un LLM lisant le dossier — médiocres seuls → *deviennent excellents combinés (AUC 0,93–0,98).* ***La valeur naît de la complémentarité des modalités****, pas de la sophistication d'un modèle.*

...mais à l'état de preuve de concept

1 · Modèle qualité & usage	2 · Validation externe multicentrique	3 · Essai d'impact prospectif	4 · Cadre réglementaire	5 · Intégration Workflow & supervision	6 · Surveillance post-déploiement	Impact clinique
●	●	○	○	○	○	○
Atteint	**Partiel**	**Non évalué**	**Non évalué**	**Non évalué**	**Non évalué**	**Non évalué**
combiné FABP4+GPT-4 · AUC 0,93	2e cohorte indépendante mais même centre, n faible	usage rétrospectif, aucune décision clinique réelle		GPT-4 en chat manuel FABP4 = RNA-seq, non dispo au lit		impact stewardship théorique estimé : les antibiotiques inappropriés auraient pu être évités chez 20/24 (83 %) et 7/7 (100 %) des patients pas-IRB traités à tort

8/8

DIAGNOSTIC ?

Population-scale cross-sectional observational study for AI-powered TB screening on one million CXRs

Munjal P et al. — 2025

npj | digital medicine

LE CONCEPT

- Dépistage TB de masse par radio de thorax (Abu Dhabi, ~2 000 radios/jour, prévalence < 0,5 %)
- AIRIS-TB (CNN EfficientNet) ne diagnostique pas : il automatise le rapport des radios normales
- Tout cliché avec anomalie est déféré au radiologue → supervision humaine native
- Métrique-reine : ne jamais manquer une TB (TB-FNR), pas l'exactitude globale

L'ÉTUDE

1,04 million de radios de thorax · Abu Dhabi · quasi-prospectif (split temporel) + 3 jeux externes · analyse d'équité · **Gold Standard : annotation radiologique**

	FNR global	Volume d'automatisation possible
Radiologues (2ième relecture)	1,85 %	—
AIRIS-TB — réglage safe	0,33 %	40 %
AIRIS-TB — réglage optimal	1,57 %	80 %
TB-FNR (cas TB manqués)	0 %	—

RÉSULTATS CLÉS

0 %

cas de TB confirmée manqués (TB-FNR) sur 1,04 million de radios

80 %

des radios automatisables (FNR 1,57 % < radiologues 1,85 %)

L'IA UTILISÉE ?

Vision par ordinateur (CNN EfficientNet) entraînée sur 10 ans de radios. Modèle classique : la valeur vient de la validation à très grande échelle (1 M de radios, split temporel) et de l'analyse d'équité

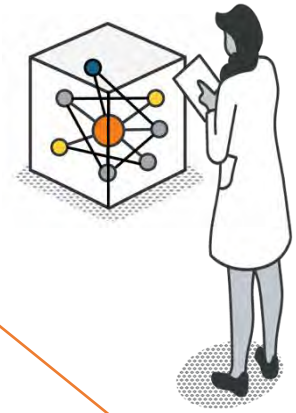
8/8

QUE NOUS APPREND CETTE ETUDE SUR L'IA ?

DIAGNOSTIC ?

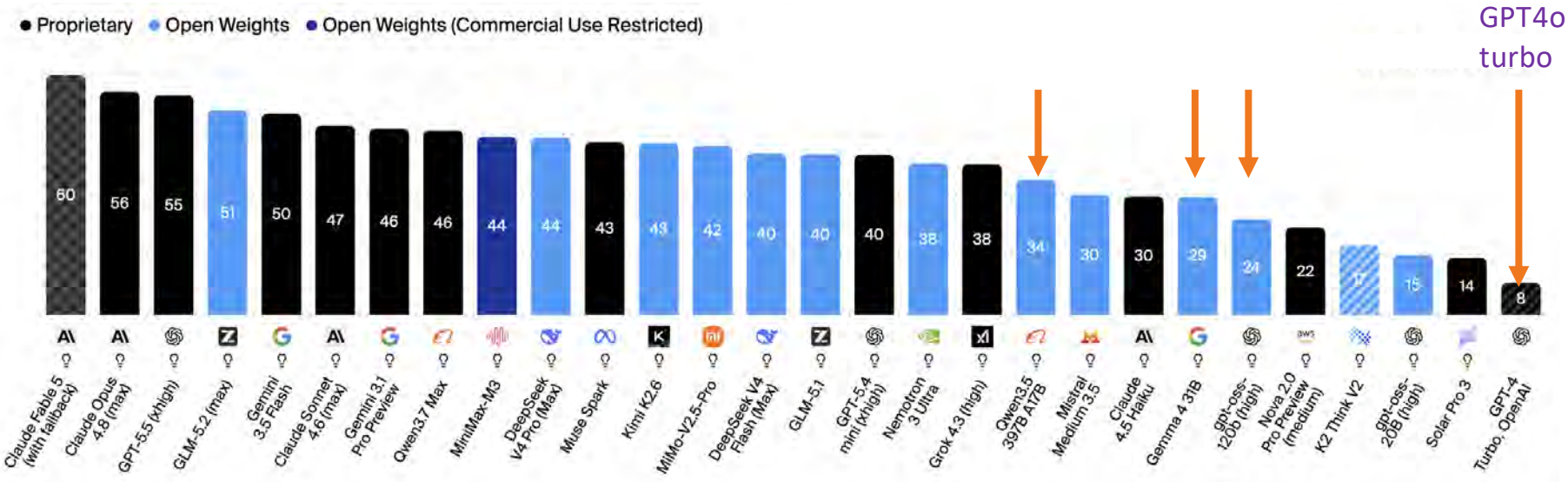
Repositionnement « malin » : l'IA ne diagnostique pas les TB, elle élimine les radiographies normales.
*La métrique-reine est la sécurité (**ne manquer aucune TB**), pas l'exactitude.*

...mais l'usage en vie réelle reste à évaluer



1 · Modèle qualité & usage	2 · Validation externe multicentrique	3 · Essai d'impact prospectif	4 · Cadre réglementaire	5. Intégration Workflow & supervision	6 · Surveillance post-déploiement	Impact clinique
Atteint CNN · AUC 98,5 % · TB-FNR 0 %	**Atteint** 1 M radios + 3 jeux externes + équité	**Testé — insuffisant** « Quasi-prospectif » (test temporel sur 2 ans de données réelles a posteriori	**Partiel** Alignement explicite sur le WHO Target Product Profile (≥90 % sensibilité, ≥70 % spécificité — dépassé), reporting MI-CLAIM, comité d'éthique d'Abu Dhabi.	**Partiel** Workflow conceptualisé mais non encore déployé	**Partiel** Le test sur 2 cycles annuels capture explicitement le drift saisonnier et la dérive de distribution	**Partiel** Non en production Impact charge calculé par simulation

L'OBSOLESCENCE DES MODÈLES VA PLUS VITE
QUE NOTRE CAPACITE A PRODUIRE DES PREUVES DE LEUR UTILITE CLINQUE



Artificial Analysis Intelligence Index v4.1

...

HORS HOPITAL ?

EDUCATION THERAPEUTIQUE, PREVENTION, RELATION PATIENT, SANTE SEXUELLE...

Impact clinique

Atteint

nature medicine

Article https://doi.org/10.1038/s41591-025-03618-6

A vaccine chatbot intervention for parents to improve HPV vaccination uptake among middle school girls: a cluster randomized trial

JOURNAL OF MEDICAL INTERNET RESEARCH Li et al

Original Paper

Effects of a Theory- and Evidence-Based, Motivational Interviewing–Oriented Artificial Intelligence Digital Assistant on Vaccine Attitudes: A Randomized Controlled Trial

Yan Li[1,2,3], PhD; Mengqi Li[1,2], PhD; Janelle Yorke[1,4], PhD; Daniel Bressington[5], PhD; Joyce Chung[1], PhD; Yao-Jie Xie[1,6], PhD; Lin Yang[1], PhD; Mengting He[1], MS; Tsz-Ching Sun[1], MS; Angela Y M Leung[1], PhD

JOURNAL OF MEDICAL INTERNET RESEARC

Original Paper

Evaluation of an Artificial Intelligence Conversational Chatbot to Enhance HIV Preexposure Prophylaxis Uptake: Development and Usability Internal Testing

Jun Tao[1], PhD; Ellie Pavlick[2], PhD; Amaris Grondin[2], BS; Josue D Bustamante[3], BS; Harrison Martin[1], BA; Hannah Parent[1], MPH; Natalie Fenn[1,4], PhD; Alexi Almonte[1], BA; Amanda Maguire-Wilkerson[1], DrPH; Mofan Gu[5], PhD; Jack Rusley[6,7], MHS, MD; Bryce K Perler[1], MSTR, MD; Tyler Wray[8], PhD; Amy S Nunn[8,9], SCD; Philip A Chan[1,8,9], MS, MD

Conclusions

***1**. La valeur clinique n'est pas seulement dans l'algorithme*

Intégration dans le processus de travail
Espace sociotechnique
Architecture, Combinaison
Qualité des données
Multimodalité

***2.** Ce qui peut nous amener le plus de transformation ne se publie pas forcément dans notre discipline*

RESEARCH ARTICLES

MACHINE LEARNING

Performance of a large language model on the reasoning tasks of a physician

Peter G. Brodeur[1]†, Thomas A. Buckley[2]†, Zahir Kanjee[1], Ethan Goh[3,4], Evelyn Bin Ling[5], Priyank Jain[6], Stephanie Cabral[1,7], Raja-Elie Abdulnour[8], Adrian D. Haimovich[9], Jason A. Freed[10], Andrew Olson[11], Daniel J. Morgan[12,13], Jason Hom[5], Robert Gallo[14,15], Liam G. McCoy[1,16,17], Haadi Mombini[18], Christopher Lucas[1], Misha Fotoohi[1], Matthew Gwiazdon[1], Daniele Restifo[1], Daniel Restrepo[19], Eric Horvitz[20,21], Jonathan Chen[3,4,22]‡, Arjun K. Manrai[2]‡*, Adam Rodman[1]‡*

Science, Avril 2026

Article

Towards autonomous medical artificial intelligence agents

https://doi.org/10.1038/s41586-026-10575-5
Received: 29 April 2025
Accepted: 18 May 2026
Published online: 17 June 2026

Dyke Ferber[1,2], Lars Hilgers[3,4], Christiane Höper[5], Benedict Kinny-Köster[4,6], Jan-Niklas Eckardt[3,7], Katharina Egger-Heidrich[3], Marius Bill[3], Martin M. K. Schneider[3], Jan Clusmann[2,3,8,9], Lejla Kadric[3], Marcel Oehme[3], Maximilian Mayrhofer-Schmid[10,11], Alexander Oeser[11], Georg Wölflein[2], Isabella C. Wiest[2,9], Jan Moritz Middeke[3,7], A. John Iafrate[12], Daniel Truhn[13], Dirk Jäger[3] & Jakob Nikolas Kather[2,3,8,9]

nature

https://doi.org/10.1038/s41

Accelerated Article Preview

Towards Conversational AI for Disease Management

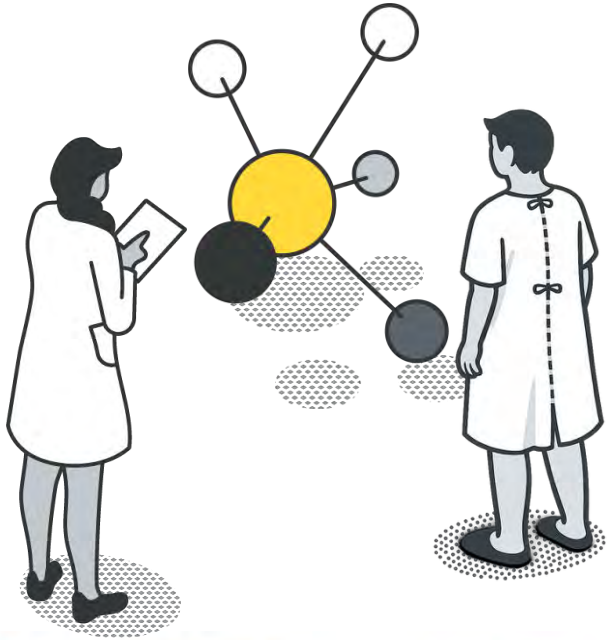
Juin 2026

***3.** Dans la recherche comme dans les soins ce que le calcul ne réalise pas doit focaliser notre attention*

La compréhension de l'impact réel, de l'utilité
L'appréciation du contexte
La gouvernance de l'incertitude
La gestion des biais, des risques
La qualité et la sécurité des soins
La recherche de nos angles morts
L'engagement relationnel

Merci

De votre attention



Groupe Infectiologie Digit@le