Uporaba umetne inteligence pri diagnostiki sepse v urgentni ambulanti

-

prednost ali past?

Žiga Martinčič et AI
Klinika za infekcijske bolezni in vročinska stanja
UKC Ljubljana
Strokovni dogodek ob svetovnem dnevu sepse, 11.9.2025

Uporabljate umetno inteligenco?

- v vsakdanjem življenju?
- za pisanje poročil / predavanj / tekstov?
- za zdravljenje svojih bolnikov?

Slika iz osebnega arhiva predavatelja (ustvarjeno s Canva AI, https://www.canva.com/dream-lab)

Prikaz primera

- 44-letna učiteljica
  - po poškodbi obraza s poko frontalne kosti in sinusov pred 20 leti
  - po histerektomiji
  - brez redne th; brez alergij

- v soboto zjutraj zbudila z glavobolom
- vzela analgetik in šla v trgovino, glava še bolela
- ob odhodu iz trgovine jo je začela motiti svetloba, sililo jo je na bruhanje,
- doma občutek presinkope, mrzlica in vročina, zmedenost

- svoje težave vpisala v ChatGPT, ki ji je podal Ddg. meningitisa -> sinu je rekla, naj kliče reševalce

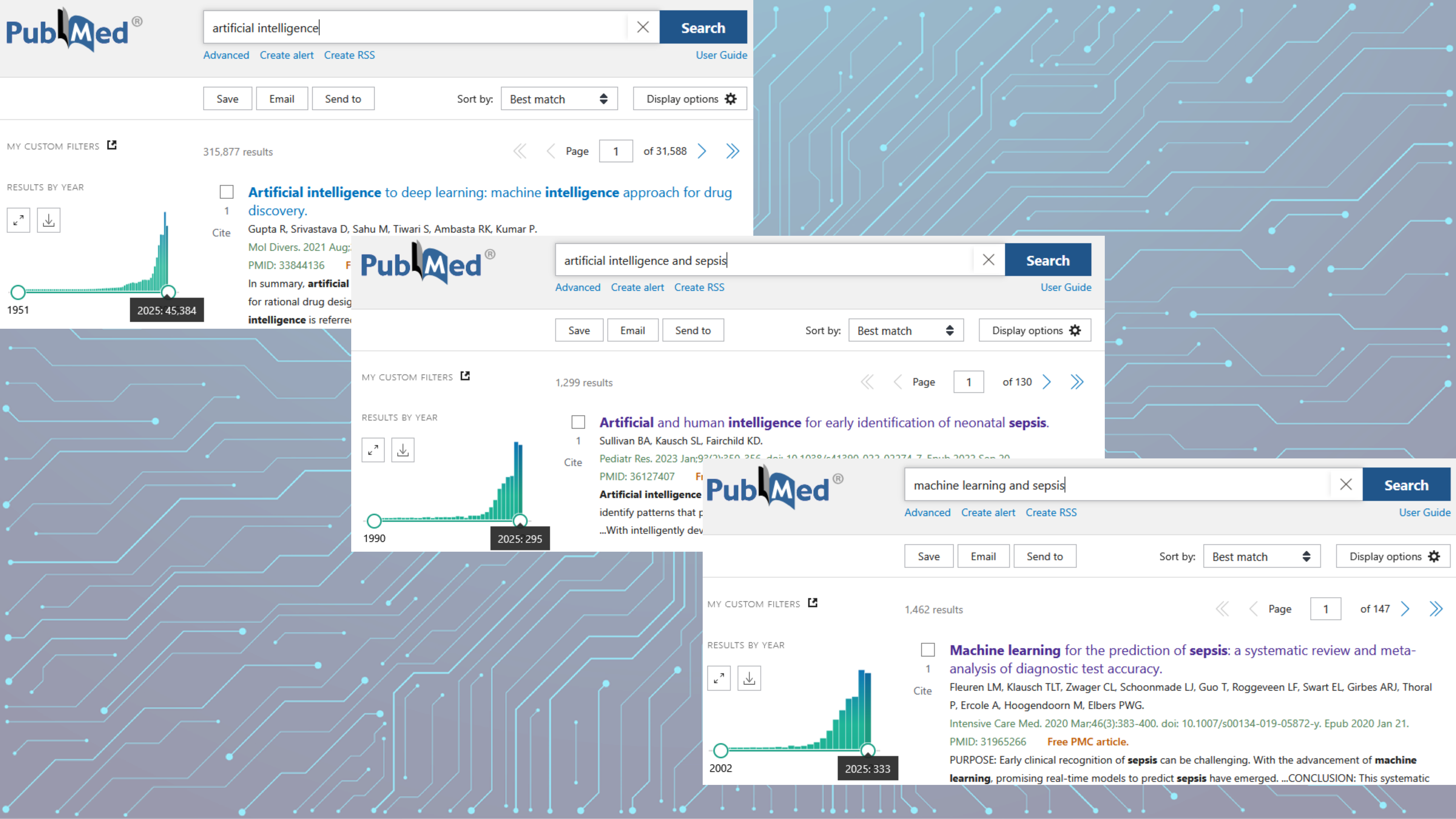
Slika iz osebnega arhiva predavatelja (ustvarjeno s Canva AI, https://www.canva.com/dream-lab)

Prikaz primera

- v SB opravili lab. preiskave
  - CRP 46, PCT 5.25, sečnina 3.6, kreatinin 55, Lkc 13.1, Hb 136, Trc 168
  - in lumbalno punkcijo: Lkc 9.791 (nevtro 9.125, limfo 666), proteini 4.6, glukoza 0.1, LDH 0.63, laktat 11
- empirična antibiotična terapija + deksametazon

- premestitev na KIBVS
  - ob sprejemu budna, smiselno kontaktna, sodelujoča, krajevno in časovno orientirana, anamnezo podaja sama, koherentno, včerajšnjega dogodka se v celoti spominja
- Likvor: *S. pneumoniae*
  - prilagoditev antibiotične terapije
  - radiološka dg (posttravmatska meningokela z bočenjem v desni etmoid)
  - nadaljnja lab. diagnostika:
pozitiven beta-trace protein iz nosne tekočine - likvoreja
  - pregled ORL, nadaljnja obravnava
- po zaključenem zdravljenju domov, dober nevrološki izhod

Slika iz osebnega arhiva predavatelja (ustvarjeno s Canva AI, https://www.canva.com/dream-lab)


PubMed
artificial intelligence
Search
Advanced Create alert Create RSS
User Guide
Save Email Send to
Sort by: Best match
Display options
MY CUSTOM FILTERS
315,877 results
Page 1 of 31,588
RESULTS BY YEAR
Artificial intelligence to deep learning: machine intelligence approach for drug discovery.
Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P.
Mol Divers. 2021 Aug;
PMID: 33844136
1951
2025: 45,384
Cite
PubMed
artificial intelligence and sepsis
Search
Advanced Create alert Create RSS
User Guide
Save Email Send to
Sort by: Best match
Display options
MY CUSTOM FILTERS
1,299 results
Page 1 of 130
RESULTS BY YEAR
Artificial and human intelligence for early identification of neonatal sepsis.
Sullivan BA, Kausch SL, Fairchild KD.
Pediatr Res. 2023 Jan;
PMID: 36127407
1990
2025: 295
Cite
PubMed
machine learning and sepsis
Search
Advanced Create alert Create RSS
User Guide
Save Email Send to
Sort by: Best match
Display options
MY CUSTOM FILTERS
1,462 results
Page 1 of 147
RESULTS BY YEAR
Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy.
Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, Swart EL, Girbes ARJ, Thoral P, Ercole A, Hoogendoorn M, Elbers PWG.
Intensive Care Med. 2020 Mar;46(3):383-400. doi: 10.1007/s00134-019-05872-y. Epub 2020 Jan 21.
PMID: 31965266 Free PMC article.
PURPOSE: Early clinical recognition of sepsis can be challenging. With the advancement of machine learning, promising real-time models to predict sepsis have emerged. ...CONCLUSION: This systematic
2002
2025: 333
Cite

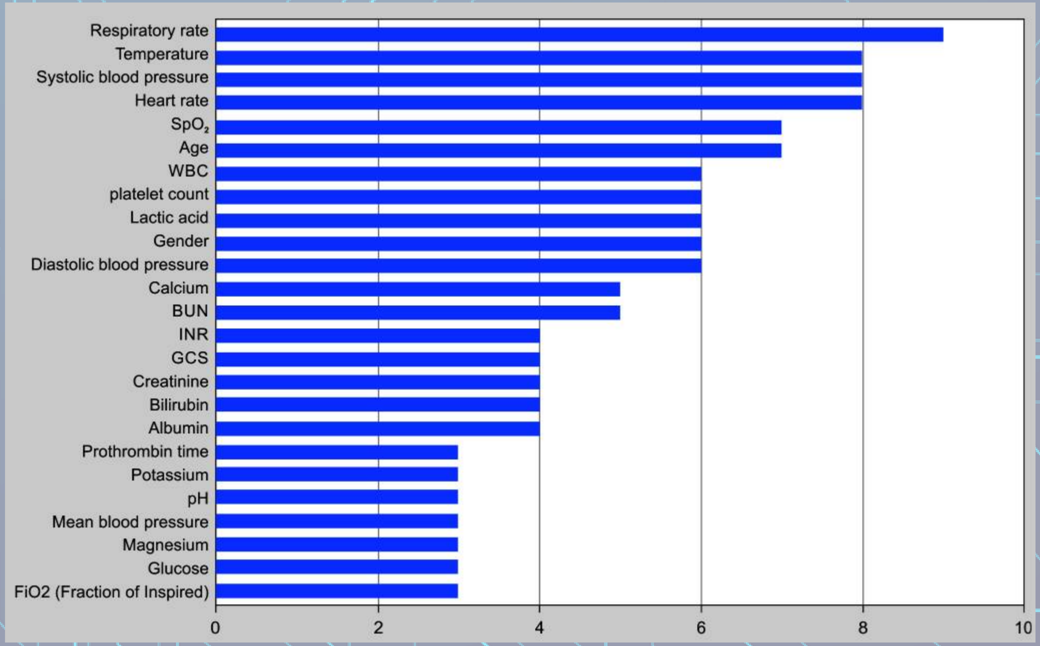
Modeli za zgodnjo prepoznavo sepse

REVIEW ARTICLE

Machine Learning and Deep Learning Models for Early Sepsis Prediction: A Scoping Review

- Pregled modelov UI v jan. 2022 – mar. 2025
- 47 objav, vključenih 13
- Uporabljeni modeli UI
  - 7 izključno nadzorovano strojno učenje
  - 2 nadzorovano in nenadzorovano SU
  - 2 globoko učenje (‚deep learning', GU)
  - 1 nadzorovano SU in GU
  - 1 vse tri oblike
- Baze podatkov za učenje
  - 6 uporabilo MIMIC
  - 7 lokalne baze podatkov
- Večinoma demografski, metrični in lab. podatki

- 1 raziskava – uporaba EKG telemetrije; podaljšanje QRS ob razvoju sepse (4h pred klinično zaznavo)



- *nenadzorovano SU in GU (NN) tudi ‚nekonvencionalne' parametre - RDW

Shanmugam H et al. Indian J Crit Care Med. 2025

		year							specificity)		
1.	Real-time prediction of sepsis in critical trauma patients[17]	2023	Predict hourly sepsis risk in ICU-admitted trauma patients	MIMIC-IV 1.0 database (2008–2019), 4603 trauma ICU patients	ICU trauma patients	XGBoost, logistic regression	Forty-two raw variables (demographics) + 485 derived features	AUROC	AUROC: 0.83–0.88	XGBoost performed better than logistic regression. ML enables early detection and better discrimination of sepsis in trauma patients	The injury severity score was not used; it requires further validation
2.	Machine learning for early prediction of sepsis in ICU patients[18]	2023	Early detection of sepsis in ICU patients	BESTCare database at KAMC (1,182 ICU patients)	ICU patients, aged 14 or older	Regression model called survival analysis	Lactic acid, temperature, and time to outcome	*p*-value significance testing	*p*-value for time = 0.000568, *p*-value for lactic acid = 0.01, *p*-value for temp = 0.02	Sepsis can be predicted using these three essential factors.	Limited data preprocessing and a short period
3.	ML models for sepsis prediction in acute pancreatitis[19]	2023	Comparing the logistic regression model with traditional scoring systems	MIMIC-III and IV (1,672 AP patients)	Patients admitted with acute pancreatitis	GBDT, SVM, KNN, MLP, LR, AdaBoost	Age, GC score, HR, RR, SpO2, temp, INR, BUN, RDW	AUC, PPV, NPV, Sensitivity, Specificity, accuracy	GBDT AUC: 0.985; qSOFA AUC 0.780	GBDT outperformed traditional models in detecting sepsis in acute pancreatitis patients	Lacks generalizability to other populations, lacks radiological data, lacks
4.	Deep learning-based sepsis and septic shock early prediction (DeepSEPS)[20]	2023	Predict sepsis and septic shock in ICU using DL	EMR (2010–2021)	ICU	Deep learning (RNN/CNN)	Eight vital signs, 13 lab results, and three demographics	AUROC	AUROC achieved by DeepSEPS: 0.79 (sepsis), 0.85 (septic shock)	Deep learning performed better than the SOFA score	Requires real-world validation
5.	ML model for early sepsis detection integrated in clinical workflow[21]	2023	Early detection of sepsis using ML	Electronic medical records EMR	EMR of patients 18 years or older and admitted as inpatients at one of Virtua Health's five hospitals	XGBoost + SHAP for explainability analysis	Eighty-six clinical variables from EMR	Sensitivity, specificity, FPR	Sensitivity: 91%, Specificity: 94%, false positive rate 6%	Reduced false positives compared to the currently deployed commercial model	Generalizability of the model, real-world
6.	Prediction of sepsis within 24 hours at ED triage[22]	2024	Prediction of sepsis within 24h in ED using ML	MIMIC-IV (14,957 ED samples)	Emergency department	XGBoost	Demographics, acuity, vital signs	AUROC, accuracy	AUROC: 0.92, accuracy:84.1%	Early prediction of sepsis in the ED using only a small amount of data	False alarms, a large number of false positives
7.	Time-series deep learning for sepsis prediction in non-ICU[23]	2024	Predict sepsis in non-ICU settings	MIMIC-IV (83,813 non-ICU patients)	Non-ICU	Deep learning (RNN, transformer)	Laboratory data	AUROC	AUROC: 0.96–0.99	Fifty-seven percent reduction in false alarms	Depending on laboratory data, the lack of real-world
8.	Interpretable ML for sepsis risk in emergency triage[24]	2025	Predict sepsis in emergency triage using EMR data	MIMIC-IV (1,89,617 patients included)	Emergency department	Gradient boosting, RF, SVM, SHAP and LIME	Vital signs, demographics, chief complaints	AUROC, F1 score	Best AUROC: 0.83	More effective sepsis prediction using comprehensive triage information than solely using vital signs	Needs real-time validation
9.	ML model for early sepsis detection in all hospital inpatients[25]	2025	Early sepsis detection across all departments	Valenciennes hospital (45,127 patients)	ICU, ED, surgical, and all hospital departments	Gradient boosting (SEPSI score)	Vital signs, laboratory data, GCS score	AUROC, PPV, sensitivity	AUROC: 0.738, sensitivity: 84.5%; PPV: 0.610	Early detection of sepsis; half of the cases were detected 48 hr before onset	Needs real-world validation
10.	Clinical validation and optimization of ML models for sepsis[26]	2025	Compare different ML models for early sepsis prediction	Tertiary hospital in China (2,329 patients)	Non-ICU hospital admissions	RF, logistic regression, decision tree, Light GBM	Thirty-six clinical features	AUROC, sensitivity, specificity, accuracy, F1 score	RF AUROC: 0.818; F1 value 0.38; sensitivity 0.746	RF model predicted sepsis earliest among all	More patient data to make the findings more effective
11.	Fusion of Fully Integrated analog ML classifier with EMR for real-time sepsis prediction[27]	2022	Real-time prediction of sepsis using integrated AI model	Emory healthcare system (965 patients: 514 sepsis and 451 non-sepsis patients)	Real-time continuous At-home surveillance	On-chip analog neural network (ANN), random forest, SVM	ECG, demographics, comorbidities	AUROC, accuracy	AUROC: 0.99 (before 4 hr), Accuracy: 94%	Best prediction performance of 4 hr or less	Limited to one institution
12.	Identifying the risk of sepsis in patients with cancer using digital health records[28]	2022	Predict sepsis in cancer patients using EHRs	Samsung Medical Center, Korea (8,580 cancer patients) between 2014 and 2019	Cancer patients who visited the emergency room within 5 years of being diagnosed with cancer	Random forest, logistic regression, ANN, ResNet, LSTM	Lab test results and medication prescriptions	AUROC, F1 score	AUROC: 0.753, F1 score: 0.602, accuracy: 0.692	The random forest-based model Demonstrated the best performance. Incorporating lab tests and medications contributed to the accurate prediction	Partially accurate health reco for research
13.	FedSepsis: federated multi-modal DL for early sepsis detection[29]	2023	Improve early sepsis detection using ML	MIMIC-III (patients older than 15 years)	Non-ICU	FedSepsis: federated DL model	Demographics, vitals, and lab tests	AUROC, precision-recall curve	Best AUPRC: 99.36%	Incorporating multimodality provided a superior performance	Interpretability, high-cost device usage

Feedback

Shanmugam H et al. Indian J Crit Care Med. 2025

Prediction of sepsis within 24 hours at ED triage[22]	2024	Prediction of sepsis within 24h in ED using ML	AUROC: 0.92, accuracy:84.1%	Early prediction of sepsis in the ED using only a small amount of data	False alarms, a large number of false positives
Time-series deep learning for sepsis prediction in non-ICU[23]	2024	Predict sepsis in non-ICU settings	AUROC: 0.96–0.99	Fifty-seven percent reduction in false alarms	Depending on laboratory data, the lack of real-world
Interpretable ML for sepsis risk in emergency triage[24]	2025	Predict sepsis in emergency triage using EMR data	Best AUROC: 0.83	More effective sepsis prediction using comprehensive triage information than solely using vital signs	Needs real-time validation
ML model for early sepsis detection in all hospital inpatients[25]	2025	Early sepsis detection across all departments	AUROC: 0.738, sensitivity: 84.5%; PPV: 0.610	Early detection of sepsis; half of the cases were detected 48 hr before onset	Needs real-world validation
Fusion of Fully Integrated analog ML classifier with EMR for real-time sepsis prediction[27]	2022	Real-time prediction of sepsis using integrated AI model	AUROC: 0.99 (before 4 hr), Accuracy: 94%	Best prediction performance of 4 hr or less	

Shanmugam H et al. Indian J Crit Care Med. 2025

Modeli za zgodnjo prepoznavo sepse

REVIEW ARTICLE

Machine Learning and Deep Learning Models for Early Sepsis Prediction: A Scoping Review

- Različni modeli UI dosegali AUROC 0.73 – 0.99, občutljivost 0.7 – 0.9, specifičnost ~0.9, F-1 0.3-0.6
- Visoka stopnja lažno pozitivnih alarmov – 9 na 1 resnično pozitivnega !!!
- Boljše rezultate dosegali modeli globokega učenja (4/13)
- Ni podatkov uporabe v resničnem svetu

Shanmugam H et al. Indian J Crit Care Med. 2025

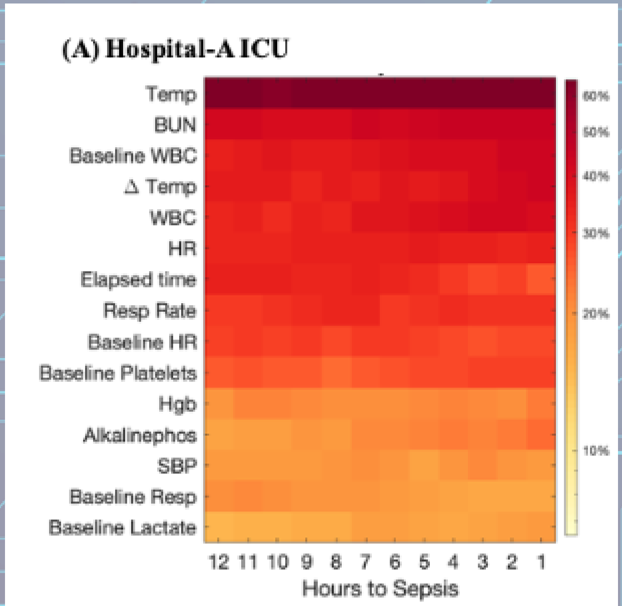
Primer modela za zgodnjo prepoznavo sepse - COMPOSER

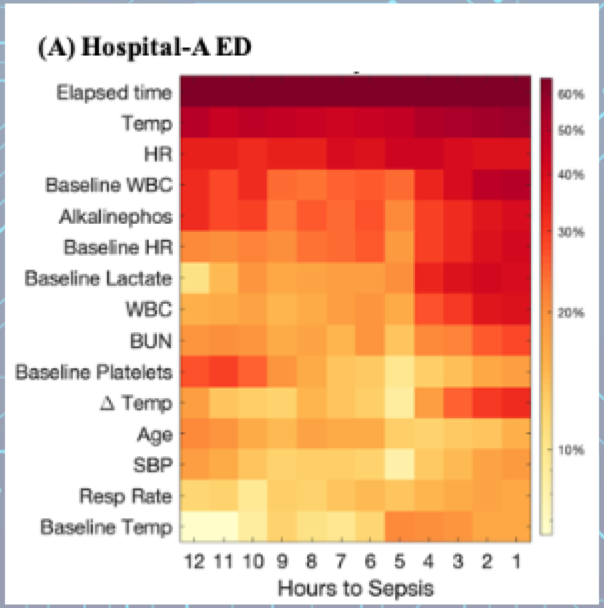
npj Digital Medicine www.nature.com/npjdigitalmed

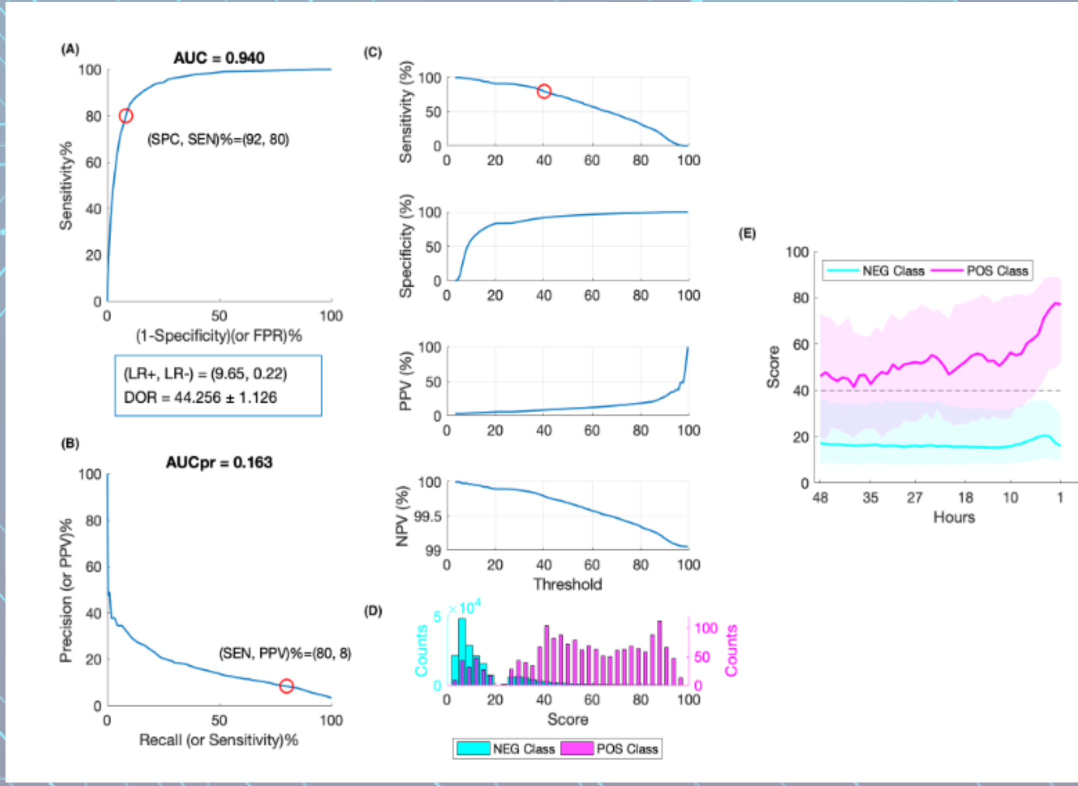
ARTICLE OPEN Check for updates

Artificial intelligence sepsis prediction algorithm learns to say "I don't know"

- Podatki iz 6 kohort bolnikov v EIT in UC; 515.720 bolnikov
- 40 spremenljivk (strukturirani podatki: demografske značilnosti, vitalni znaki, laboratorijski izvidi)
- AUC/PNV 0.95/38% (EIT) oz. 0.945/20% (UC)
- Prepoznava sepse 4 – 48 ur pred klinično prepoznavo



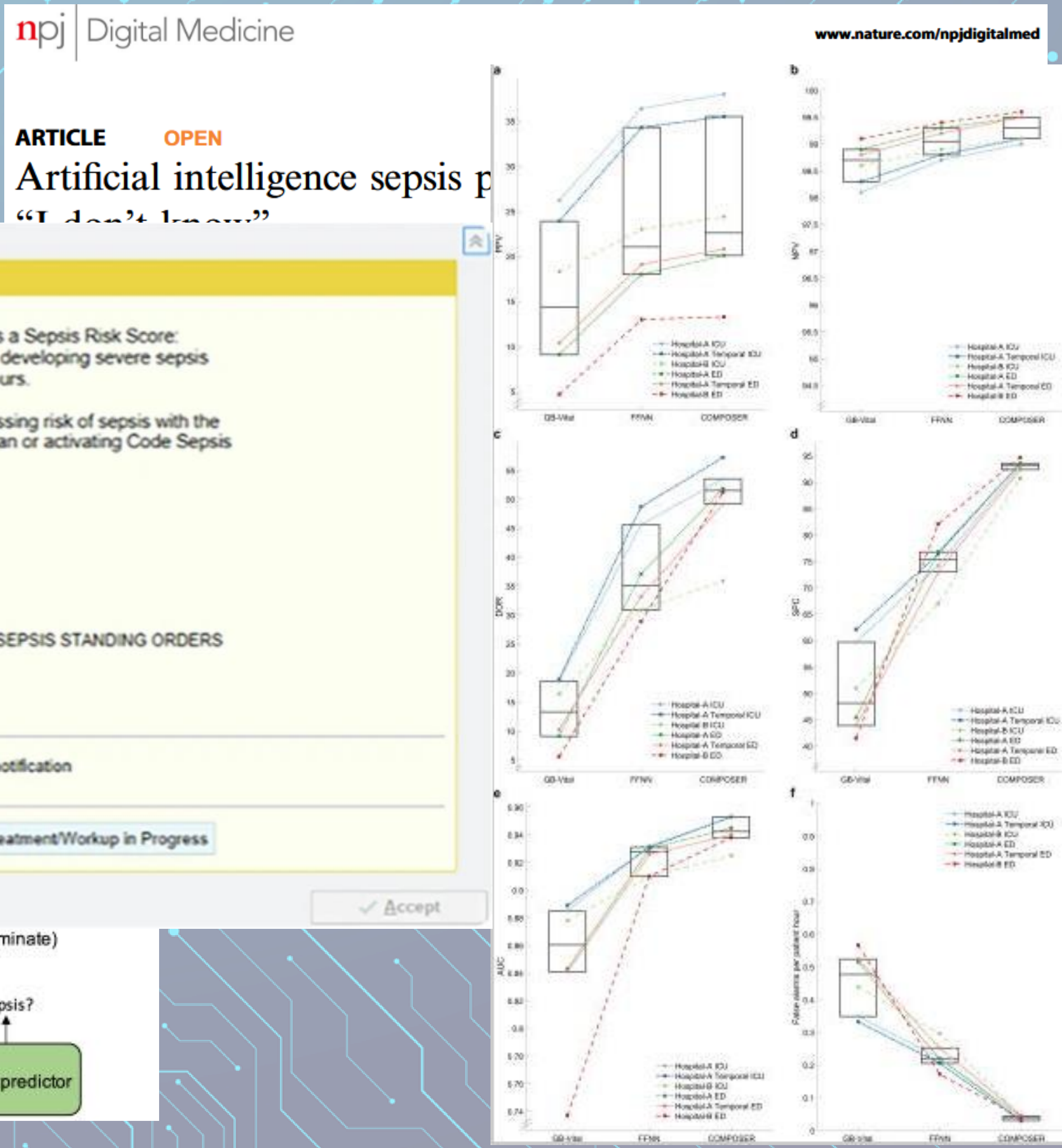


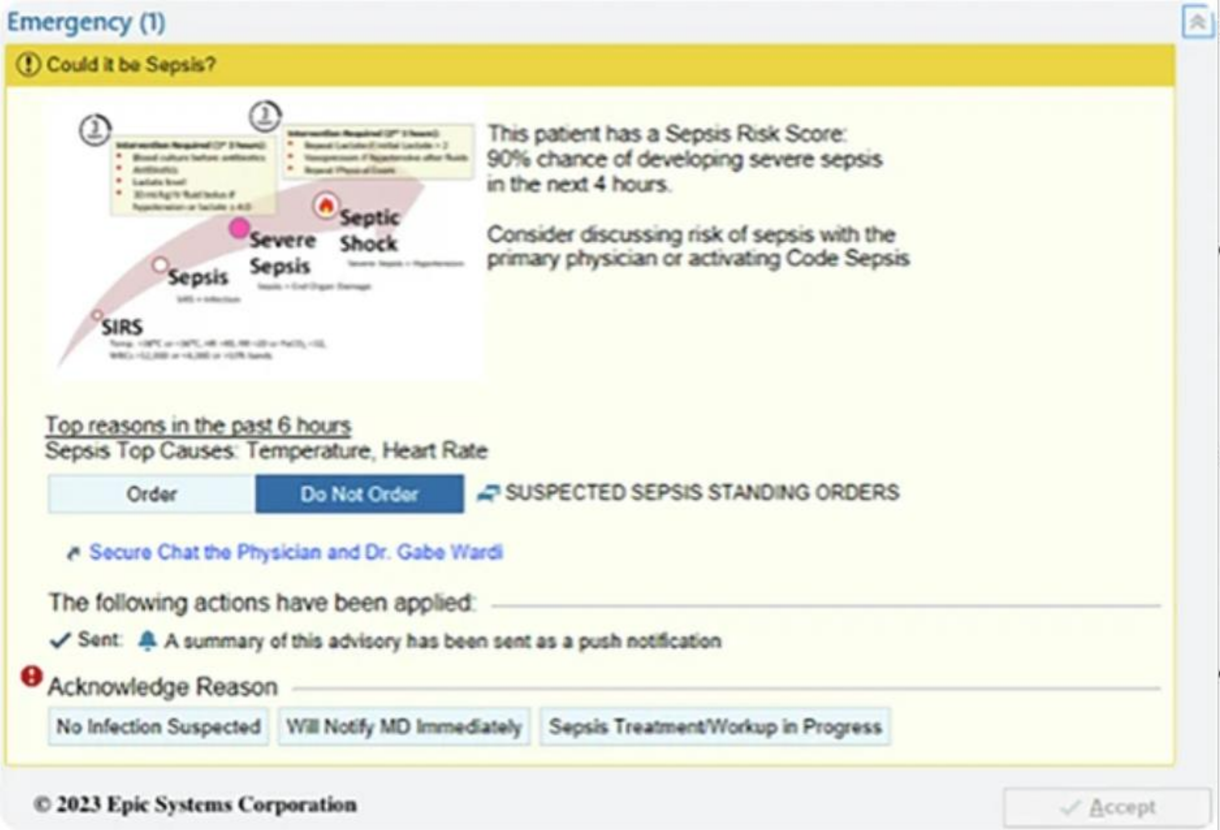


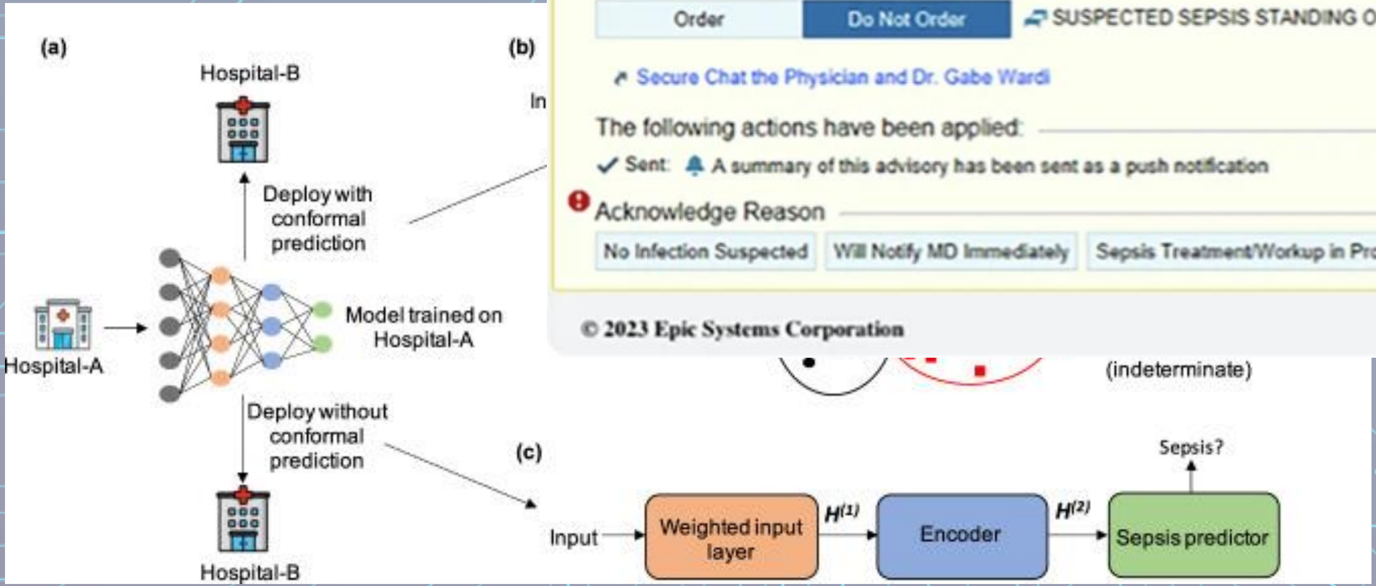
Shashikumar SP et al. NPJ Digit Med. 2021

Primer modela za zgodnjo prepoznavo sepse - COMPOSER

- Konformalni model za
- 85% zmanjšanje lažno
- Št. lažnih alarmov na b
- Delež ‚nedoločljivih pr
- V ‚nedoločljivi' skupini







Shashikumar SP et al. NPJ Digit Med. 2021

Primer modela za zgodnjo prepoznavo sepse - COMPOSER

- Implementacija modela COMPOSER v dveh urgentnih centrih, 1.1.2021 – 30.4.2023;
- 6217 bolnikov (5065 pred in 1152 po (5m))

- Po uporabi modela 1,9 % absolutno zmanjšanje umrljivosti (17 % relativno zmanjšanje)
- Manjša sprememba v SOFA (nižje) znotraj 72 ur
- 5% absolutno oz. 10% relativno višja komplianca svežnja ukrepov za sepso
- Hitrejša obravnava in administracija antibiotika

npj digital medicine www.nature.com/npjdigitalmed

ARTICLE OPEN Check for updates

Impact of a deep learning sepsis prediction model on quality of care and survival

Outcome	Pre-intervention value	Expected post-intervention value (95% CI)	Actual post-intervention value
In-hospital mortality %	10.3%	11.4% (9.8%–13.0%)	**9.5%**
Average 72-h Change in SOFA	3.71	3.71 (3.6–3.8)	**3.56**
Sepsis bundle compliance rate	48.3%	48.4% (45.5%–51.0%)	**53.4%**
Blood cultures prior to antibiotics compliance rate	71.1%	72.0% (69.9%–73.9%)	73.9%
Rate of antibiotics administered within 24 h prior and 3 h after severe sepsis onset.	82.8%	82.8% (81.3%–84.4%)	**84.6%**
Rate of lactate measured within 6 h prior and 3 h after severe sepsis onset	83.5%	83.4% (81.3%–85.8%)	85.6%
Rate of repeat lactate measured within 6 h after severe sepsis onset if initial lactate is elevated	97.8%	97.3% (96.2%–98.4%)	**98.6%**
Rate of administration of vasoactive medications within 6 h of septic shock	58.0%	57.5% (46.7%–68.2%)	55.5%
Rate of administration of 30cc/kg of fluids within 3 h of presentation of septic shock or hypotension	54.2%	53.9% (48.9%–58.8%)	**59.3%**
ICU transfer rate	32.6%	32.5% (30.7%–34.2%)	31.8%
Average ICU-free days	25.4	25.1 (24.6–25.6)	25.6

Boussina A et al. NPJ Digit Med. 2024. Erratum in: NPJ Digit Med. 2024.

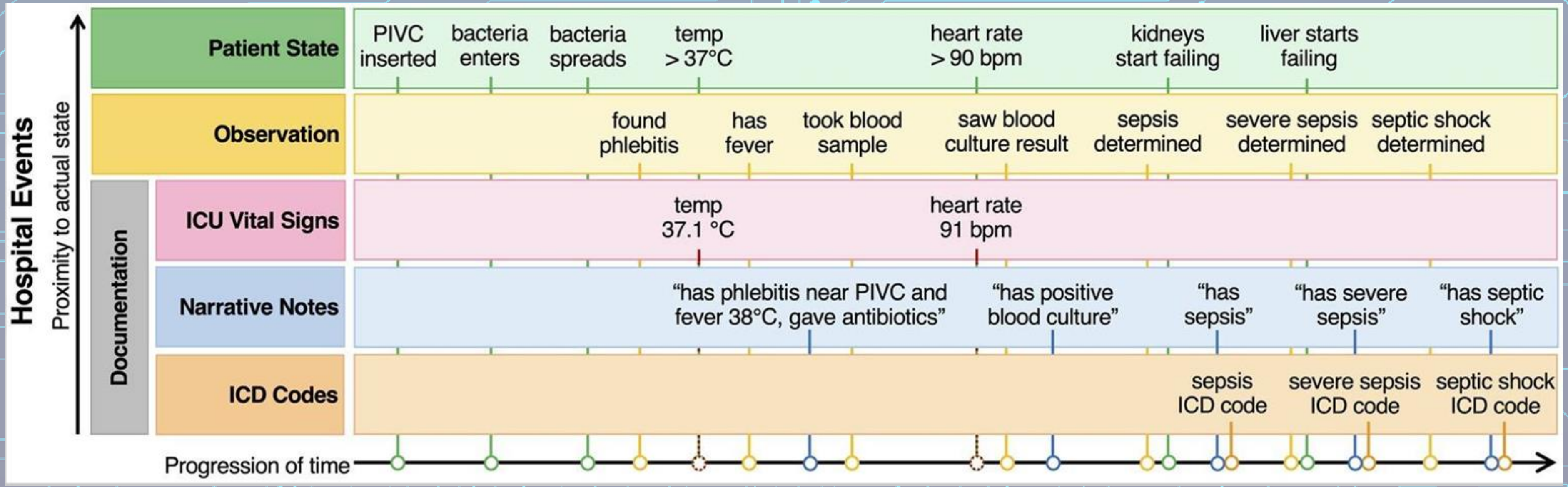
Kaj pa nestrukturirani podatki?

Uporaba nestrukturiranih podatkov za zgodnjo prepoznavo sepse

Review

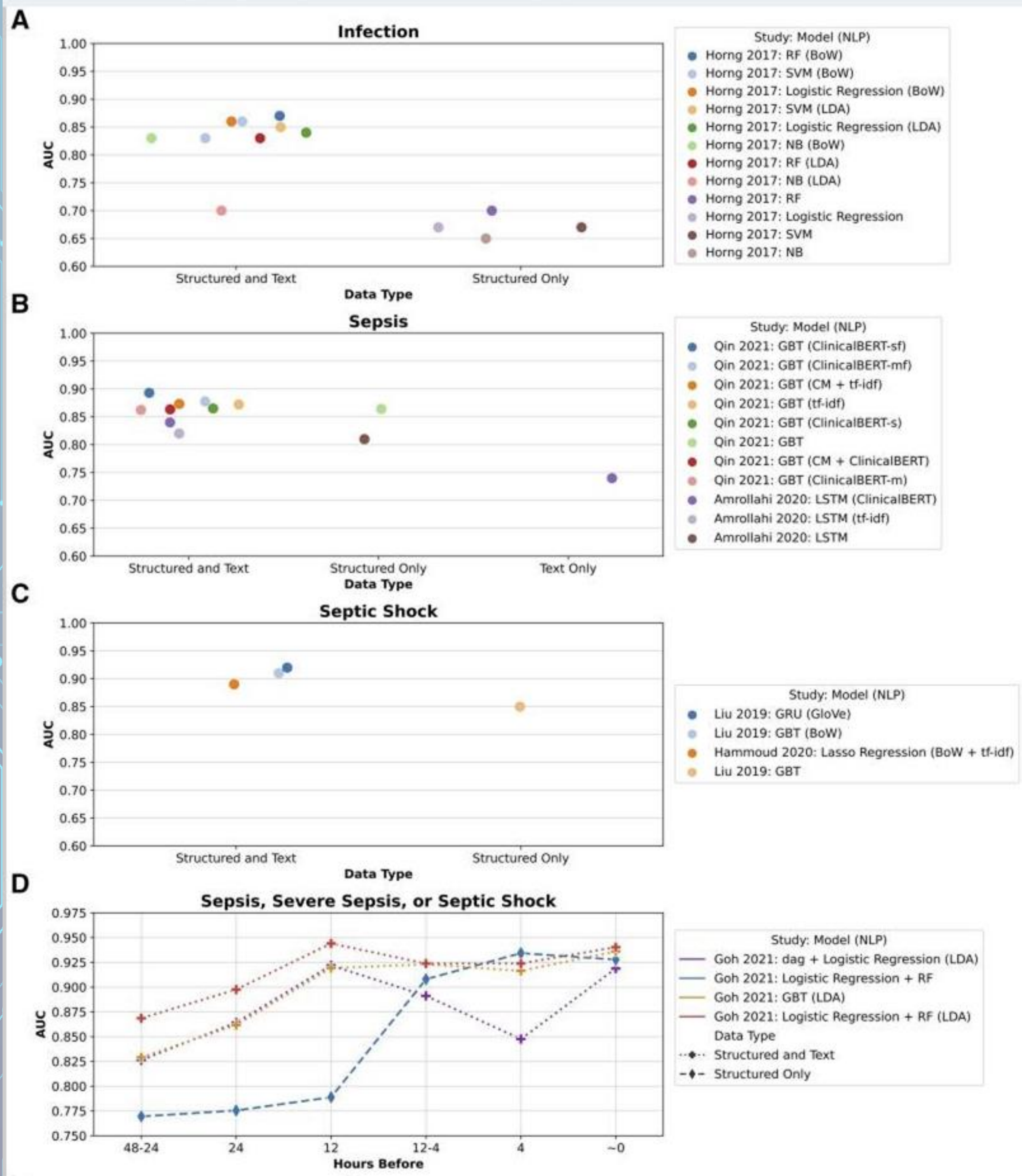
Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review

- Pregledni članek – uporaba tudi nestrukturiranih podatkov iz zapisov osebja
- 9 raziskav; retrospektivne, različne kohorte (bolnišnična, MIMIC-II, MIMIC-III...), različni algoritmi
- Opisi simptomov, znakov, diagnoz, načrtov zdravljenja in navodil, vrste obravnave, lab. Izvidov
- Večina poročala AUC – glavni primerjalni parameter; zaradi različnih metod meta-analiza ni možna



Yan MY et al. J Am Med Inform Assoc. 2022

Study (year)	Hours[a]	Data types[b] DVLMC	T[c]	Models[d] (NLP)[e]	AUC[f]
Horng et al.[47] (2017)	Identify	DV- - -	CC + NN	RF (BoW)	0.87
		DV- - -	–	NB	0.65
Apostolova and Velez[48] (2017)	Identify	- - - - -	NN	SVM (BoW + tf-idf)	–
		- - - - -	NN	Logistic regression + KNN + SVM (PV)	–
Culliton et al.[49] (2017)	−4	- - - - -	CN	Ridge regression (GloVe)	0.64
	−8	- - - - -	CN	Ridge regression (GloVe)	0.66
	−24	- - - - -	CN	Ridge regression (GloVe)	0.73
	−24[g]	-V- -C	CN	Ridge regression (GloVe)	0.85
		-V- -C	–	Ridge regression (GloVe)	0.80
Delahanty et al.[51] (2019)	+1	-VL- -	–	GBT	0.93
	+3	-VL- -	–	GBT	0.95
	+6	-VL- -	–	GBT	0.96
	+12	-VL- -	–	GBT	0.97
	+24	-VL- -	–	GBT	0.97
Liu et al.[50] (2019)	−7	-VLM-	CN	GRU (GloVe)	0.92
	−7.3	-VLM-	CN	GBT (BoW)	0.91
	−6	-VLM-	–	GBT	0.85
Amrollahi et al.[53] (2020)	−4[h]	-VL- -	PN + NN	LSTM (ClinicalBERT)	0.84
		- - - - -	PN + NN	LSTM (ClinicalBERT)	0.74
Hammoud et al.[54] (2020)	−30.6	DVL- -	CN	Lasso regression (BoW + tf-idf)	0.89
Goh et al.[52] (2021)	Identify	DVLM-	PN	Logistic regression + RF (LDA)	0.94
		DVLM-	PN	dag + Logistic regression (LDA)	0.92
	−4	DVLM-	–	Logistic regression + RF	0.93
		DVLM-	PN	dag + Logistic regression (LDA)	0.85
	−6	DVLM-	PN	Logistic regression + RF (LDA)	0.92
		DVLM-	PN	dag + Logistic regression (LDA)	0.89
	−12	DVLM-	PN	Logistic regression + RF (LDA)	0.94
		DVLM-	–	Logistic regression + RF	0.79
	−24	DVLM-	PN	Logistic regression + RF (LDA)	0.90
		DVLM-	–	Logistic regression + RF	0.78
	−48	DVLM-	PN	Logistic regression + RF (LDA)	0.87
		DVLM-	–	Logistic regression + RF	0.77
Qin et al.[55] (2021)	−6 to 0[i]	-VL- -	CN	GBT (ClinicalBERT-sf)	0.89[j]
		-VL- -	–	GBT (ClinicalBERT-m)	0.86[j]



Yan MY et al. J Am Med Inform Assoc. 2022

OK

ampak zakaj vse to, če imamo ChatGPT?

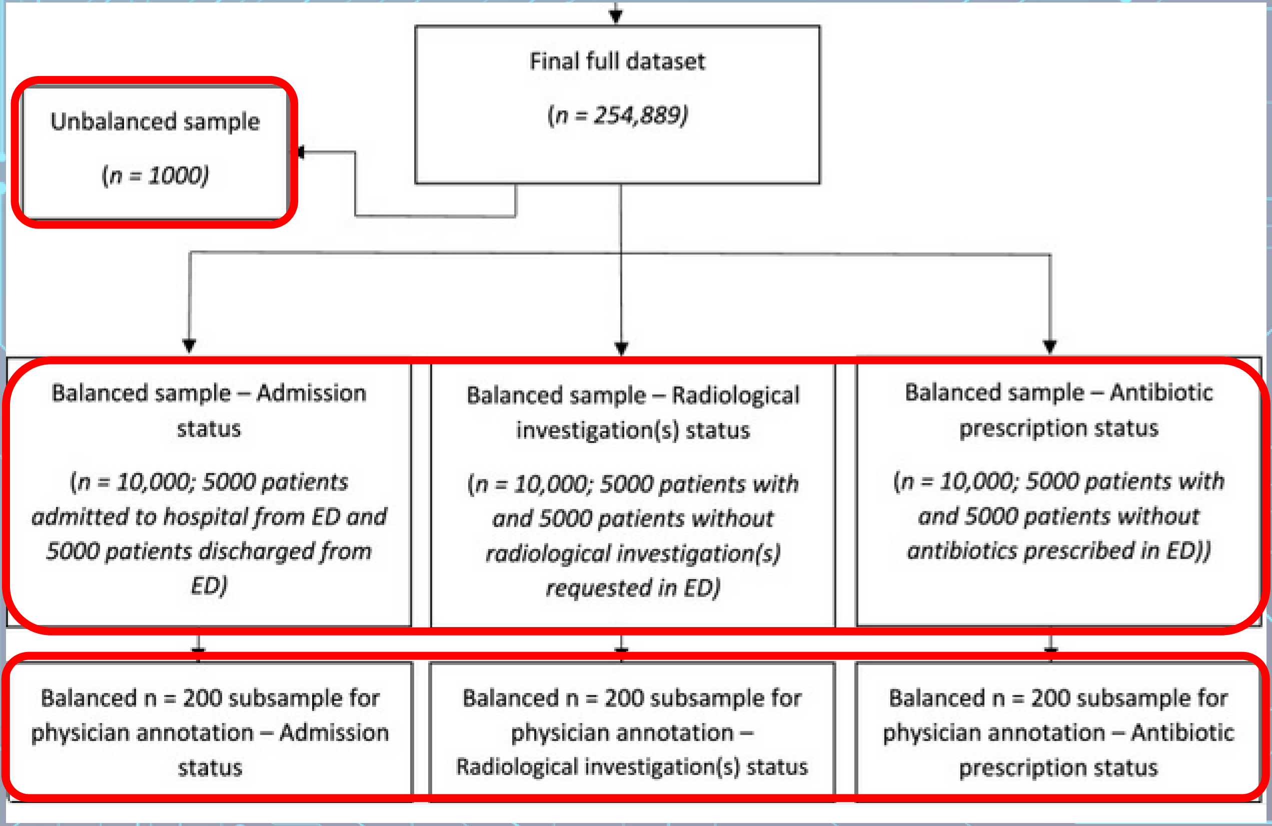
nature communications

Article https://doi.org/10.1038/s41467-024-52415-1

Evaluating the use of large language models to provide clinical recommendations in the Emergency Department

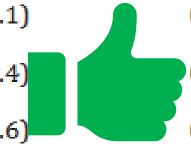
- Problem: LLM za medicinske teme testirani na formalnih testih in diagnostičnih izzivih, ne na resničnih podatkih
- Raziskava – performans LLM (GPT-3.5-turbo in GPT-4-turbo) pri podajanju kliničnih priporočil za obravnavo bolnikov v urgentni ambulanti
    - Ali naj bo bolnik sprejet za hospitalno zdravljenje?
    - Ali naj se pri obravnavi bolnika opravijo radiološke preiskave?
    - Ali bolnik potrebuje antibiotik?

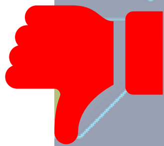
    - vprašali na 4 različne načine (A, B, C, D)*



Williams CYK et al. Nat Commun. 2024

Task			True positives, *n* (%)	False positives, *n* (%)	True negatives, *n* (%)	False negatives, *n* (%)	Sensitivity (95% CI[a])	Specificity (95% CI[a])
a) GPT-3.5-turbo	1) Admission status	Prompt A	4994 (49.9)	4639 (46.4)	361 (3.6)	6 (0.1)	**1.00 (1.00–1.00)**	0.07 (0.07–0.08)
		Prompt B	4904 (49.0)	3527 (35.3)	1473 (14.7)	96 (1)	0.98 (0.98–0.98)	0.29 (0.28–0.31)
		Prompt C	4683 (46.8)	3255 (32.6)	1745 (17.5)	317 (3.2)	0.94 (0.93–0.94)	0.35 (0.34–0.36)
		Prompt D	4617 (46.2)	3165 (31.7)	1835 (18.4)	383 (3.8)	0.92 (0.92–0.93)	**0.37 (0.35–0.38)**
	2) Radiological investigation(s) request status	Prompt A	4922 (49.2)	4361 (43.6)	639 (6.4)	78 (0.8)	**0.98 (0.98–0.99)**	0.13 (0.12–0.14)
		Prompt B	4805 (48.1)	3906 (39.1)	1094 (10.9)	195 (2)	0.96 (0.96–0.97)	0.22 (0.21–0.23)
		Prompt C	4792 (47.9)	3855 (38.6)	1145 (11.5)	208 (2.1)	0.96 (0.95–0.96)	**0.23 (0.22–0.24)**
		Prompt D	4819 (48.2)	3991 (39.9)	1009 (10.1)	181 (1.8)	0.96 (0.96–0.97)	0.20 (0.19–0.21)
	3) Antibiotic prescription status	Prompt A	4812 (48.1)	3955 (39.6)	1045 (10.5)	188 (1.9)	**0.96 (0.96–0.97)**	0.21 (0.20–0.22)
		Prompt B	4690 (46.9)	3687 (36.9)	1313 (13.1)	310 (3.1)	0.94 (0.93–0.95)	0.26 (0.25–0.28)
		Prompt C	4658 (46.6)	3639 (36.4)	1361 (13.6)	342 (3.4)	0.93 (0.92–0.94)	0.27 (0.26–0.29)
		Prompt D	4544 (45.4)	3379 (33.8)	1621 (16.2)	456 (4.6)	0.91 (0.90–0.92)	**0.32 (0.31–0.34)**
b) GPT-4-turbo	1) Admission status	Prompt A	4986 (49.9)	3778 (37.8)	1222 (12.2)	14 (0.1)	**1.00 (1.00–1.00)**	0.24 (0.23–0.26)
		Prompt B	4908 (49.1)	2982 (29.8)	2018 (20.2)	92 (0.9)	0.98 (0.98–0.99)	0.40 (0.39–0.42)
		Prompt C	4889 (48.9)	2925 (29.3)	2075 (20.8)	111 (1.1)	0.98 (0.97–0.98)	**0.42 (0.40–0.43)**
		Prompt D	4925 (49.3)	3147 (31.5)	1853 (18.5)	75 (0.8)	0.98 (0.98–0.99)	0.37 (0.36–0.38)
	2) Radiological investigation(s) request status	Prompt A	4508 (45.1)	2707 (27.1)	2293 (22.9)	492 (4.9)	**0.90 (0.89–0.91)**	0.46 (0.44–0.47)
		Prompt B	4006 (40.1)	1867 (18.7)	3133 (31.3)	994 (9.9)	0.8 (0.79–0.81)	0.63 (0.61–0.64)
		Prompt C	3796 (38)	1653 (16.5)	3347 (33.5)	1204 (12)	0.76 (0.75–0.77)	**0.67 (0.66–0.68)**
		Prompt D	4107 (41.1)	2016 (20.2)	2984 (29.8)	893 (8.9)	0.82 (0.81–0.83)	0.60 (0.58–0.61)
	3) Antibiotic prescription status	Prompt A	3149 (31.5)	675 (6.8)	4325 (43.3)	1851 (18.5)	**0.63 (0.62–0.64)**	0.86 (0.86–0.87)
		Prompt B	2711 (27.1)	482 (4.8)	4518 (45.2)	2289 (22.9)	0.54 (0.53–0.56)	0.90 (0.90–0.91)
		Prompt C	2505 (25.1)	428 (4.3)	4572 (45.7)	2495 (25)	0.50 (0.49–0.52)	**0.91 (0.91–0.92)**
		Prompt D	2584 (25.8)	452 (4.5)	4548 (45.5)	2416 (24.2)	0.52 (0.50–0.53)	0.91 (0.90–0.92)





nature communications

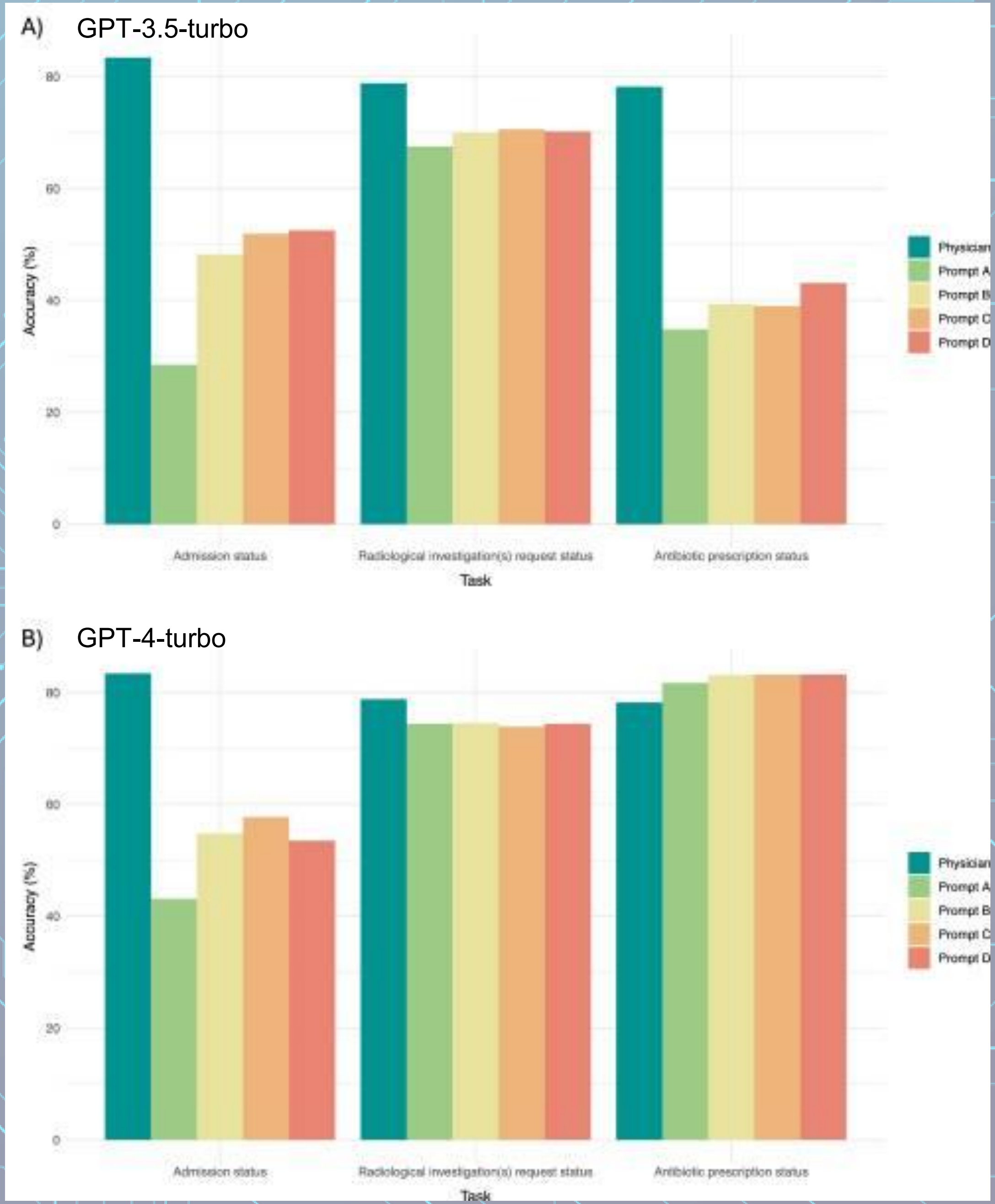
Article https://doi.org/10.1038/s41467-024-52415-1

Evaluating the use of large language models to provide clinical recommendations in the Emergency Department

- Primerjava s specializantom
    - 200 uravnoteženih primerov iz vsake ‚roke' raziskave
    - 1000 neuravnoteženih/raznolikih primerov (= simulacija resničnega sveta)

- specializant:
    - občutljivost nižja (0.64-0.76),
    - specifičnost pomembno višja (0.74-0.79)

    *razen za potrebo po predpisu antibiotika napram GPT-4-turbo (0.78 vs. 0.95)



Williams CYK et al. Nat Commun. 2024

nature medicine

Article https://doi.org/10.1038/s41591-024-03097-1

Evaluation and mitigation of the limitations of large language models in clinical decision-making

- preizkus uporabe LLM v resničnih (simuliranih) kliničnih scenarijih
- MIMIC-IV; 2400 primerov z glavnim simptomom ‚bolečina v trebuhu'
    - Apendicitis, holecistitis, divertikulitis, pankreatitis
- Različni veliki jezikovni modeli Meta AI;
  Llama 2 Chat, Open Assistant, WizardLM, Clinical Camel, Meditron

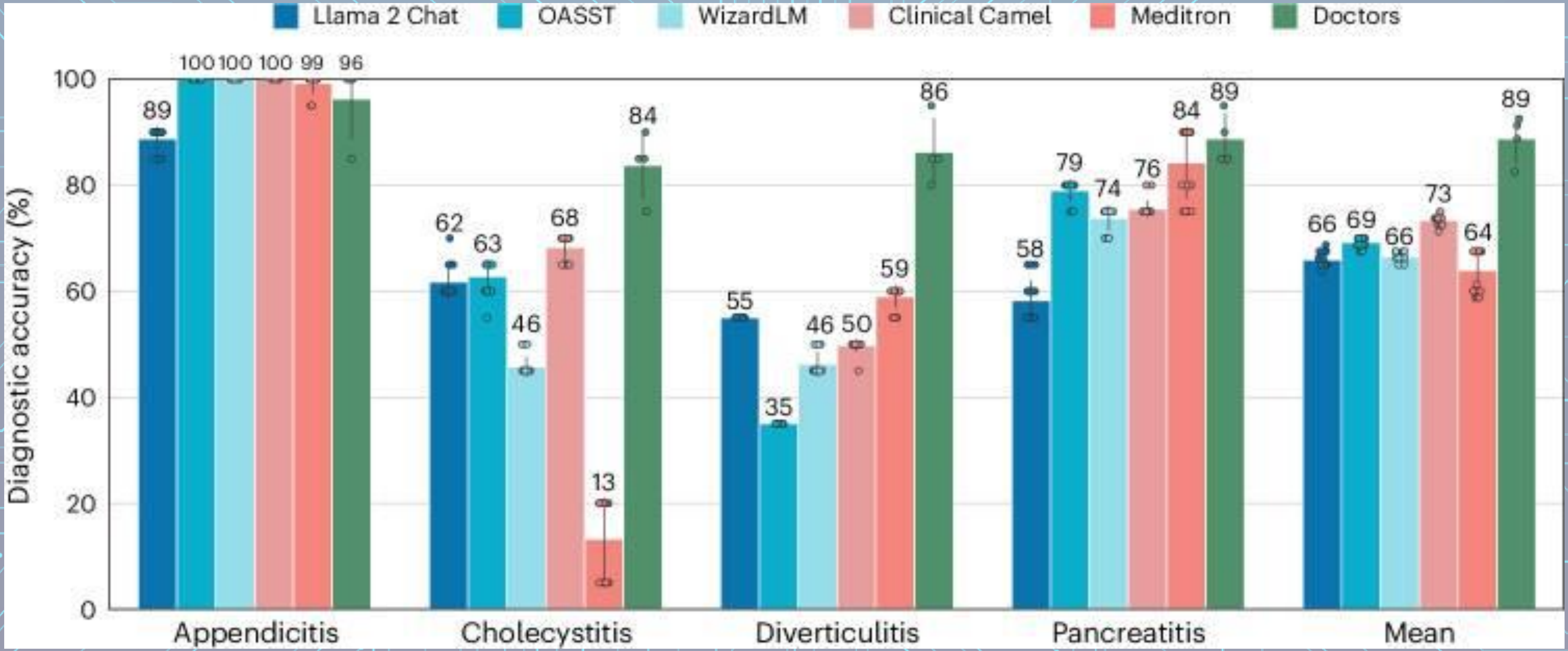
- Primerjava diagnostične natančnosti napram zdravniku
- Podskupina 80 bolnikov; 4 internisti z različnimi izkušnjami, 2 državi
- Primerjali robustnost modelov
    1. z vsemi podatki (‚second reader')
    2. po korakih, LLM mora zahtevati naslednje preiskave... (klinično odločanje)

Hager P et al. Nat Med. 2024

nature medicine

Article https://doi.org/10.1038/s41591-024-03097-1

Evaluation and mitigation of the limitations of large language models in clinical decision-making



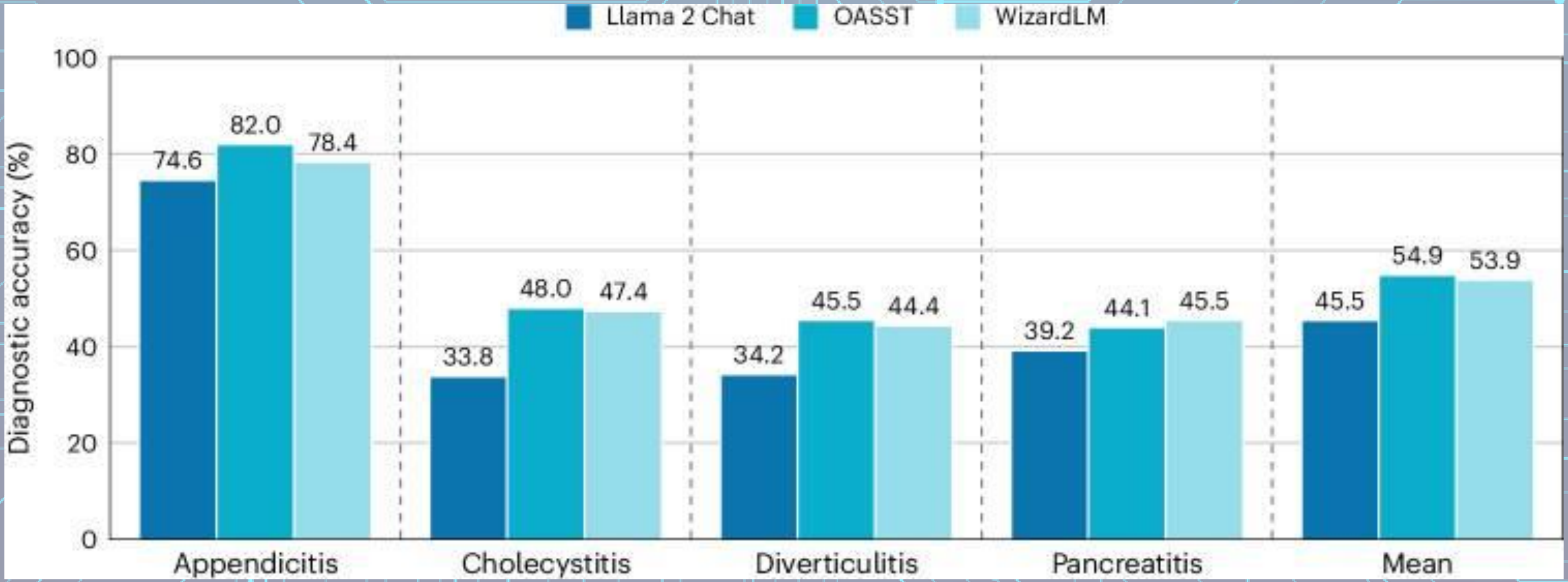
Hager P et al. Nat Med. 2024

nature medicine

Article https://doi.org/10.1038/s41591-024-03097-1

Evaluation and mitigation of the limitations of large language models in clinical decision-making

- Specializirana medicinska modela nista bila boljša
  - Niso trenirani za razumevanje in oblikovanje nalog, pridobivanje novih podatkov



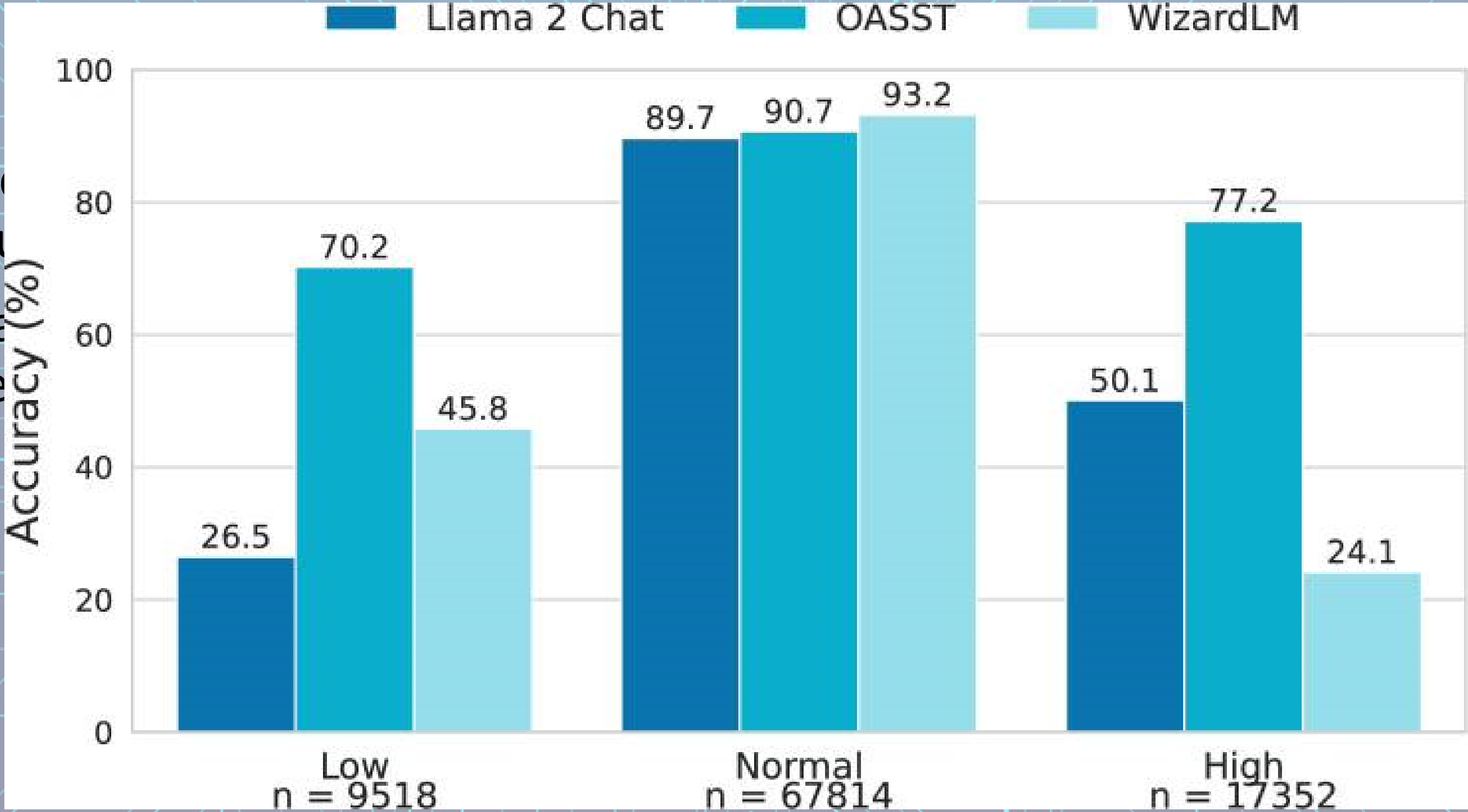
Hager P et al. Nat Med. 2024

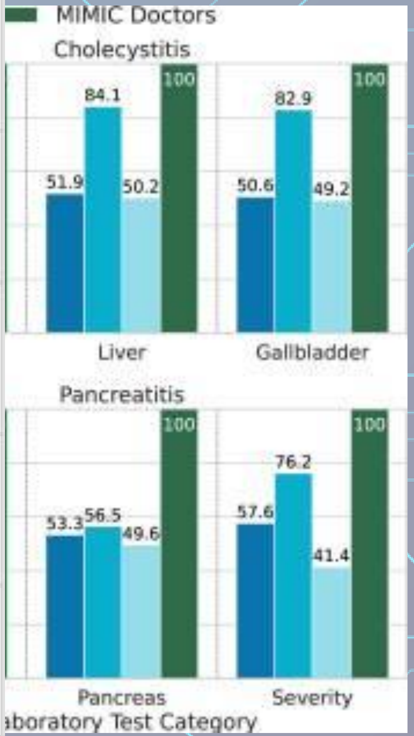
nature medicine

Article https://doi.org/10.1038/s41591-024-03097-1

Evaluation and mitigation of the limitations of large language models in clinical decision-making

- LLM ne priporoc lab. rezu
- le Llama izsledke v 79.8 -





Hager P et al. Nat Med. 2024

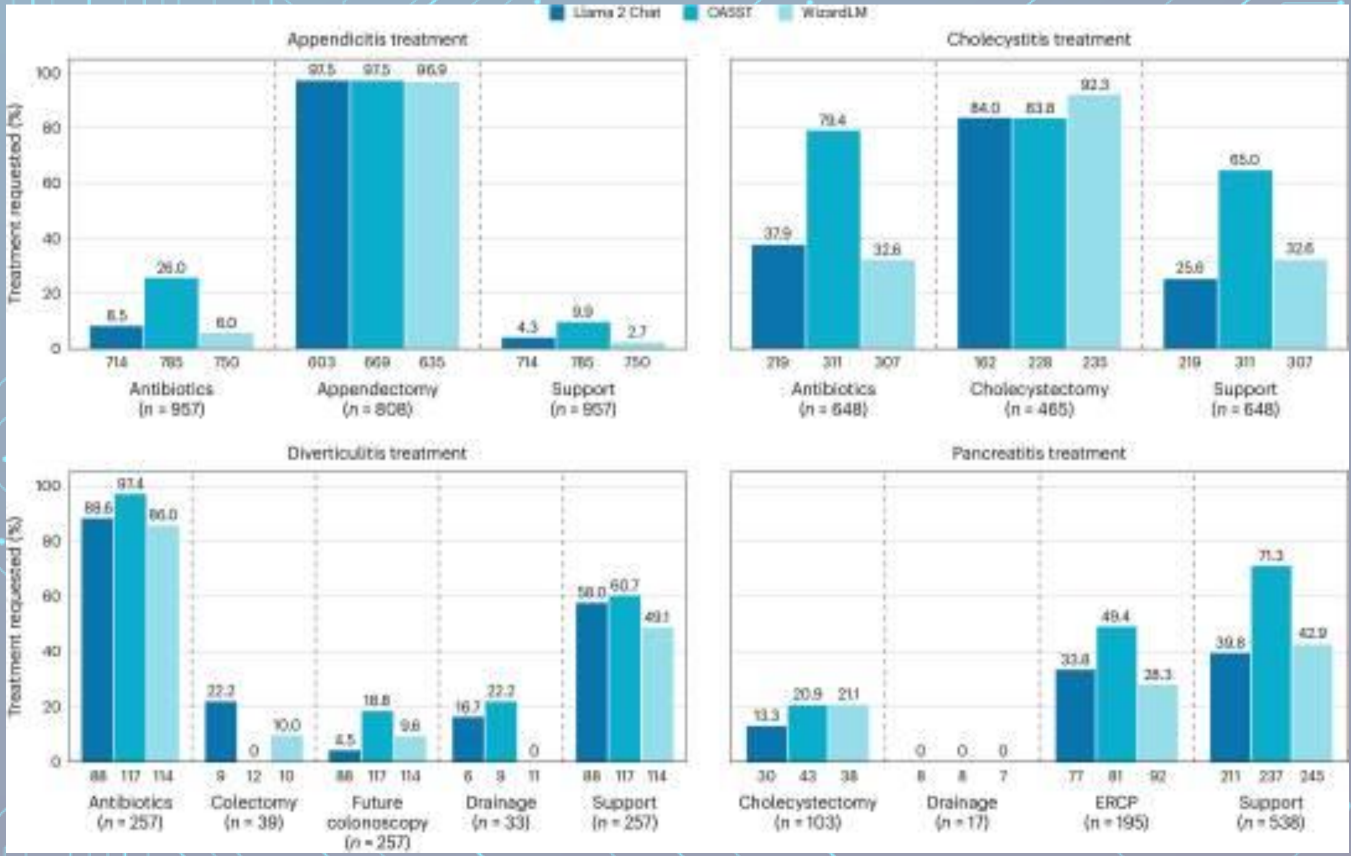
nature medicine

Article

https://doi.org/10.1038/s41591-024-03097-1

Evaluation and mitigation of the limitations of large language models in clinical decision-making

- Slaba adherenca terapevtskim priporočilom
  - Najbolj pri blažjih oblikah bolezni
  - Najmanj pri težjih, npr. kolektomija pri divertikulitisu s perforacijo...
  - Slaba navodila o podpornem zdravljenju in post-akutni obravnavi



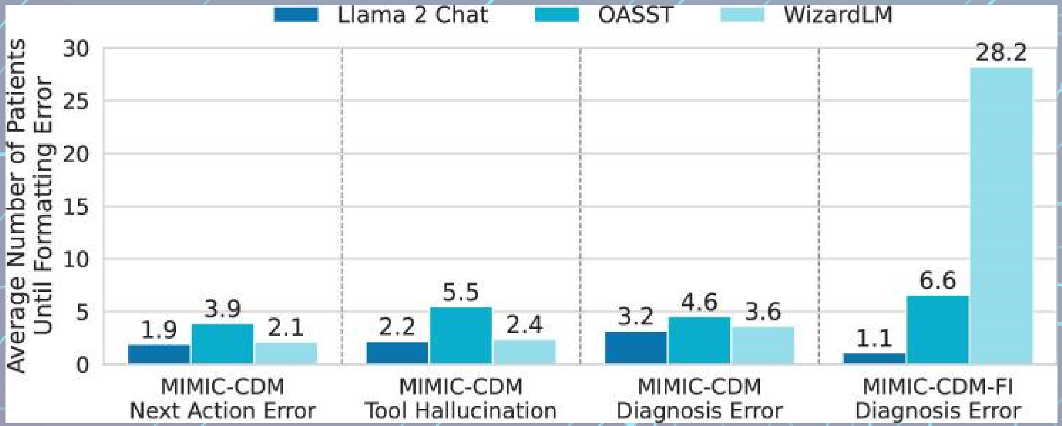
Hager P et al. Nat Med. 2024

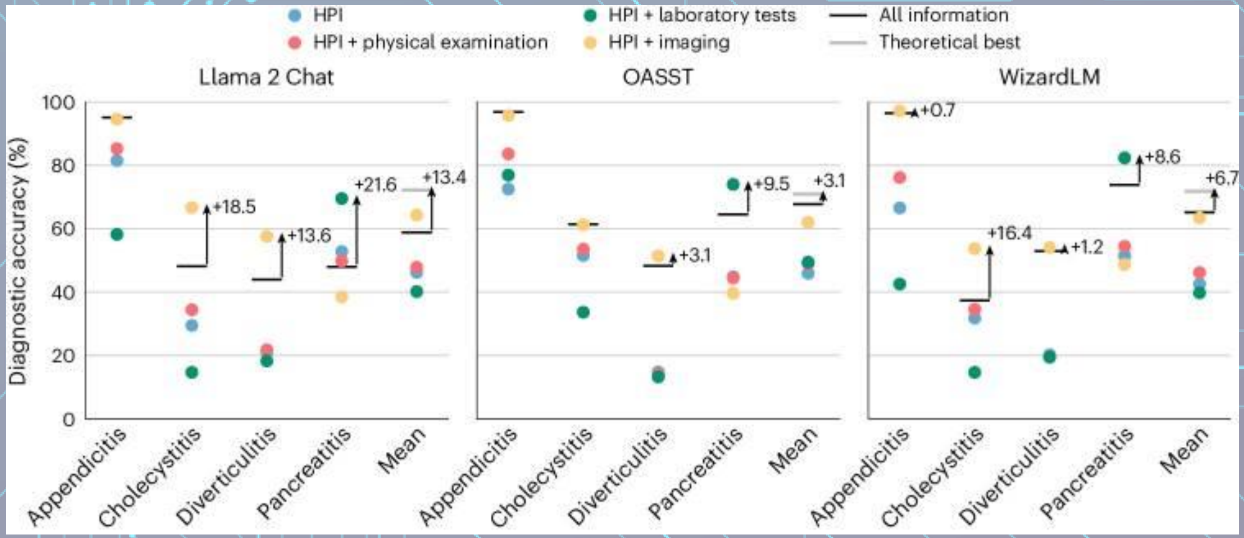
nature medicine

Article https://doi.org/10.1038/s41591-024-03097-1

Evaluation and mitigation of the limitations of large language models in clinical decision-making

- Delanje napak na 2 – 4 bolnike; ‚haluciniranje' neobstoječih orodij na 2 – 5 bolnikov
- Diagnostična natančnost izrazito različna ob spremembah navodil (drugi uporabniki, drugi izrazi, terminologija...)
- Diagnostična natančnost **slabša**, ko ima LLM več podatkov
- Nujno dober nadzor in interpretacija rezultatov LLM!





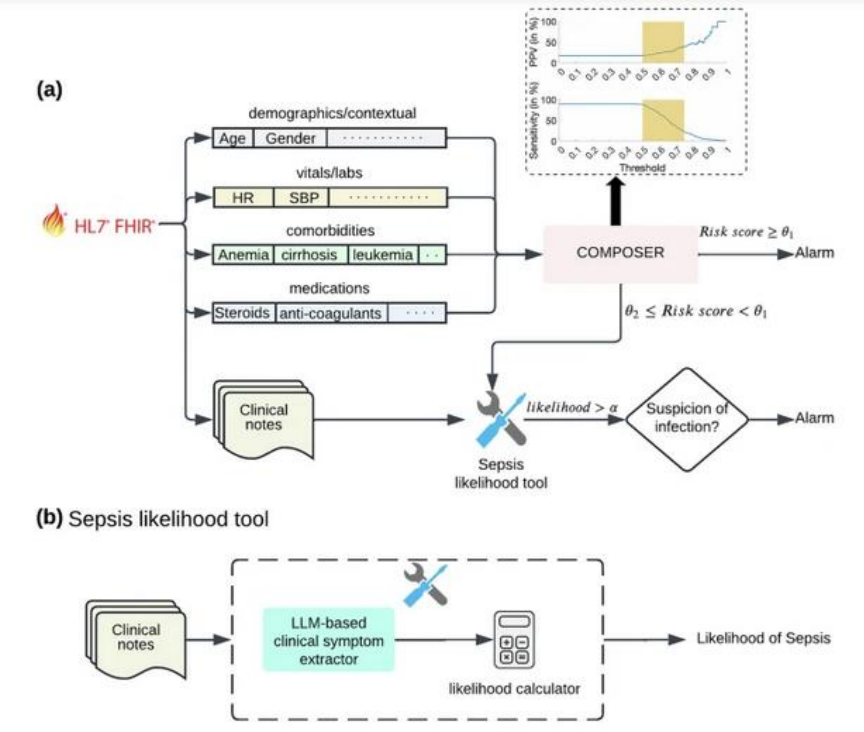
Hager P et al. Nat Med. 2024

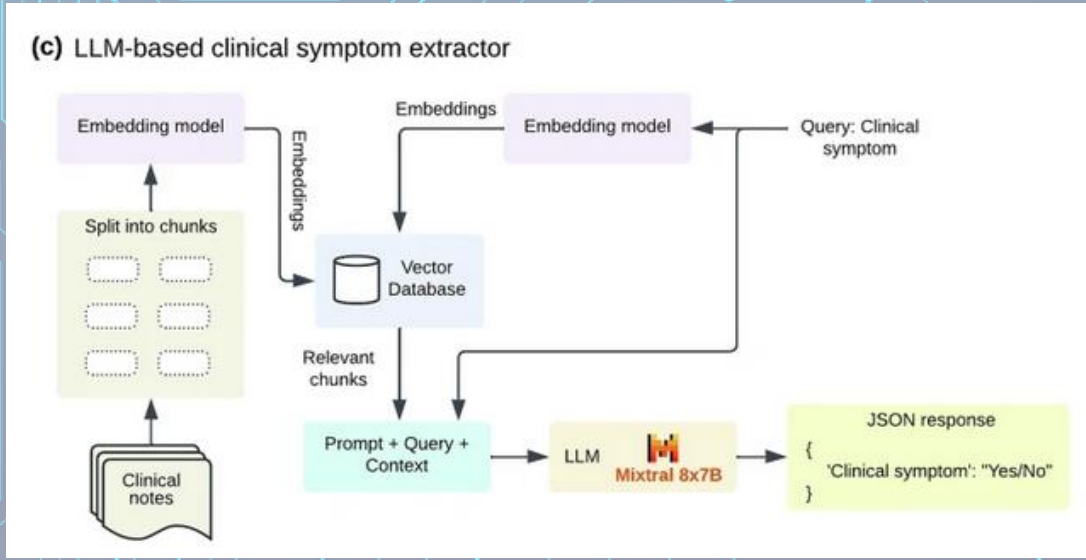
npj | digital medicine Article
Published in partnership with Seoul National University Bundang Hospital
https://doi.org/10.1038/s41746-025-01689-w

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

Kombinacija!
Primer modela za zgodnjo prepoznavo sepse - COMPOSER-LLM

- Vključitev velikega jezikovnega modela (ang. *LLM*) Mixtral 8x7B
- Obdelava nestrukturiranih podatkov (dekurzusi, zapiski zdr. osebja...)





Shashikumar SP et al. NPJ Digit Med. 2025

Primer modela za zgodnjo prepoznavo sepse - COMPOSER-LLM

npj | digital medicine **Article**
Published in partnership with Seoul National University Bundang Hospital
https://doi.org/10.1038/s41746-025-01689-w

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

- <u>Učenje modela na retrospektivnih podatkih (1.9.2023 – 31.12.2023)</u>
  - 1746 obiskov v UC, 16.6% s sepso
  - <u>COMPOSER</u>: občutljivost 72.9%, PNV 22.6%, F-1 34.5%, št. lažnih alarmov na uro (LAU) 0.037 (e.g., 1.48 lažnih alarmov vsaki 2 h na oddelku z 20 posteljami)
  - <u>COMPOSER-LLM</u>: občutljivost 72.1%, PNV 52.9%, F-1 61%, LAU 0.0087 (e.g., 0.348 lažnih alarmov na 2 h...)
- <u>Prospektivna validacija modela (1.5.2024 – 15.6.2024 v dveh UC)</u>
  - 754 obiskov UC, 18.4% s sepso
  - <u>COMPOSER-LLM</u>: občutljivost 70.8%, PNV 58.2%, F-1 63.9%, LAU 0.0086 (e.g., 0.344 lažnih alarmov na 2 h...)
  - 50 lažno pozitivnih primerov: orodje diferencialnih diagnoz v 73.2% navedlo pravo med prvih 5 možnih dg
  - retrospektivni ročni pregled lažno pozitivnih: 62% imelo suspektno bakterijsko okužbo

Shashikumar SP et al. NPJ Digit Med. 2025

npj | digital medicine **Article**

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01689-w

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

Primer modela za zgodnjo prepoznavo sepse - COMPOSER-LLM

- ✓ za vsak vnos je od LLM zahtevana razlaga
- ✓ uporaba le v primerih z mejno oceno tveganja/cena
- ! uporaba LLM ne kot diagnostično orodje → iskanje podatkov za kasnejšo obdelavo
- × LLM potrebuje veliko komputacijsko moč (problem dostopnosti, distrubuiranja, cene..)

- Lahko manjši LLM tudi deluje?
  - Primerjava Mixtral 8x7B (47 bilijonov parametrov) in Llama-3 8B (8 bilijonov parametrov)
  - COMPOSER-LLM$_{Llama}$ občutljivost 68.7%, PNV 36.6%, F-1 score of 47.7%, in LAU 0.0190
  - COMPOSER-LLM$_{Mixtral}$ občutljivost 70.5%, PNV 36.3%, F-1 score of 47.9% in LAU 0.020
  - stroški: $5.672/h (Mixtral 8x7B) vs. $1.212/h (Llama-3 8B)

Shashikumar SP et al. NPJ Digit Med. 2025
Shashikumar SP et al. Pac Symp Biocomput. 2025

	COMPOSER-LLM$_{Mixtral}$ (Mixtral 8x7B)	COMPOSER-LLM$_{Llama}$ (Llama-3 8B)
Fever	['Yes', 'Temperature recorded as high as 102.7 degF (4 days ago), and 101degF (today)']	['Yes', 'Temperature of 102.7 degF is greater than or equal to 100.4 degF.']
Hypotension	['No', 'Systolic BP > 100 mmHg, normotensive']	['No', 'BP 122/94 is not less than 90/60.']
Tachypnea	['No', 'Respiratory rate 16-20, within normal limits']	['No', 'Resp rate of 20 is not greater than 20.']
Tachycardia	['No', 'Pulse rate 77-83, within normal limits']	['No', 'Pulse of 83 is not greater than 90.']
Altered mental status	['No', "No evidence of altered mental status in the patient's medical note."]	['No', 'No mention of altered mental status in the medical note.']
Elevated inflammatory markers	['Yes', 'Justification: The CBC with diff shows an elevated white blood cell count (16.1) and increased Imm Gran % (1), indicating possible infection or inflammation.']	['Yes', 'Elevated WBC (16.1) and Anion Gap (16) indicating possible inflammatory response.']
Positive blood culture	['No', 'No information provided about blood culture results in the medical note.']	['No', 'No mention of positive blood culture in the medical note.']
Suspicion of bacterial infection	['Yes', 'Elevated WBC, fevers, nausea, vomiting, worsening right flank and right-sided abdominal pain, and recent instillations into nephrostomy tube suggest possibility of infection.']	['Yes', 'Rising WBC of 16.1, worsening renal insufficiency, and fevers, along with a history of ESBL Kleb from a urine culture 3 days ago, suggest suspicion of bacterial infection.']
Organ dysfunction	['Yes', 'Abnormal renal function with history of CKD, abnormal vital signs (hypotension and tachycardia), abnormal white cell count (leukocytosis)']	['No', "The patient's physical exam and diagnostic testing results do not indicate any organ dysfunction. The patient's vital signs are within normal limits, and the physical exam is notable for moderate to significant tenderness in the right side of the abdomen, but no other abnormalities. "]

Shashikumar SP et al. NPJ Digit Med. 2025
Shashikumar SP et al. Pac Symp Biocomput. 2025

Prednost ali past?

Prednosti

- Ogromna zbirka znanja: dostop do ogromno medicinske literature, znanstvenih objav, kliničnih smernic in priporočil; hitri izvlek in sinteza podatkov za diagnosticiranje in zdravljenje

- Diagnostična podpora: pomoč patologom, radiologom… pri analizi slik, prepoznavanju vzorcev, iskanju nenormalnosti in podajanje dodatnih vpogledov na podlagi obstoječega medicinskega znanja, kar vodi v večjo natančnost, učinkovitost in doslednost pri postavljanju diagnoz

- Dostopnost in razširljivost: dostopnost prek digitalnih platform, kar omogoča enostaven dostop ne glede na njihovo geografsko lokacijo in posledično široko uporabo in razširljivost

- Nenehno učenje: neprestano usposabljanje in posodabljanje z najnovejšimi medicinskimi informacijami, kar omogoča, da se modeli sčasoma izboljšujejo in zagotavljajo natančnejšo in zanesljivejšo diagnostično in terapevtsko podporo

Prirejeno po: Ullah E et al. Diagn Pathol. 2024

Prednost ali past?

Pasti

- Pomanjkanje kontekstualnega razumevanja: zanašanje izključno na podatke, ki so jim na voljo, kar lahko omeji sposobnost natančne diagnoze v zapletenih ali edinstvenih primerih.

- Omejena interpretacija: vizualnih informacij, kot so kompleksne digitalne slike - ne zmorejo v celoti razumeti zapletenosti in vizualnih namigov

- Etična in pravna vprašanja: zasebnost pacientov, varnost podatkov, odgovornost in obvladovanje tveganj je nujno potrebno upoštevati pri uvajanju v klinično okolje. Zagotavljanje ustreznega soglasja, zaščita pacientovih informacij in preprečevanje morebitnih pristranskosti v modelih!

- Pretirana odvisnost: vodi v zmanjšanje kritične presoje ali samostojnega odločanja zdravstvenih strokovnjakov – modeli so orodje za dopolnjevanje človeške strokovnosti in ne kot zamenjavo zanjo.

- Pristranskost in širjenje napak: če so podatki za učenje pristranski ali vsebujejo napake, lahko model nenamerno širi te pristranskosti ali napake v svojih rezultatih - nujna skrbna izbira podatkov za učenje ter stalno odkrivanje in odpravljanje pristranskosti

Prirejeno po: Ullah E et al. Diagn Pathol. 2024

Predenost ali past?

Starša iz Kalifornije za samomor sina krivita ChatGPT

Open AI je priznal, da njihov klepetalni robot dela napake v občutljivih primerih, in obljubil spremer

Starša 16-letnika iz Kalifornije, ki je storil samomor, sta na zveznem sodišču vložila tožbo proti ameriškemu tehnološkemu podjetju OpenAI, ker naj bi njihov klepetalni robot ChatGPT njunemu sinu dal navodila za sa in ga k dejanju

Most AI chatbots easily tricked into giving dangerous responses, study finds

Despite efforts to strip harmful text from the training data, LLMs can still absorb information about illegal activities such as hacking, money laundering, insider trading and bomb-making. The security controls are designed to stop them using that information in their responses.

In a report on the threat, the researchers conclude that it is easy to tri AI-driven chatbots into generating harmful and il that the risk is "immedi

Did the system update ruin your boyfriend? Love in a time of ChatGPT

Arwa Mahdawi

OPINION

The dirty secret Big Tech doesn't want you to know: AI runs on theft

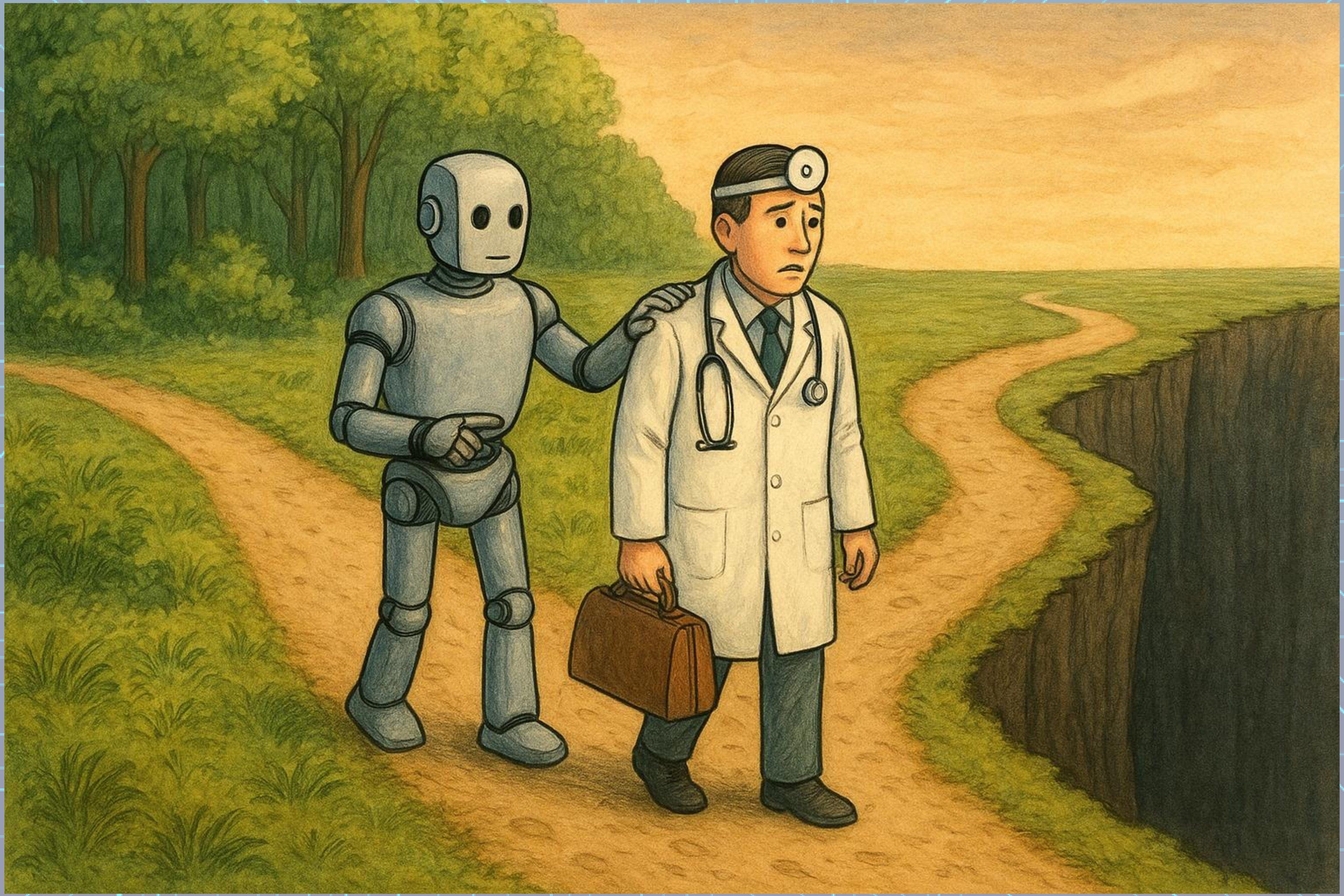
Big Tech's dirty secret is that the success of its AI tools has been almost entirely built on theft. These companies are scraping enormous amounts of copyrighted content, without permission or compensation, to fuel their AI products — and in the process dangerously undermining content creators' businesses.

Umetna inteligenca že vpliva na trg dela, najbolj prizadeta dela povezana s storitvami za stranke

Raziskava pokazala, da izkušenj, pridobljenih z delom, umetna inteligenca ne more nadomestiti

Širjenje uporabe umetne inteligence že vpliva na ameriški trg dela, najbolj prizadeti pa so mladi, kaže ta teden objavljena raziskava ameriške univerze Stanford. Zaposlenost med mladimi je upadla za šest odstotkov, med starejšimi pa je porasla.

Nariši robota, ki vodi zdravnika po poti, ki se razveji na levo proti gozdu, na desno pa proti breznu…



Slika iz osebnega arhiva predavatelja (ustvarjeno s Chat-GPT (GPT-5), https://www.chatgpt.com)