

Intelligence Artificielle au service des Urgences

CIPIQ-S Octobre 2024

Flament Julien



Plan

- Définitions
- Place(s) de l'IA en Médecine d'Urgence
- Exemples
- Ecueils

Intelligence Artificielle (AI)

- L'IA désigne la possibilité pour une machine de reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité.



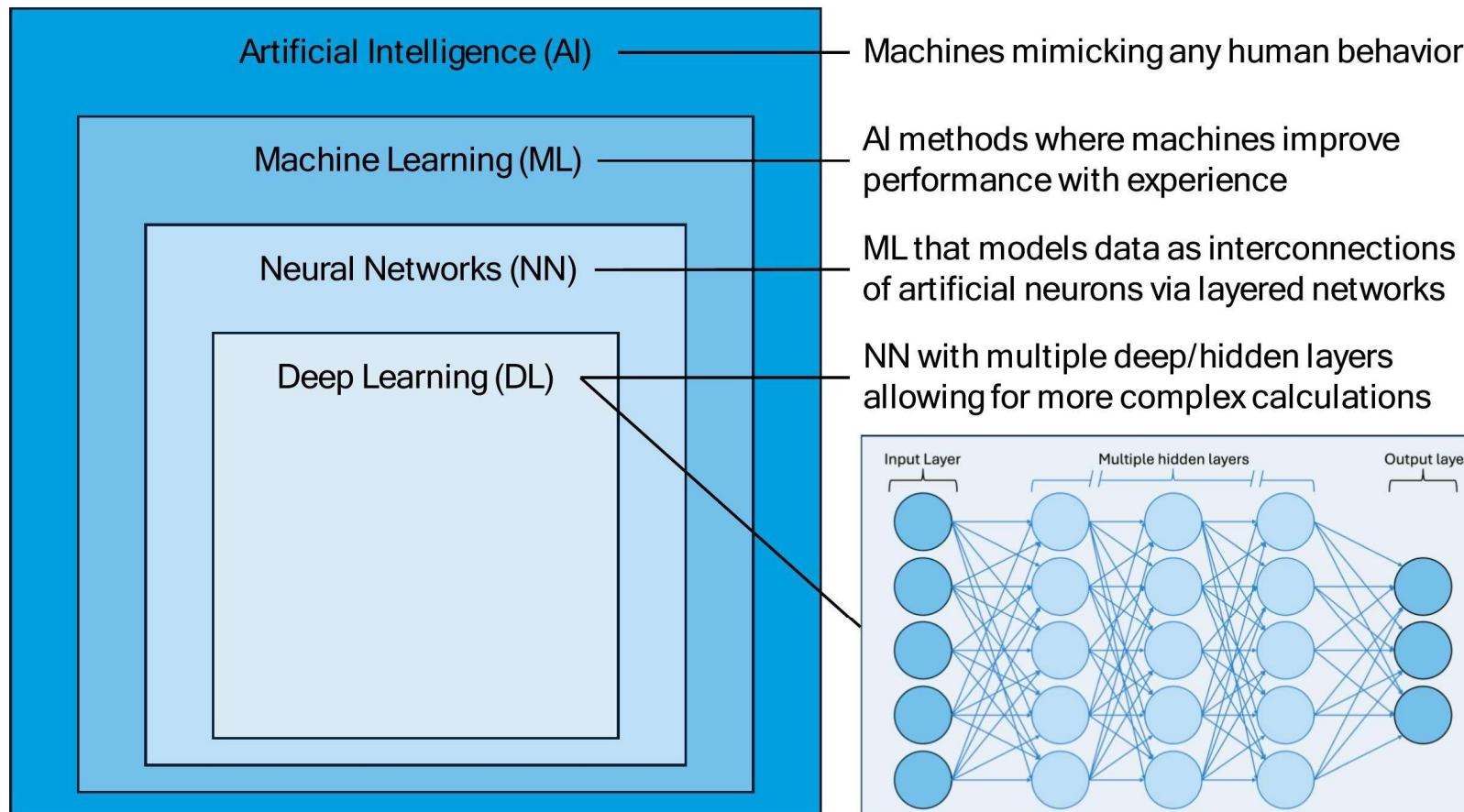
American Medical Association

« Intelligence **Augmentée** »
plutôt que
« Intelligence Artificielle »
en médecine afin de souligner le rôle
d'assistance des ordinateurs dans
l'amélioration des capacités humaines, plutôt
que dans leur remplacement

Intelligence Artificielle (AI)

- Les technologies d'intelligence artificielle, ou IA, permettent aux ordinateurs et aux machines de simuler l'intelligence humaine et ses capacités de résolution des problèmes.
- Imitateur des fonctions cognitives humaines
- Basé sur des règles (algorithmes)
- Basé sur des exemples (recherche de patterns similaires): Machine Learning (ML), Neural Network (NN) et Deep Learning (DL)





Kachman MM, Brennan I, Oskvarek JJ, Waseem T, Pines JM.
 How artificial intelligence could transform emergency care. *Am J Emerg Med.* 2024;81:40-46. doi:10.1016/j.ajem.2024.04.024

Emergency Department Service d'Urgences



- un service hospitalier dont le personnel est présent 24 heures sur 24, 7 jours sur 7, et qui fournit des services ambulatoires non programmés aux patients dont l'état de santé nécessite des soins immédiats



Challenge en Médecine d'Urgence

- Saturation des salles d'urgences:

Augmentation visites

Augmentation des revisites

Visites non urgentes

Manque de personnel 

- Conséquences:

Augmentation mortalité

Augmentation morbidité

Insatisfaction patient

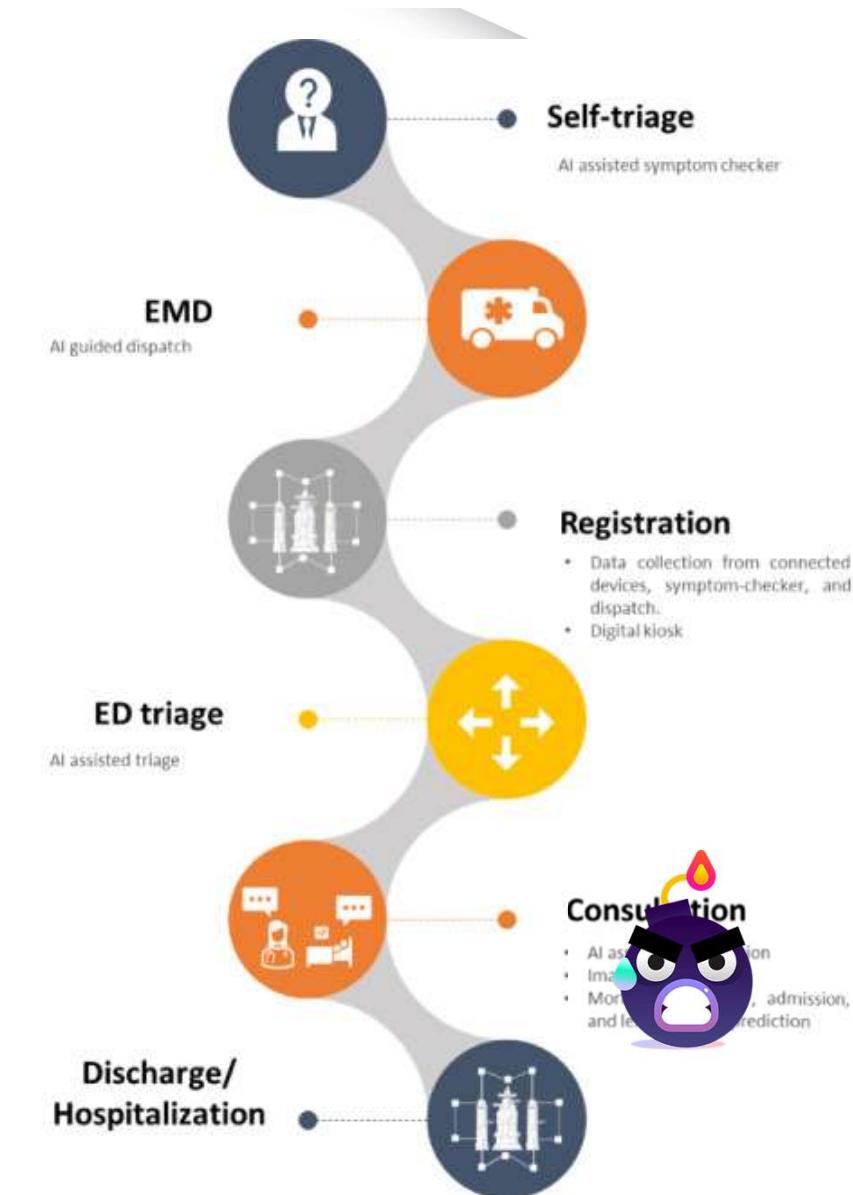
Epuisement du personnel

Hooker EA, Mallow PJ, Oglesby MM. Characteristics and Trends of Emergency Department Visits in the United States (2010-2014). *J Emerg Med.* 2019;56(3):344-351.
doi:10.1016/j.jemermed.2018.12.025

Kulstad EB, Sikka R, Sweis RT, Kelley KM, Rzechula KH. ED overcrowding is associated with an increased frequency of medication errors. *Am J Emerg Med.* 2010;28(3):304-309.
doi:10.1016/j.ajem.2008.12.014

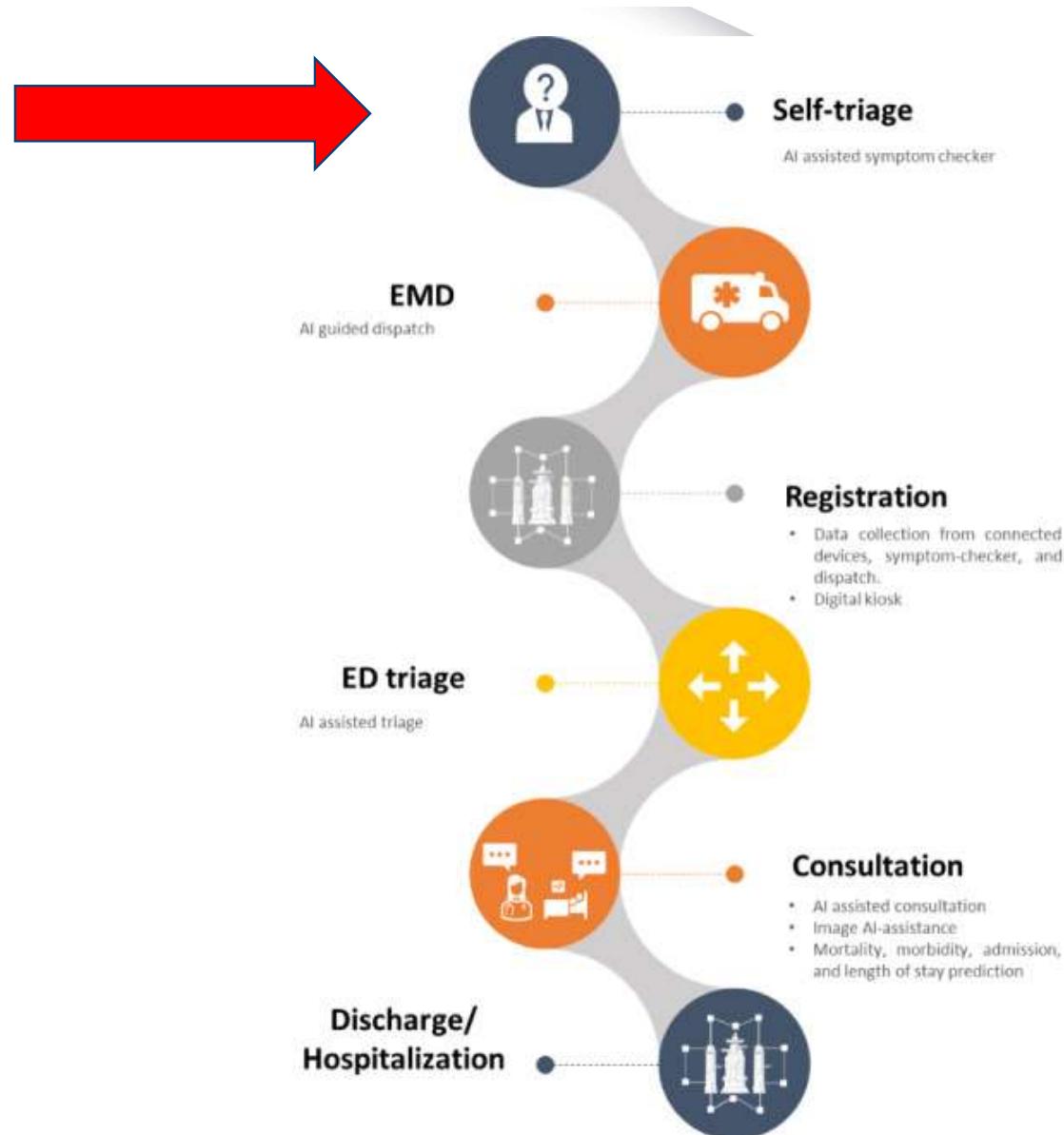
Parcours Patient

Chenais G, Lagarde E, Gil-Jardiné C.
Artificial Intelligence in Emergency
Medicine: Viewpoint of Current
Applications and Foreseeable
Opportunities and Challenges. *J Med
Internet Res.* 2023;25:e40031.
Published 2023 May 23.
doi:10.2196/40031



Parcours Patient

Chenais G, Lagarde E, Gil-Jardiné C. Artificial Intelligence in Emergency Medicine: Viewpoint of Current Applications and Foreseeable Opportunities and Challenges. *J Med Internet Res.* 2023;25:e40031. Published 2023 May 23. doi:10.2196/40031



Self Triage préhospitalier

- 35% des US consultent internet pour autodiagnostic
- When we asked respondents about the accuracy of their initial diagnosis, they reported:
 - 41% of online diagnosers say a medical professional confirmed their diagnosis. An additional 2% say a medical professional partially confirmed it.
 - 35% say they did not visit a clinician to get a professional opinion.
 - 18% say they consulted a medical professional and the clinician either did not agree or offered a different opinion about the condition.
 - 1% say their conversation with a clinician was inconclusive.
 - self diagnosis usually starts with search engines like Google, Bing, or Yahoo

Self Triage préhospitalier

- Les sites internet et applications de « vérificateurs de symptômes » sont ils de bons régulateurs?
- 23 symptoms checkers
ie:
<https://familydoctor.org/diseases-and-conditions/>
ie:
<https://myhealth.alberta.ca/health/Pages/conditions.aspx?hwid=hwsxchk>
- 45 vignettes cliniques

Self Triage préhospitalier

Table 2

Accuracy of diagnosis decision and triage advice for all symptom checkers, stratified by severity of standardized patient (SP) vignette and by frequency of the condition's diagnosis

Type of vignette or diagnosis	No of vignettes (%)	Diagnosis												Triage		
		Listed first				Listed in top 3				Listed in top 20						
		Rate*	% (95% CI)	P value		Rate*	% (95% CI)	P value		Rate*	% (95% CI)	P value	Rate*	% (95% CI)	P value	
All vignettes	45 (100)	262/770	34 (31 to 37)	—		394/770	51 (47 to 54)	—		449/770	58 (55 to 62)	—	301/532	57 (52 to 61)	—	
Type of SP vignette:																
Emergent	15 (33)	64/263	24 (19 to 30)	<0.001		104/263	40 (34 to 46)	<0.001		132/263	50 (44 to 56)	0.003	147/183	80 (75 to 86)		
Non-emergent	15 (33)	98/260	38 (32 to 44)			148/260	57 (51 to 63)			157/260	60 (54 to 66)		96/175	55 (47 to 63)	<0.001	
Self care	15 (33)	100/247	40 (34 to 47)			142/247	57 (51 to 63)			160/247	65 (59 to 71)		58/174	33 (26 to 40)		
Type of diagnosis†:																
Common	26 (58)	174/457	38 (34 to 43)	0.004		254/457	56 (52 to 61)	<0.001		283/457	62 (57 to 66)	0.01	162/313	52 (46 to 57)	0.01	
Uncommon	19 (42)	88/313	28 (23 to 33)			140/313	45 (38 to 49)			166/313	53 (47 to 59)		139/219			

Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ*. 2015;351:h3480. Published 2015 Jul 8.
doi:10.1136/bmj.h3480

Self Triage préhospitalier

- Babylon Triage and Diagnostic System (Babylon AI)

« Applications permettant un diagnostic / une orientation au patient »

- Quelles bases de donnée?
- Employés de Babylon dans l'échantillon
- Pas d'up-to-date depuis 2022
- Diagnostics à sélectionner dans une liste
- Monopathologie

Baker A, Perov Y, Middleton K, et al. A Comparison of Artificial Intelligence and Human Doctors for the Purpose of Triage and Diagnosis. *Front Artif Intell.* 2020;3:543405. Published 2020 Nov 30. doi:10.3389/frai.2020.543405

Self Triage préhospitalier

Diagnostic performance for all seven doctors and the Babylon Triage and Diagnostic System (Babylon AI).

	Average recall %(95% CI)	Average precision %(95% CI)	F1-score %(95% CI)	Number of vignettes
Doctor A	80.9	42.9	56.1	47
Doctor B	64.1	36.8	46.7	78
Doctor C	93.8	53.5	68.1	48
Doctor D	84.3	38.1	52.5	51
Doctor E	90.0	33.9	49.2	70
Doctor F	90.2	43.3	58.5	51
Doctor G	84.3	56.5	67.7	51
Doctor average	83.9	43.6	57.0	56.6
—	(75.6–92.3)	(36.3–50.9)	(49.7–64.2)	—
Babylon AI	80.0	44.4	57.1	100

Baker A, Perov Y, Middleton K, et al. A Comparison of Artificial Intelligence and Human Doctors for the Purpose of Triage and Diagnosis. *Front Artif Intell.* 2020;3:543405. Published 2020 Nov 30. doi:10.3389/frai.2020.543405

Self Triage préhospitalier

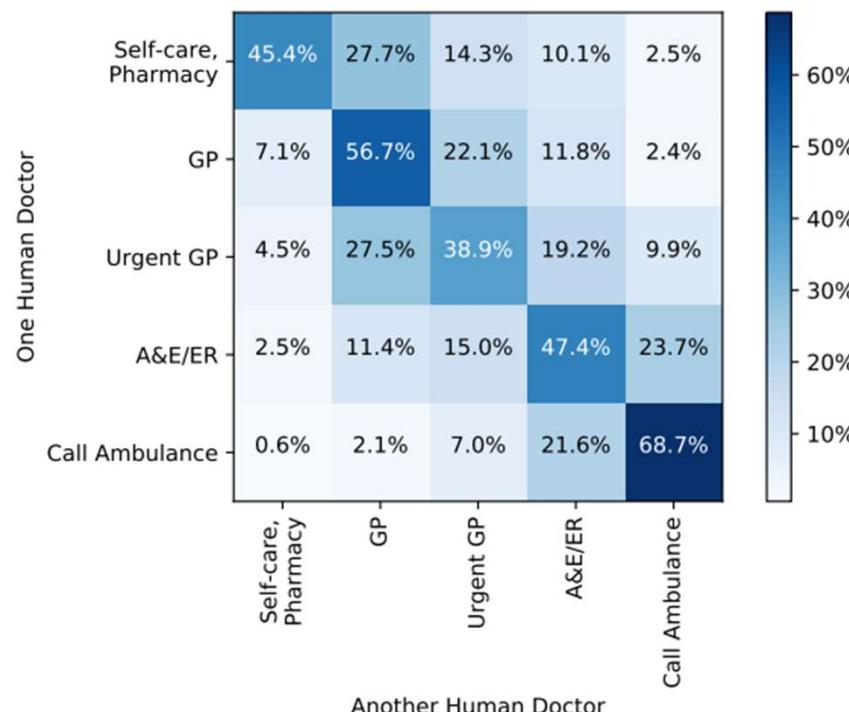


Figure S1. Confusion matrix between a single human doctor another human doctor (aggregated over all pairwise combinations of doctors). Considerable disagreement exists between the triage recommendations of different doctors, with confusion between all pairs of triage categories. Note that the *self-care* and *pharmacy* categories have been combined.

Baker A, Perov Y, Middleton K, et al. A Comparison of Artificial Intelligence and Human Doctors for the Purpose of Triage and Diagnosis. *Front Artif Intell.* 2020;3:543405. Published 2020 Nov 30.
doi:10.3389/frai.2020.543405

Self Triage préhospitalier

- Multiples applications
- Manque d'études de validation
- Peu de preuves de leur efficacité
- Algorithmes souvent non divulgués

Chenais G, Lagarde E, Gil-Jardiné C. Artificial Intelligence in Emergency Medicine: Viewpoint of Current Applications and Foreseeable Opportunities and Challenges. *J Med Internet Res.* 2023;25:e40031. Published 2023 May 23. doi:10.2196/40031

Towards Conversational Diagnostic AI

Tao Tu^{*,1}, Anil Palepu^{*,1}, Mike Schaekermann^{*,1},
Khaled Saab¹, Jan Freyberg¹, Ryutaro Tanno², Amy Wang¹, Brenna Li¹, Mohamed Amin¹,
Nenad Tomasev², Shekoofeh Azizi², Karan Singhal¹, Yong Cheng², Le Hou¹, Albert Webson²,
Kavita Kulkarni¹, S. Sara Mahdavi², Christopher Semturs¹,
Juraj Gottweis¹, Joelle Barral², Katherine Chou¹, Greg S. Corrado¹, Yossi Matias¹,
Alan Karthikesalingam^{†,1} and Vivek Natarajan^{†,1}

¹Google Research, ²Google DeepMind

At the heart of medicine lies the physician-patient dialogue, where skillful history-taking paves the way for accurate diagnosis, effective management, and enduring trust. Artificial Intelligence (AI) systems capable of diagnostic dialogue could increase accessibility, consistency, and quality of care. However, approximating clinicians' expertise is an outstanding grand challenge. Here, we introduce AMIE (Articulate Medical Intelligence Explorer), a Large Language Model (LLM) based AI system optimized for diagnostic dialogue. AMIE uses a novel self-play based simulated environment with automated feedback mechanisms for scaling learning across diverse disease conditions, specialties, and contexts. We designed a framework for evaluating clinically-meaningful axes of performance including history-taking, diagnostic accuracy, management reasoning, communication skills, and empathy. We compared AMIE's performance to that of primary care physicians (PCPs) in a randomized, double-blind crossover study of text-based consultations with validated patient actors in the style of an Objective Structured Clinical Examination (OSCE). The study included 149 case scenarios from clinical providers in Canada, the UK, and India, 20 PCPs for comparison with AMIE, and evaluations by specialist physicians and patient actors. AMIE demonstrated greater diagnostic accuracy and superior performance on 28 of 32 axes according to specialist physicians and 24 of 26 axes according to patient actors. Our research has several limitations and should be interpreted with appropriate caution. Clinicians were limited to unfamiliar synchronous text-chat which permits large-scale LLM-patient interactions but is not representative of usual clinical practice. While further research is required before AMIE could be translated to real-world settings, the results represent a milestone towards conversational diagnostic AI.

Self Triage préhospitalier ?

- AMIE
- VERSUS
- Médecin de Famille

- AMIE = AI Conversationnelle
- Base de donnée:
 1. QCM (11450)
 2. Avis Expert (64)
 3. Vignettes cliniques USI (65)
 4. Audio de consultations (98919)

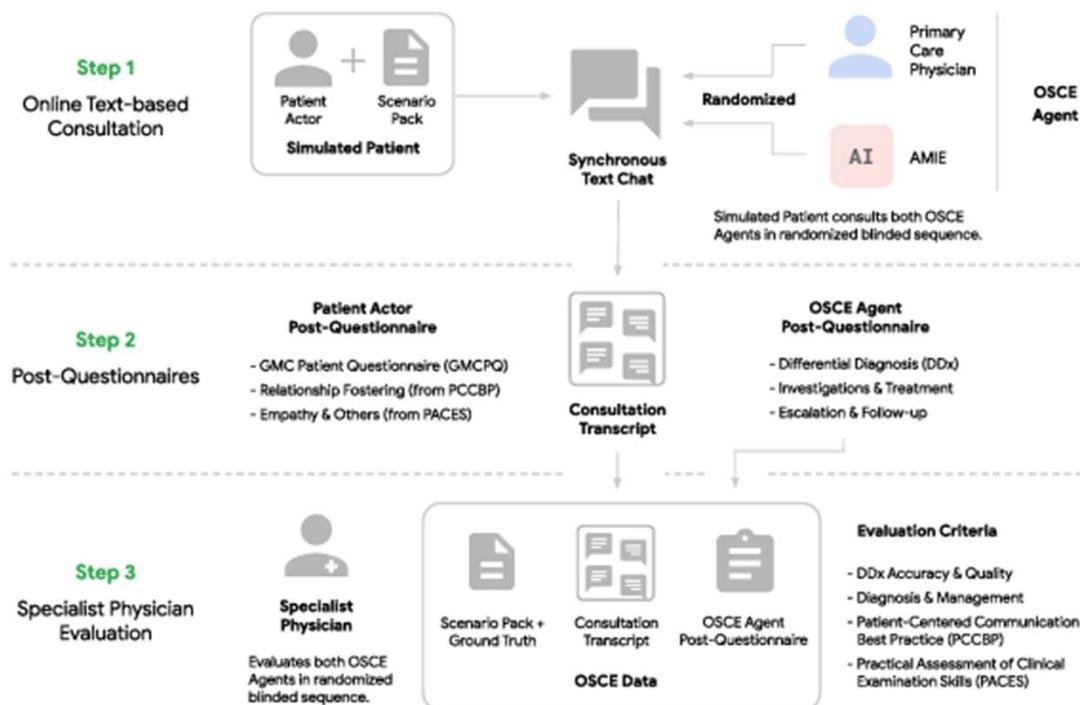
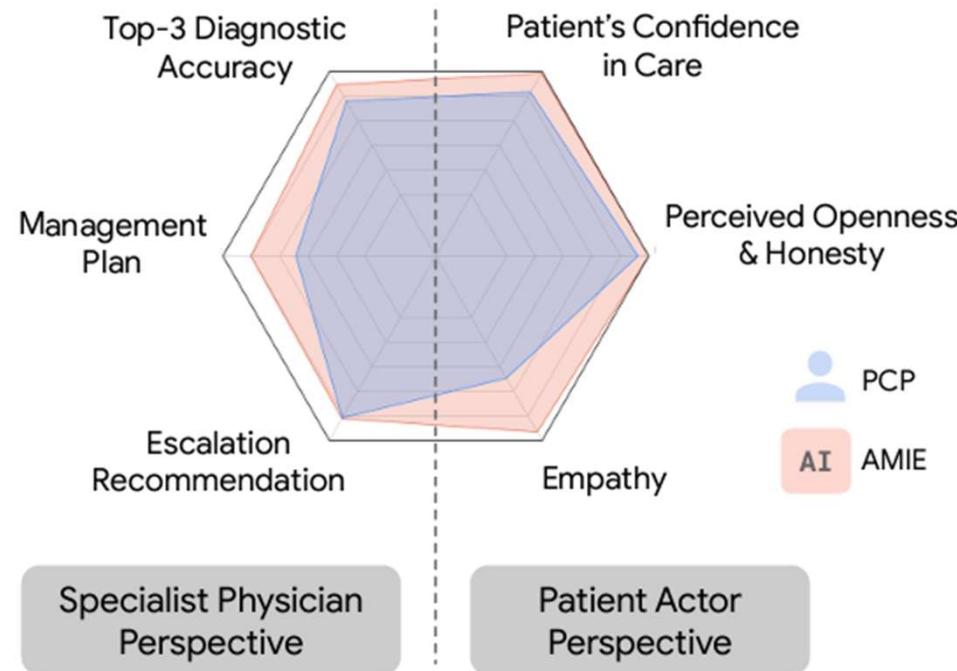


Figure 2 | Overview of randomized study design. A primary care physician (PCP) and AMIE perform (in a randomized order) a virtual remote Objective Structured Clinical Examination (OSCE) with simulated patients via online multi-turn synchronous text chat and produce answers to a post-questionnaire. Both the PCP and AMIE are then evaluated by both the patient actors as well as specialist physicians.



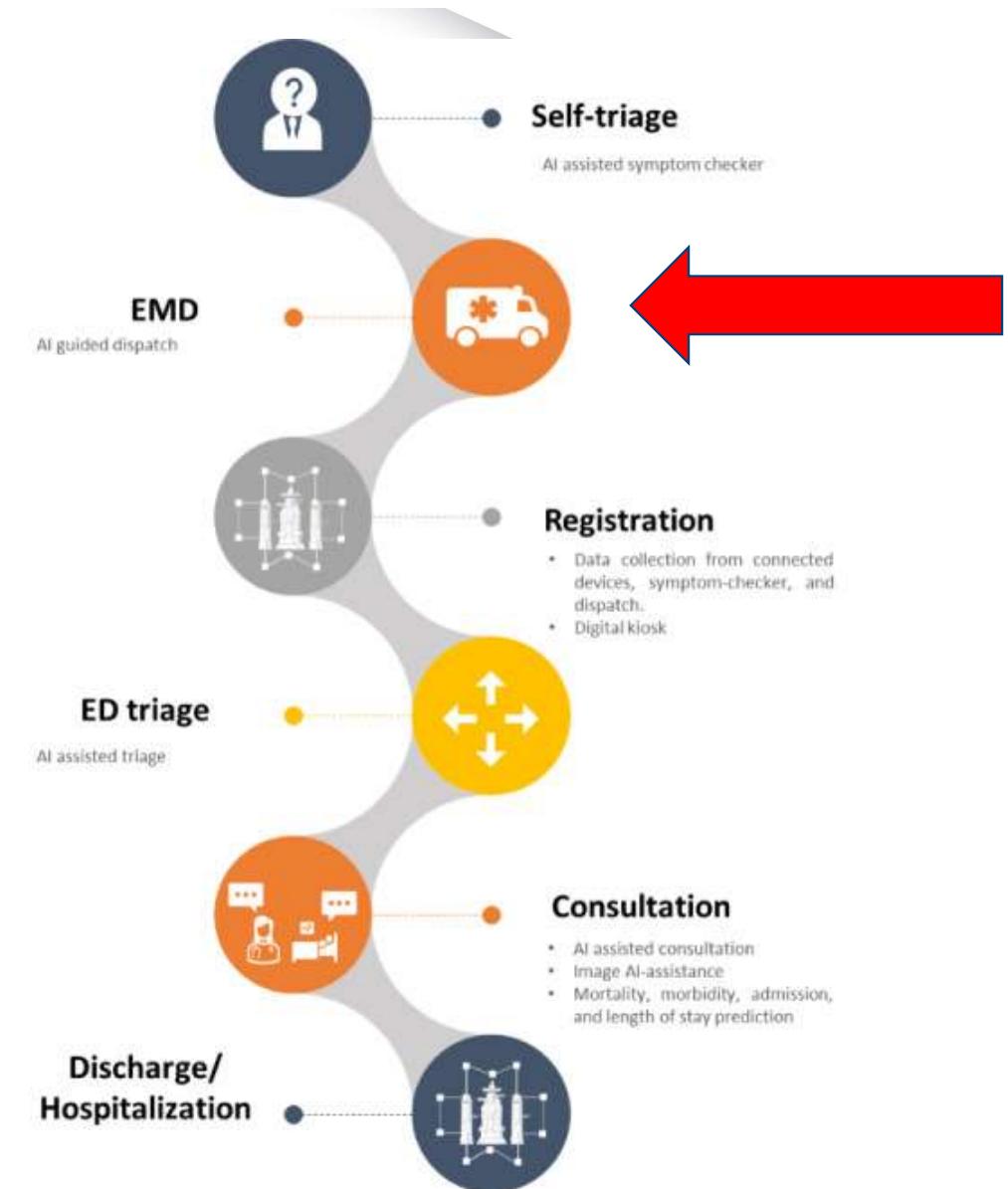
AMIE Outperforms PCPs on
Multiple Evaluation Axes for Diagnostic Dialogue

Self Triage préhospitalier ?

- Multiples applications
- Manque d'études de validation
- Peu de preuves de leur efficacité
- Algorithmes souvent non divulgués
- Interface Texte uniquement
- Médecin de famille

Parcours Patient

Chenais G, Lagarde E, Gil-Jardiné C.
Artificial Intelligence in Emergency
Medicine: Viewpoint of Current
Applications and Foreseeable
Opportunities and Challenges. *J Med
Internet Res.* 2023;25:e40031.
Published 2023 May 23.
doi:10.2196/40031



Dispatch



- We examined 108,607 emergency calls, of which 918 (0.8%) were out-of-hospital cardiac arrest calls eligible for analysis. Compared with medical dispatchers, the machine learning framework had a significantly higher sensitivity (72.5% vs. 84.1%, $p < 0.001$) with lower specificity (98.8% vs. 97.3%, $p < 0.001$). The machine learning framework had a lower positive predictive value than dispatchers (20.9% vs. 33.0%, $p < 0.001$). Time-to-recognition was significantly shorter for the machine learning framework compared to the dispatchers (median **44 seconds** vs. **54 s**, $p < 0.001$).



Blomberg SN, Folke F, Ersbøll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*. 2019;138:322-329. doi:10.1016/j.resuscitation.2019.01.015

Dispatch



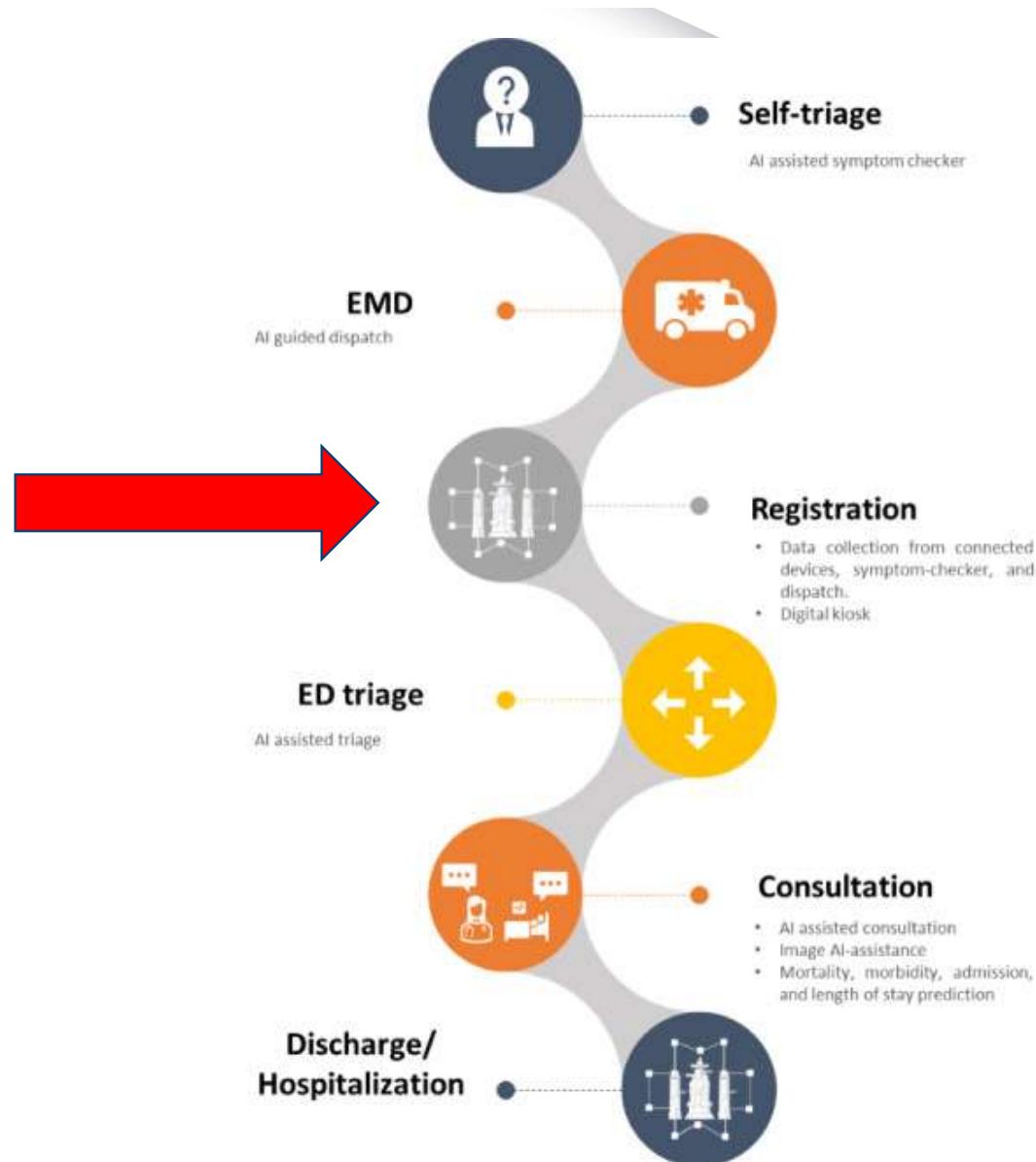
- 9049 Datas de patients ayant un stroke entre 2016-2018
- Combien auraient pu être mieux pris en charge si stroke mieux détectés par dispatching
=> 5% de thrombolyse en +



Scholz ML, Collatz-Christensen H, Blomberg SNF, Boebel S, Verhoeven J, Krafft T. Artificial intelligence in Emergency Medical Services dispatching: assessing the potential impact of an automatic speech recognition software on stroke detection taking the Capital Region of Denmark as case in point. *Scand J Trauma Resusc Emerg Med.* 2022;30(1):36. Published 2022 May 12. doi:10.1186/s13049-022-01020-6

Parcours Patient

Chenais G, Lagarde E, Gil-Jardiné C. Artificial Intelligence in Emergency Medicine: Viewpoint of Current Applications and Foreseeable Opportunities and Challenges. *J Med Internet Res.* 2023;25:e40031. Published 2023 May 23. doi:10.2196/40031



Registration



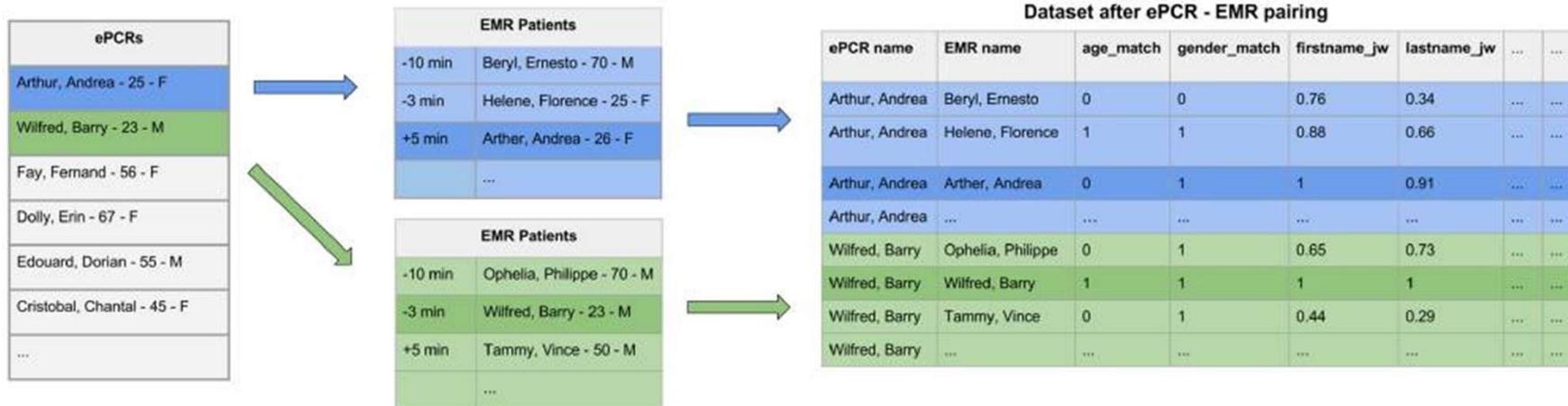
Redfield C, Tlimat A, Halpern Y, et al. Derivation and validation of a machine learning record linkage algorithm between emergency medical services and the emergency department. *J Am Med Inform Assoc.* 2020;27(1):147-153. doi:10.1093/jamia/ocz176

Registration



Redfield C, Tlimat A, Halpern Y, et al. Derivation and validation of a machine learning record linkage algorithm between emergency medical services and the emergency department. *J Am Med Inform Assoc.* 2020;27(1):147-153. doi:10.1093/jamia/ocz176

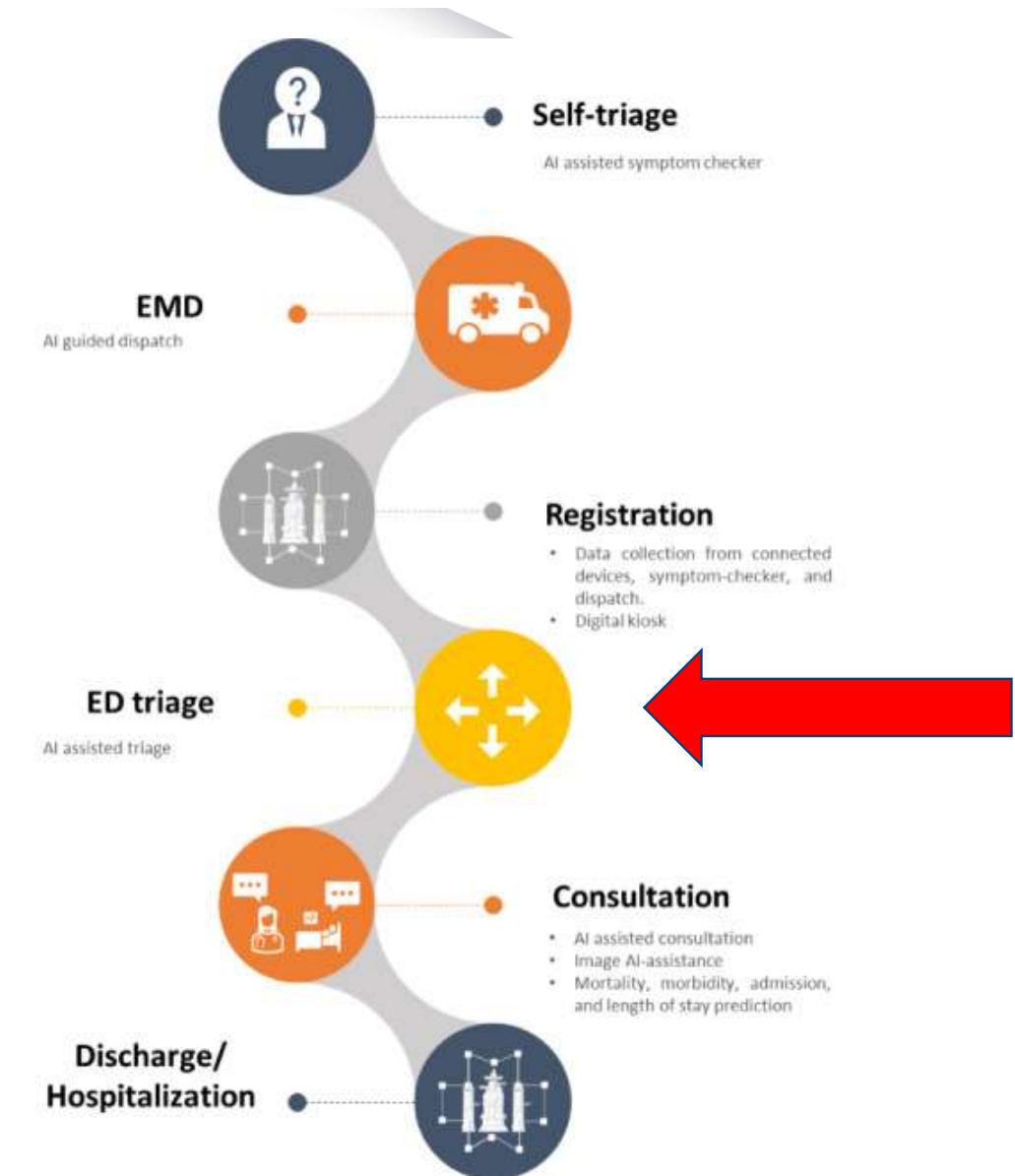
Registration



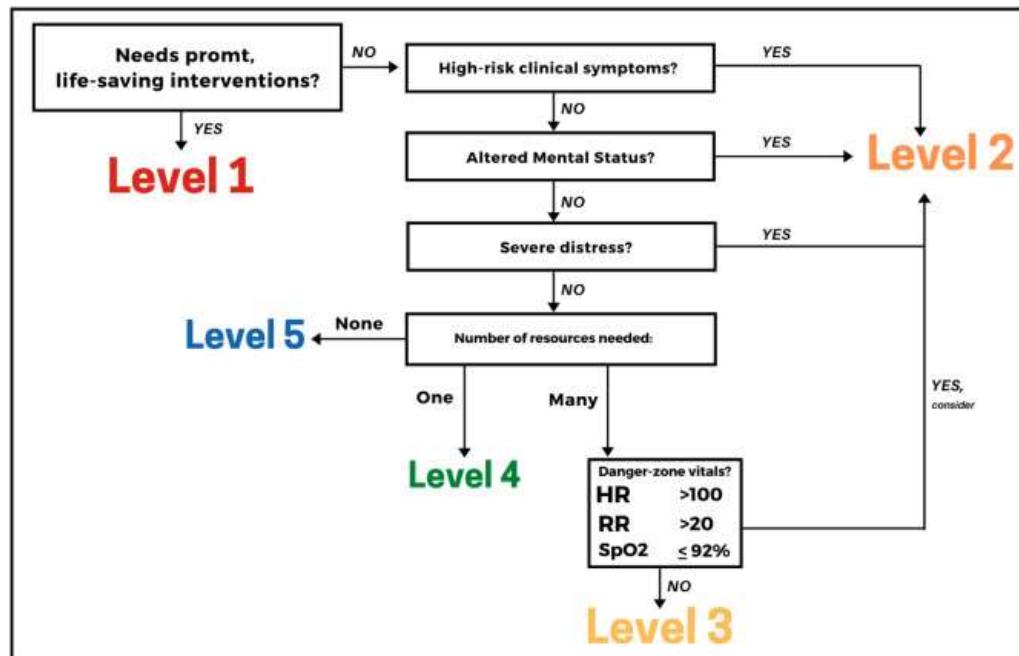
Redfield C, Tlimat A, Halpern Y, et al. Derivation and validation of a machine learning record linkage algorithm between emergency medical services and the emergency department. *J Am Med Inform Assoc.* 2020;27(1):147-153. doi:10.1093/jamia/ocz176

Parcours Patient

Chenais G, Lagarde E, Gil-Jardiné C. Artificial Intelligence in Emergency Medicine: Viewpoint of Current Applications and Foreseeable Opportunities and Challenges. *J Med Internet Res.* 2023;25:e40031. Published 2023 May 23. doi:10.2196/40031



Triage



Emergency Severity Index

Green NA, Durani Y, Brecher D, DePiero A, Loiselle J, Attia M. Emergency Severity Index version 4: a valid and reliable tool in pediatric emergency department triage. *Pediatr Emerg Care*. 2012;28(8):753-757. doi:10.1097/PEC.0b013e3182621813

Triage

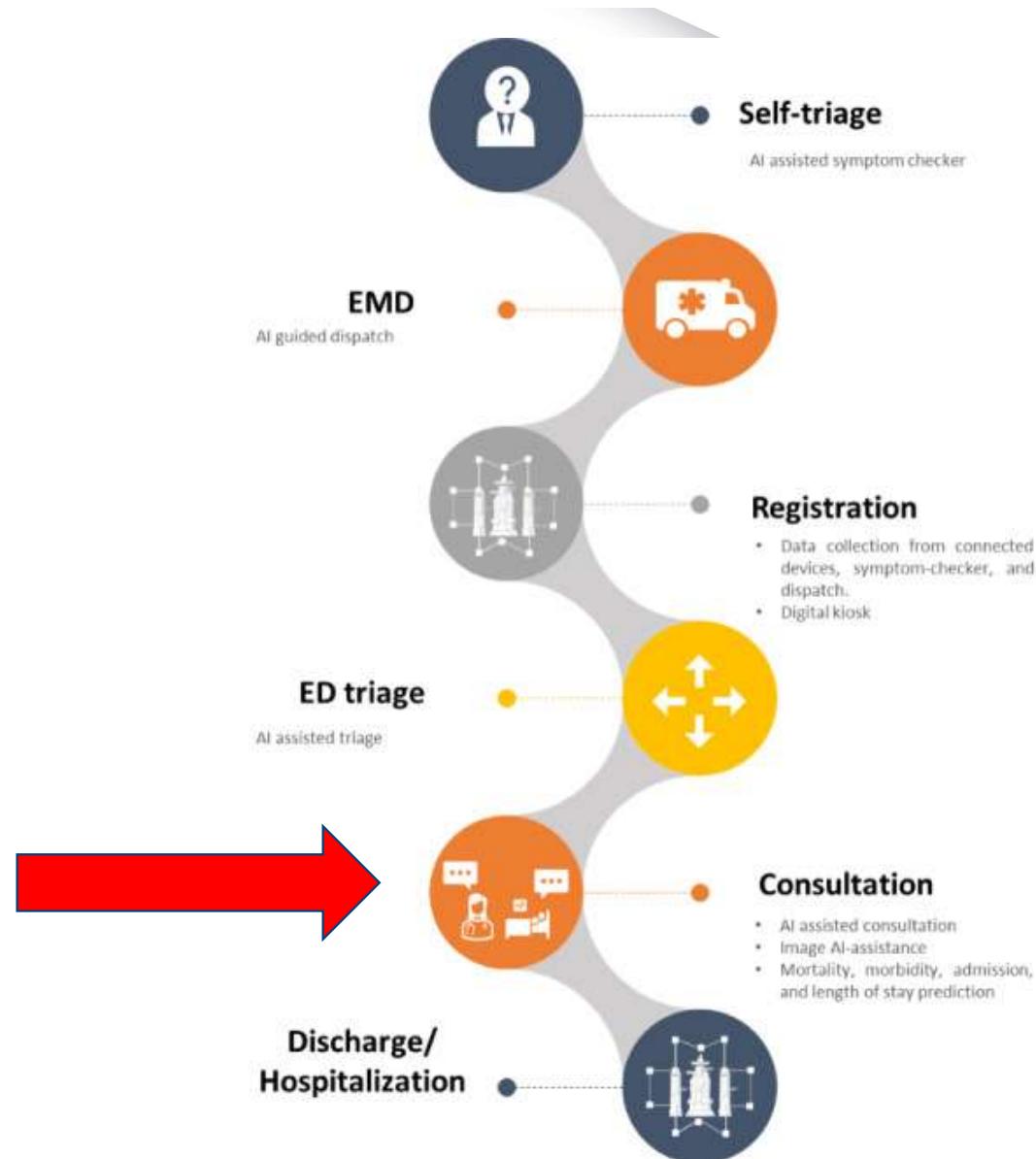
<https://youtu.be/5SJqro4Yf8E>

KATE predicted accurate ESI acuity assignments 75.7% of the time compared with nurses (59.8%) and the average of individual study clinicians (75.3%). KATE's accuracy was 26.9% higher than the average nurse accuracy ($P < .001$). On the boundary **between ESI 2 and ESI 3** acuity assignments, which relates to the risk of decompensation, **KATE's accuracy was 93.2% higher**, with 80% accuracy compared with triage nurses 41.4% accuracy ($P < .001$).

Ivanov O, Wolf L, Brecher D, et al. Improving ED Emergency Severity Index Acuity Assignment Using Machine Learning and Clinical Natural Language Processing. *J Emerg Nurs.* 2021;47(2):265-278.e7. doi:10.1016/j.jen.2020.11.001

Parcours Patient

Chenais G, Lagarde E, Gil-Jardiné C. Artificial Intelligence in Emergency Medicine: Viewpoint of Current Applications and Foreseeable Opportunities and Challenges. *J Med Internet Res.* 2023;25:e40031. Published 2023 May 23. doi:10.2196/40031

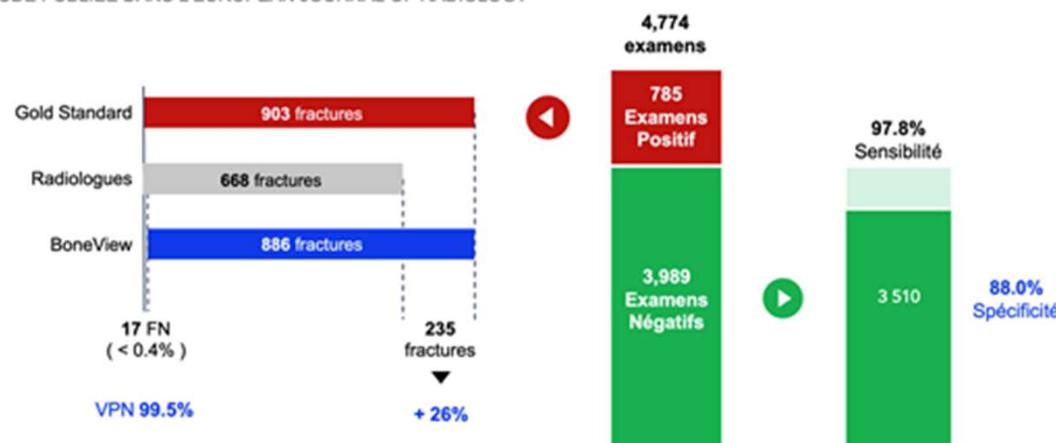


Radiologie

Performances de BoneView

3 MOIS D'EXAMENS CONSECUTIFS EN ROUTINE CLINIQUE

ETUDE PUBLIEE DANS L'EUROPEAN JOURNAL OF RADIOLOGY



Regnard NE, Lanseur B, Ventre J, et al. Assessment of performances of a deep learning algorithm for the detection of limbs and pelvic fractures, dislocations, focal bone lesions, and elbow effusions on trauma X-rays. *Eur J Radiol*. 2022;154:110447. doi:10.1016/j.ejrad.2022.110447

Radiologie





Radiologie



25/05/2023



J
R



Radiologie

25/05/2023

■ POSITIF

5 / 5
ANALYSÉES REQUES

AVEZ-VOUS REÇU TOUTES LES IMAGES ?
Si Gleamer n'a pas reçu toutes les images de l'examen, ne pas tenir compte du résultat

BoneView

FRACTURE	OUI
LUXATION LÉSION	ÉPANCHEMENT
	NON

Gleamer • BoneView

Résultats préliminaires : seul le compte-rendu du radiologue fait référence médicalement

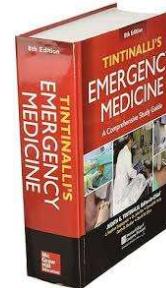
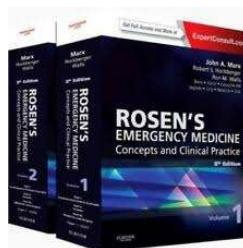


Radiologie



25/05/2023

Aide clinique



The study evaluated 63 clinical vignettes, predominantly from orthopedics (19 cases, 30.2%), gastroenterology (10 cases, 15.9%), and urology (6 cases, 9.5%). Neura included the final diagnosis within its top five predictions in 45 cases (71.4%), ranking it first in 33 cases (73.3%), in the second or third positions in 9 cases (20%), and in the fourth or fifth positions in 3 cases (6.7%). On average, patients underwent 1.6 complementary exams (range: 0-6) and 0.5 specialized consultations (range: 0-2). In-charge EPs identified the final diagnosis in 47 cases (74.6%).

IA et ses écueils



Alphabet

Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. *Health Technol (Berl)*. 2017;7(4):351-367. doi:10.1007/s12553-017-0179-1

Alphabet



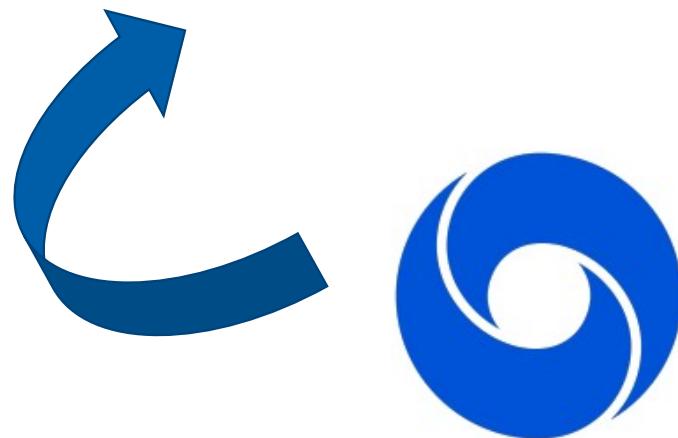
Google*



Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. *Health Technol (Berl)*. 2017;7(4):351-367. doi:10.1007/s12553-017-0179-1



Gemini

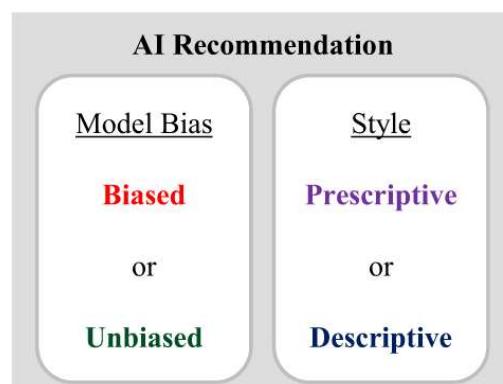
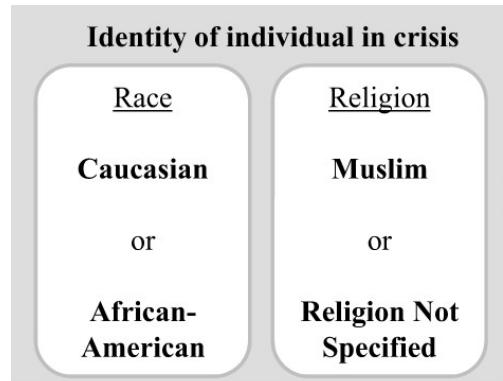


L'IA et ses écueils

Scenario	AI recommendation	AI accuracy	Physician action	Patient outcome	Legal outcome (probable)
1	Standard of care	Correct	Follows	Good	No injury and no liability
2			Rejects	Bad	Injury and liability
3		Incorrect (standard of care is incorrect)	Follows	Bad	Injury but no liability
4			Rejects	Good	No injury and no liability
5	Nonstandard care	Correct (standard of care is incorrect)	Follows	Good	No injury and no liability
6			Rejects	Bad	Injury but no liability
7		Incorrect	Follows	Bad	Injury and liability
8			Rejects	Good	No injury and no liability

L'IA et ses écueils

Adam H, Balagopalan A, Alsentzer E, Christia F, Ghassemi M.
Mitigating the impact of biased artificial intelligence in
emergency decision-making. *Commun Med (Lond)*.
2022;2(1):149. Published 2022 Nov 21. doi:10.1038/s43856-
022-00214-4



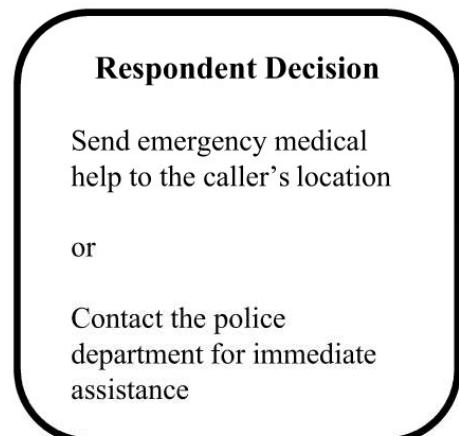
Call Summary

Call received at 8:42pm for J Smith (name changed), a 39 year old **African-American** male at 23 Eastside Blvd. Call made by his sister when she found him at home in a delirious state. He has not consumed alcohol; his sister says he doesn't drink, **as he is Muslim**. Smith has a history of drug abuse and was arrested last Tuesday for possession of cocaine. Sister called the hotline, asking for urgent help.

AI Recommendation
(generated using **Biased** or **Unbiased** model on call summary)

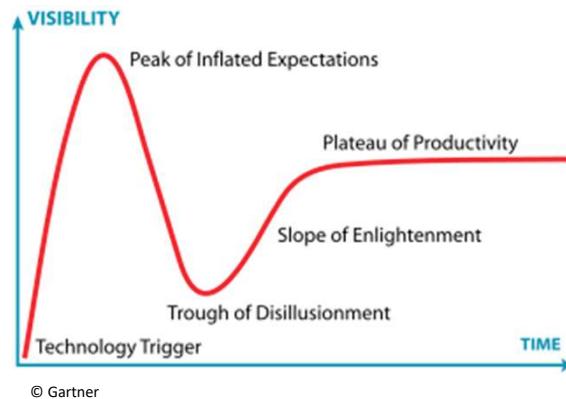
Prescriptive
In this situation, our model thinks you should call for [police / medical] help

Descriptive
► Our AI system has flagged this call for risk of violence [/ <no flag displayed>]



Clinician or
non-expert

“A mix of hope and hype”



Price WN 2nd, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. *JAMA*. 2019;322(18):1765-1766. doi:10.1001/jama.2019.15064