Department of Neurological Surgery
Department of Radiology
NYU Institute for Neuroscience
NYU Center for Data Science
Global AI Frontier Lab

AI in Stroke Care: Where and When

Eric K. Oermann, M.D.
Associate Professor of Neurosurgery, Radiology, and Data Science
Associate Program Director, Neurosurgery Residency Program
Director of the OLAB, Health AI Research Lab at NYU Langone Health
Investigator, Neuroscience Institute, NYU
Principal Investigator, Global AI Frontier Lab
Chair of Data Science, Congress of Neurological Surgeons
Associate Editor for Data Science, Neurosurgery

Disclosure Statement of Financial Interest

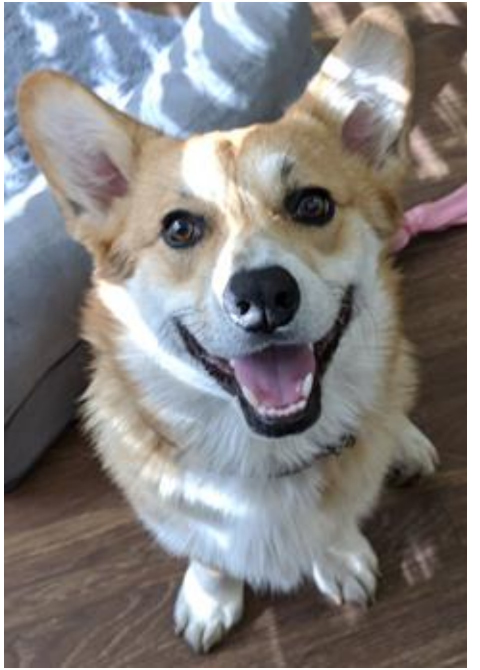
Within the past 12 months, I or my spouse/partner have had a financial interest/arrangement or affiliation with the organization(s) listed below.

Grant/research support: NIH, Keck Foundation, GAIFL

Consultant: Google Inc., Sofinnova Inc., ATEC

Equity: Artisight Inc., MarchAI, Delvi

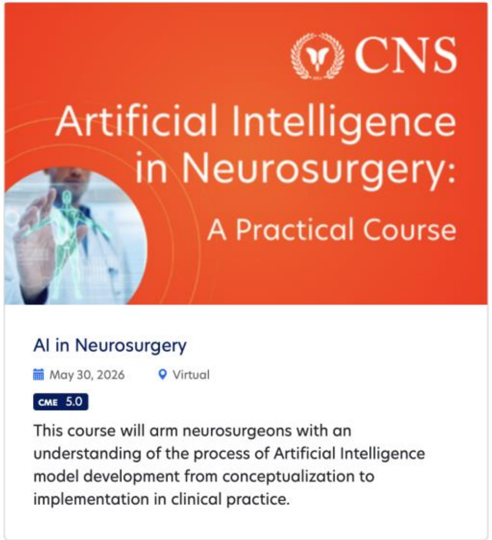
Employment: Merck & Co., Eikon Therapeutics



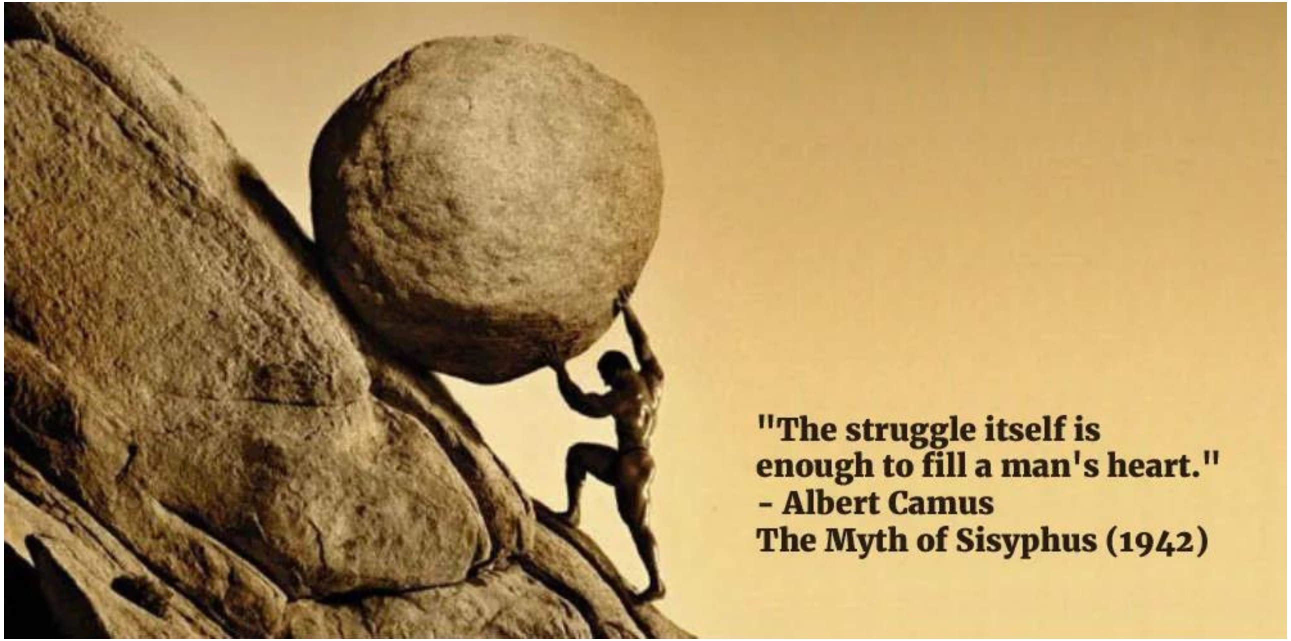
NYU Langone Health

Khoi D. Than

Annual Meeting Chair



Stroke: It’s not where, but when



Neurosurgery
is a gateway for
technology in Medicine

Neurosurgical technology drives ***Medical*** innovation

Stereotactic Coordinates
(Horsley & Clarke, 1908)
3D brain mapping system enabling precise targeting for surgeries and biopsies

Operating Microscope
(Jacobson & Suarez, 1960)
Revolutionized microsurgery, now standard in ophthalmology, ENT, and plastic surgery

Deep Brain Stimulation
(Benabid, 1987)
Neurostimulation treating Parkinson's, epilepsy, OCD, and chronic pain disorders

Radiosurgery
(Leksell, 1960s-70s)
Gamma Knife and CyberKnife are standard of care for a wide range of cancers and much of Radiation Oncology.

Awake Brain Mapping
(Penfield, 1930s-50s)
Real-time functional mapping during surgery, preserving speech and motor function

Ultrasonic Aspirator (CUSA)
(Flamm, 1970s)
Ultrasound-based tissue removal, now used in liver, kidney, and general surgery

Endovascular surgery
(Serbinenko & Guglielmi, 1970s-90s)
Catheter-based stroke treatment revolutionizing interventional cardiology and radiology

Image-Guided Navigation
(Roberts, Bucholz, 1980s-90s)
Computer-assisted surgical guidance, now essential in spine, ENT, and orthopedic surgery

The future…
(**You**)
Brain computer interfaces, artificial intelligence, surgical robotics… there has never been a more exciting time to be a neurosurgeon.

The **writing** of neurosurgical software is the future…

Past or Present?

0148-396X/90/2705-0837$02.00/0
NEUROSURGERY

Vol. 27, No. 5, 1990
Printed in U.S.A.

Technical note

Computer-Aided Intelligence: Application of an Expert System to Brachial Plexus Injuries

Wink S. Fisher III, M.D.

Department of Neurosurgery, Naval Hospital Bethesda, Bethesda, Maryland

When confronted with a patient with a brachial plexus injury, how often as neurosurgeons do we consult an atlas to confirm the anatomy of the brachial plexus and then attempt to establish the location of the lesion? Similar difficulties are encountered with lumbar and sacral plexus lesions. In a project organized to assist the neurosurgeon in this time-consuming task, a computer program that can rapidly determine the site of a lesion in a brachial, lumbar, or sacral plexus injury was created. Using known anatomical pathways (37 clinically relevant upper and 20 lower extremity muscle innervations), and relying solely upon the neurological motor examination, rapid computer-assisted diagnosis is possible. When more than one final common pathway lesion occurs (for example, multiple root avulsions of the brachial plexus), possible lesion sites can be obtained. An interactive dialogue between the user and the program helps to determine the location of the lesion. The program can be run on any IBM-compatible personal computer and is presented as an instrument that provides assistance in cases of complex peripheral nerve injuries, when expert consultants are unavailable. In addition, it can be used as an aid to learning and as a review of basic neuroanatomy. (*Neurosurgery* 27:837–843, 1990)

Key words: Artificial intelligence, Brachial plexus, Computer application, Expert system, Nerve injury

NYU Langone Health

Past or Present?

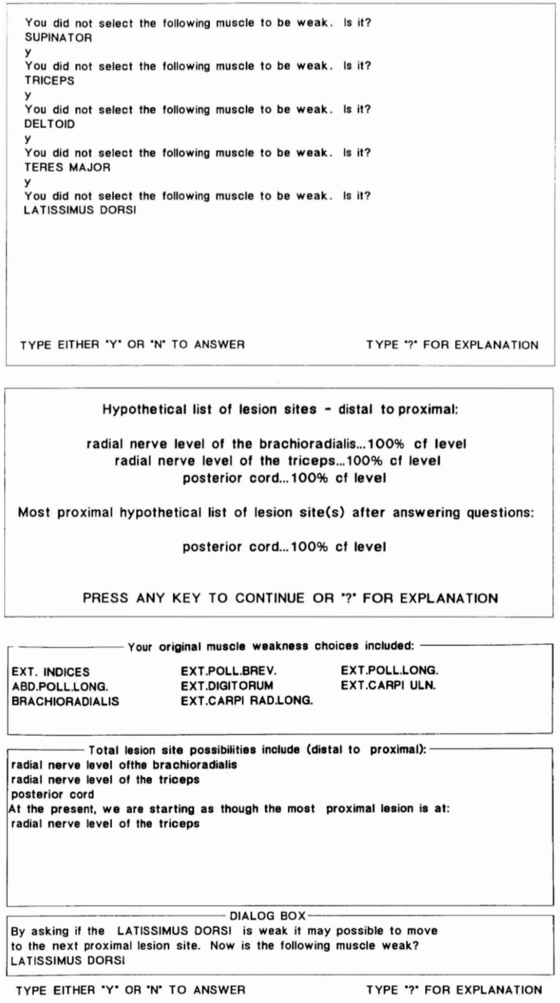
- **Computer assisted diagnosis**
- **Using a Chatbot**
- **Interpretable feedback**
- **… focus on augmented intelligence**

If it said "ChatGPT" this paper could have been written **yesterday**.

It describes ***all of the current trends*** in modern **Medical AI research**.

***Neurosurgeons have been doing medical AI with chatbots for 30 years*.**

NYU Langone Health



Ways for neurosurgeons to get involved in medical AI research (when you can't write code*)

- Defining *clinically relevant* hypotheses, datasets, and evaluations
- *Deployment of AI* of any kind in clinical practice and reporting on *patient outcomes* or studying AI safety in healthcare.
- Incorporating insights from neuroscience & neurosurgery into the fundamentals of AI design
- *I'm coming back to this later

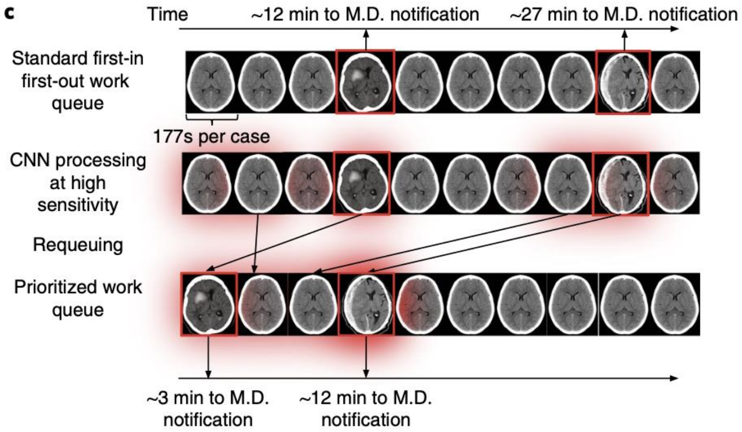
AI in Stroke

nature medicine
LETTERS
https://doi.org/10.1038/s41591-018-0147-y

Automated deep-neural-network surveillance of cranial images for acute neurologic events

Design

- 37,236 head CTs; 3D-CNN with weak NLP-derived labels
- Randomized, double-blinded, prospective trial in a simulated clinical environment
- Compared AI triage vs standard radiology workflow
- Kicked off "AI for radiology workflow improvement" trend for critical neurological diseases



Data: *3D-CNN triage 1.2 s vs radiology workflow 174 s (≈150× speed-up)*

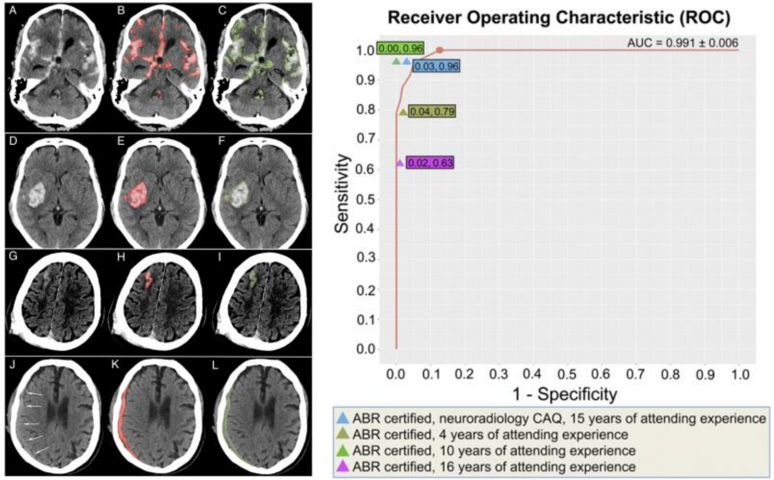
Why it matters: *First evaluation of an AI triage system for acute neurologic findings. The 150× speed-up is the data point that launched a thousand workflow studies.*

Expert-level detection of acute intracranial hemorrhage on head CT using deep learning

Kuo W, Häne C, Mukherjee P, Malik J, Yuh EL. PNAS 2019;116(45):22737–22745

Design

- Single-stage, fully convolutional network (PatchFCN)
- Detection + localization of acute ICH on NCHCT
- Compared against 4 US board-certified radiologists
- UCSF / UC Berkeley collaboration



Data: *AUC 0.991 (deep learning) vs 0.957 (radiologist average)*

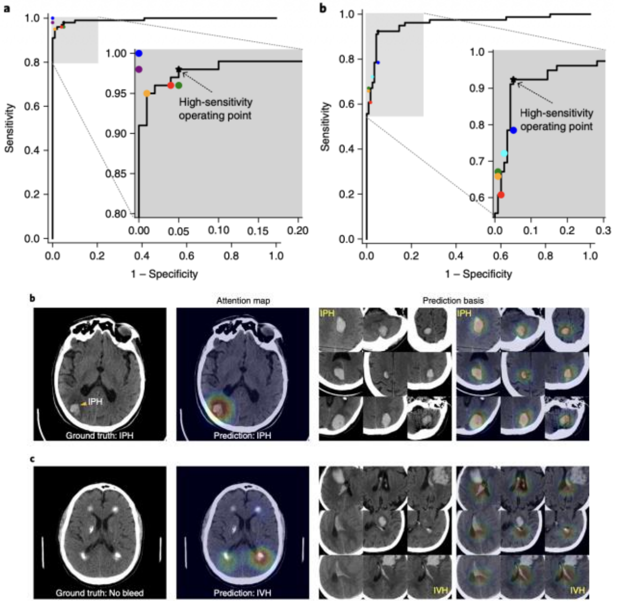
Why it matters: *The paper that made "expert-level" a credible claim for stroke imaging AI — and the first to show DL catching findings humans missed.*

NYU Langone Health

2019 — Explainable AI from Small Datasets

An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets

Lee H, Yune S, Mansouri M, et al. Nature Biomedical Engineering 2019;3:173–185



Design

- Attention-based CNN with atlas-based interpretability maps
- Trained on modest institution-specific data — not millions of images
- Evaluated on retrospective + prospective external datasets
- Tested head-to-head against 4 radiologists

Data: *AUC 0.97 · sensitivity 96 % · specificity 95 % (institution-specific training)*

Why it matters: *Addressed the black-box objection head-on. Most hospitals don't have Mount Sinai-scale data — Lee et al. showed you don't need it.*

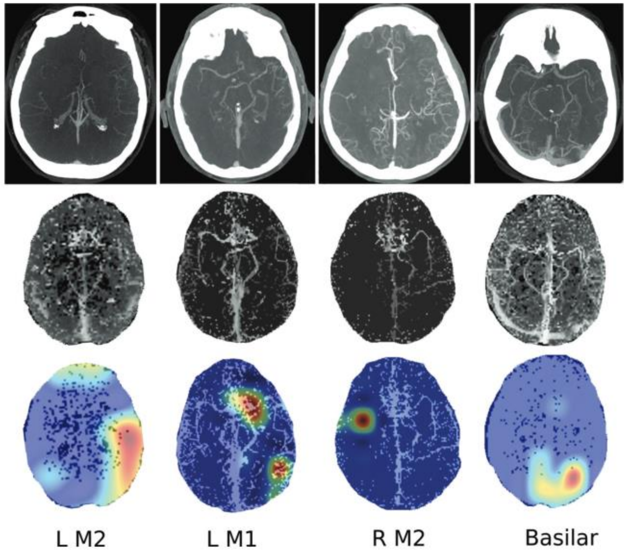
2020 — Deep Learning Finds LVOs on CTA

Detecting large vessel occlusion at multiphase CT angiography by using a deep convolutional neural network

Stib MT, Vasquez J, ..., McTaggart RA. Radiology 2020;297(3):640–649

Design

- 540 CTAs (270 LVO+, 270 LVO-) from a multicenter cohort
- Multiphase CTA (arterial, peak venous, late venous)
- CNN with MIP-based pre-processing of contrast-enhanced vasculature
- Validated against catheter angiography gold standard



Data: *Sensitivity by phase combination: single 0.77 · two 0.86 · three 1.00*

Why it matters: *Multiphase beats single-phase CTA for AI detection of LVO, just as it does for human readers. A reminder that AI is only as good as the imaging pipeline feeding it.*

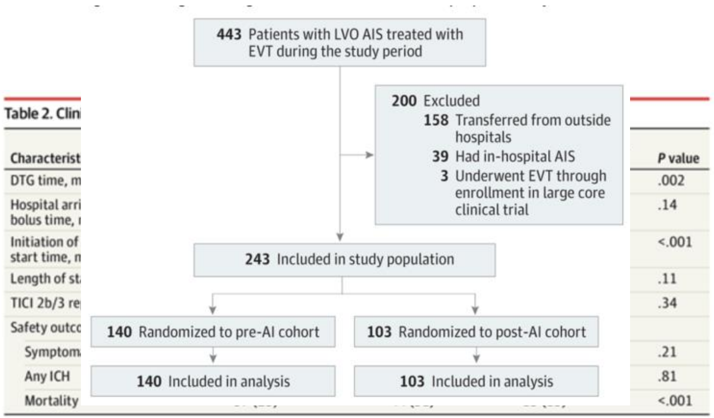
2023 — Cluster RCT: AI Improves Workflow

Automated large vessel occlusion detection software and thrombectomy treatment times: a cluster randomized clinical trial

Martinez-Gutierrez JC, Kim Y, Salazar-Marioni S, ..., Sheth SA.
JAMA Neurol 2023;80:1182–1190

Design

- Cluster randomized, stepped-wedge design
- 4 comprehensive stroke centers in greater Houston
- 243 LVO patients treated with EVT over 12 months
- Viz.ai automated detection + secure group messaging



Implementation of the AI algorithm was associated with a reduction in DTG time by 11.2 minutes

Why it matters: *The first randomized trial to show AI-driven workflow reduces time to thrombectomy. 11 minutes saved ≈ 11 weeks of disability-free life per patient (Stib's rule of thumb).*

2023 — Clinical LLMs at Health-System Scale

Article

Health system-scale language models are all-purpose prediction engines

https://doi.org/10.1038/s41586-023-06160-y

Received: 14 October 2022

Accepted: 2 May 2023

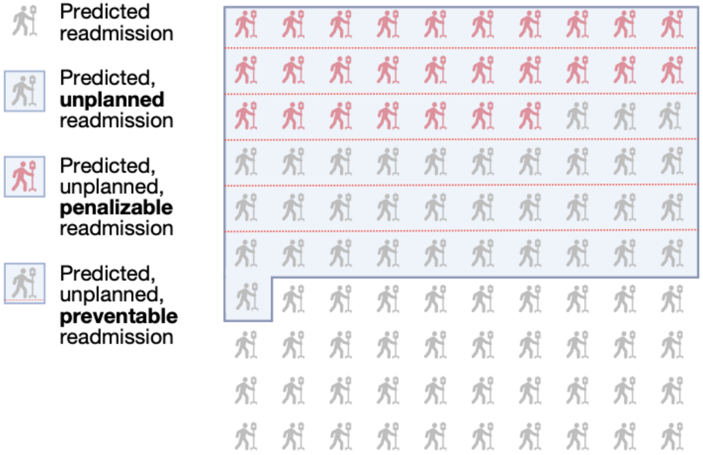
Published online: 7 June 2023

Open access

Lavender Yao Jiang[1,2], Xujin Chris Liu[1,3], Nima Pour Nejatian[4], Mustafa Nasir-Moin[1], Duo Wang[5], Anas Abidin[4], Kevin Eaton[6], Howard Antony Riina[1], Ilya Laufer[1], Paawan Punjabi[6], Madeline Miceli[6], Nora C. Kim[1], Cordelia Orillac[1], Zane Schnurman[1], Christopher Livia[1], Hannah Weiss[1], David Kurland[1], Sean Neifert[1], Yosef Dastagirzada[1], Douglas Kondziolka[1], Alexander T. M. Cheung[1], Grace Yang[1,2], Ming Cao[1,2], Mona Flores[4], Anthony B. Costa[4], Yindalon Aphinyanaphongs[5,7], Kyunghyun Cho[2,8,9,10] & Eric Karl Oermann[1,2,11]

Design

- NYUTron — 109 M-parameter BERT pre-trained on 4.1 B words of NYU clinical notes (387 k patients)
- Fine-tuned on 5 tasks (readmit, mortality, LOS, comorbidity, denial)
- Prospective deployment at 6 NYU Langone inpatient sites
- First clinical LLM integrated into production EHR at this scale



Data: NYUTron had an AUC of 78.70% in a prospective, single-arm, non-interventional trial with recall of 82.3% and precision of 20.6%.

Why it matters: *Unstructured notes beat structured data on the outcomes hospitals actually track. For stroke, this means readmission risk, complications, and disposition can be predicted from the H&P alone.*

Defining clinically relevant hypotheses

Sully Chen

nature medicine

Large language models in clinical medicine: an LLM-powered large-scale systematic review

Sully F. Chen, Anton Alyakin, Andreas Seas, Eunice Yang, Joanne J. Choi, Jin Vivian Lee, Amelia Chen, Pranav Warman, Rochelle Bitolas, Robert J. Steele, Daniel Alber, Eric K. Oermann

Accepted for publication

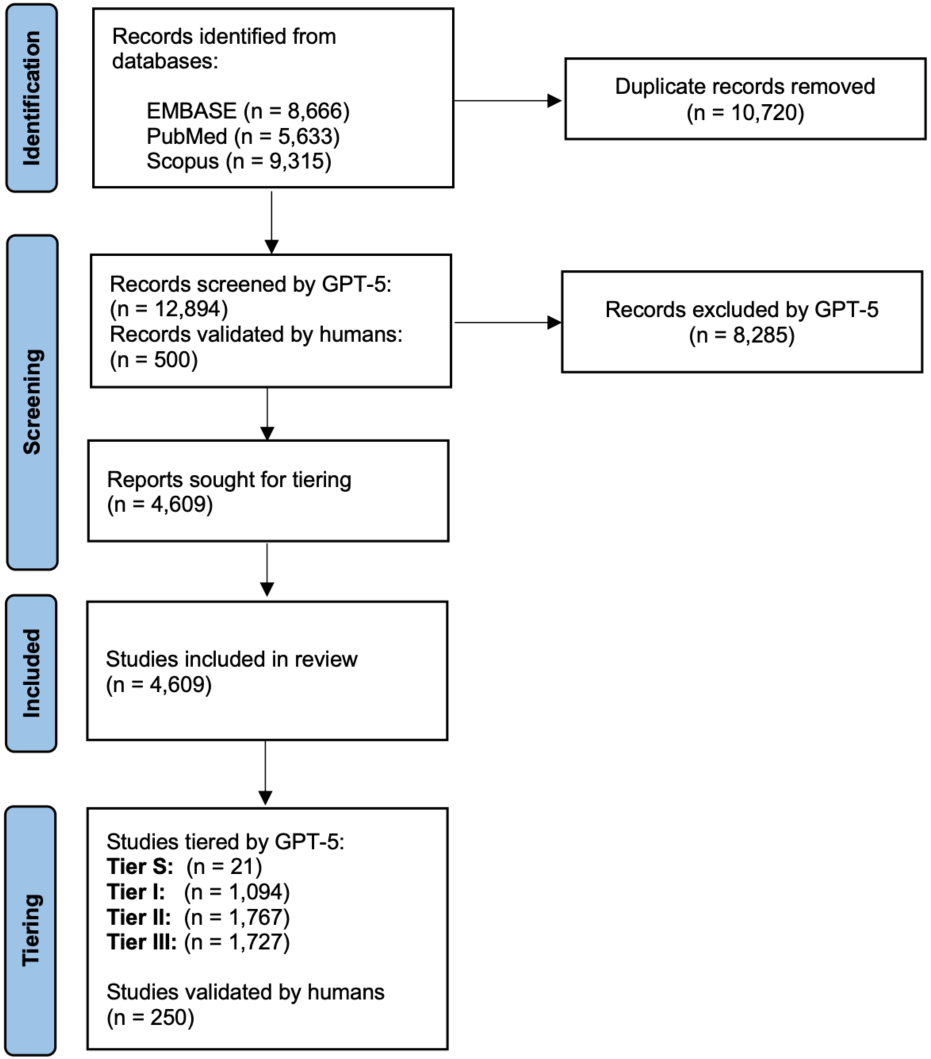
The reality of our field (artificial intelligence)…

Since 2020, ~5,000 papers published on *just* LLMs in Medicine

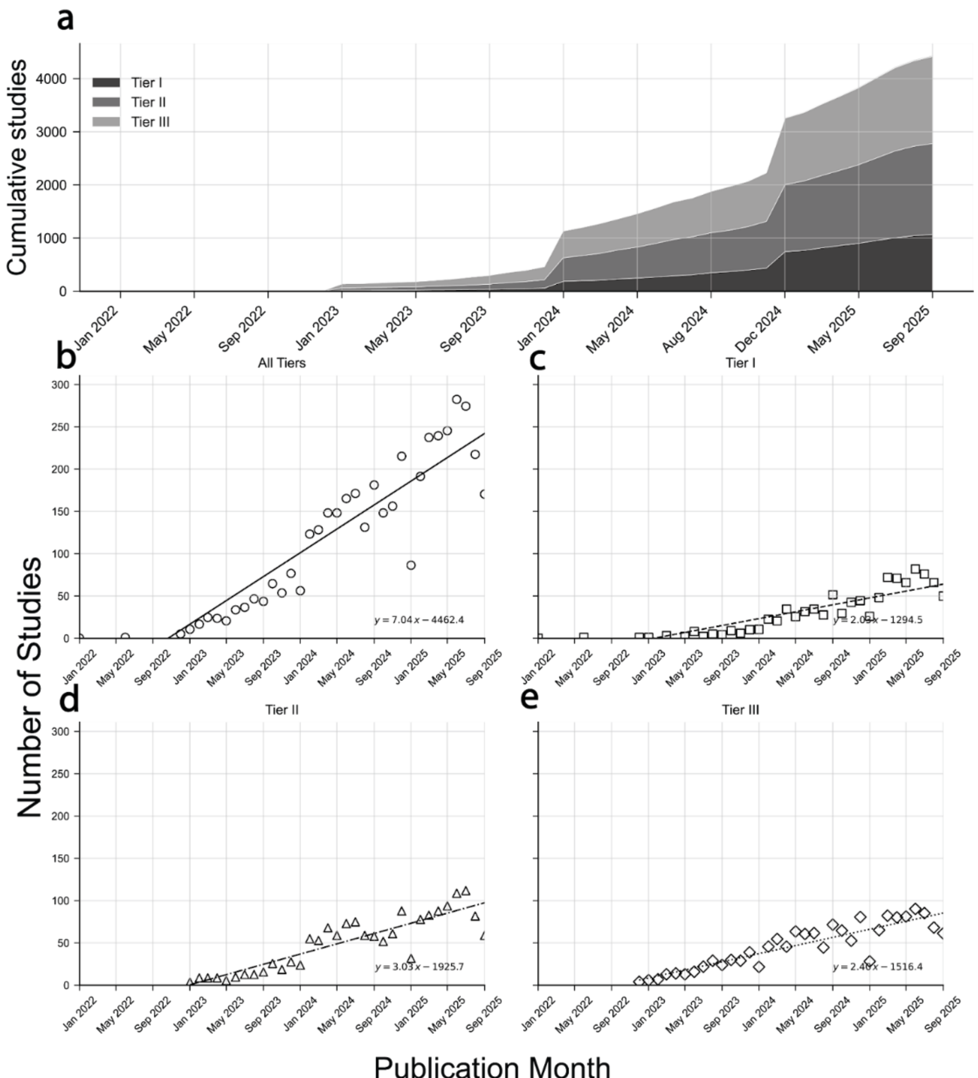
Majority test on multiple choice medical exams

Very few *studies* generated actual medical evidence

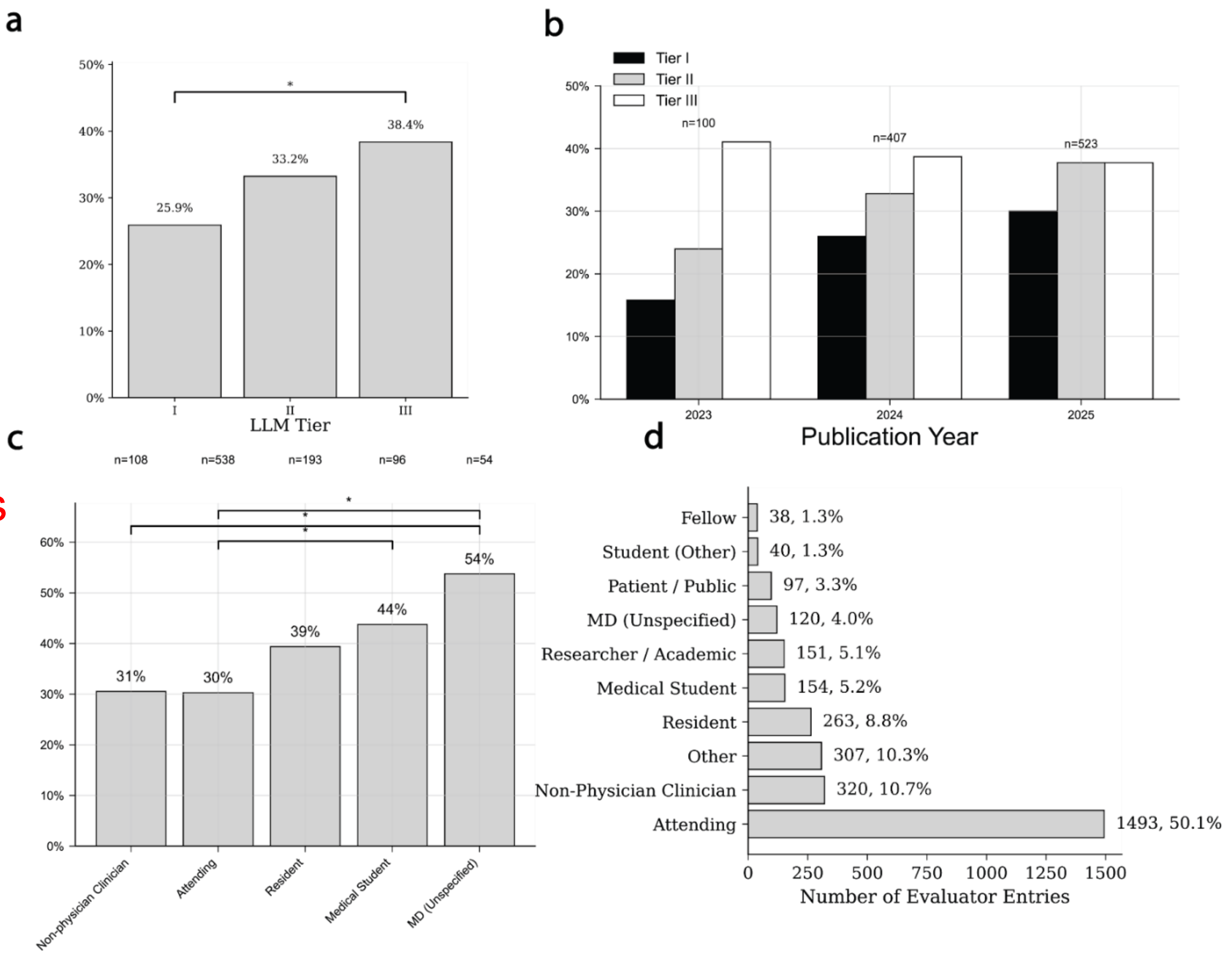
NYU Langone Health


Identification
Records identified from databases:
EMBASE (n = 8,666)
PubMed (n = 5,633)
Scopus (n = 9,315)
Duplicate records removed (n = 10,720)
Screening
Records screened by GPT-5: (n = 12,894)
Records validated by humans: (n = 500)
Records excluded by GPT-5 (n = 8,285)
Reports sought for tiering (n = 4,609)
Included
Studies included in review (n = 4,609)
Tiering
Studies tiered by GPT-5:
Tier S: (n = 21)
Tier I: (n = 1,094)
Tier II: (n = 1,767)
Tier III: (n = 1,727)
Studies validated by humans (n = 250)

4 papers are
published *per day*



NYU Langone Health



LLM’s are less impressive when compared against attendings

NYU Langone Health

Nandan Lad

PLOS One

PLOS One

Publish About Browse

Search

advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Large-scale multi-omic biosequence transformers for modeling protein–nucleic acid interactions

Sully F. Chen, Robert J. Steele, Glen M. Hocky, Beakal Lemeneh, Shivanand P. Lad, Eric K. Oermann

Published: February 2, 2026 • https://doi.org/10.1371/journal.pone.0341501

scientific reports

Explore content ⌄ About the journal ⌄ Publish with us ⌄

Shrutika Singh

nature › scientific reports › articles › article

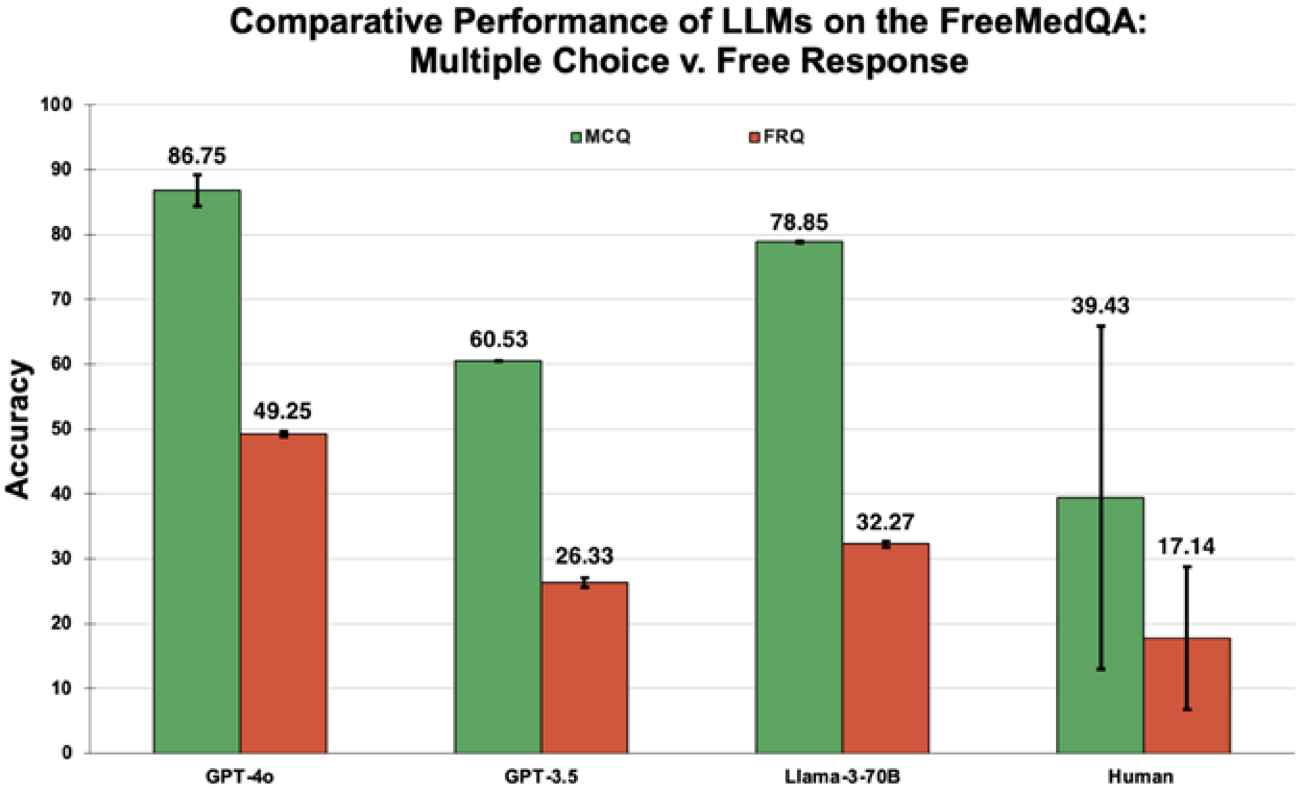
Article | Open access | Published: 26 November 2025

The pitfalls of multiple-choice questions in generative AI and medical education

Shrutika Singh, Anton Alyakin, Daniel Alexander Alber, Jaden Stryker, Ai Phuong S. Tong, Karl Sangwon, Nicolas Goff, Mathew De La Paz, Miguel Hernandez-Rovira, Ki Yun Park, Eric Claude Leuthardt & Eric Karl Oermann

Scientific Reports **15**, Article number: 42096 (2025) | Cite this article

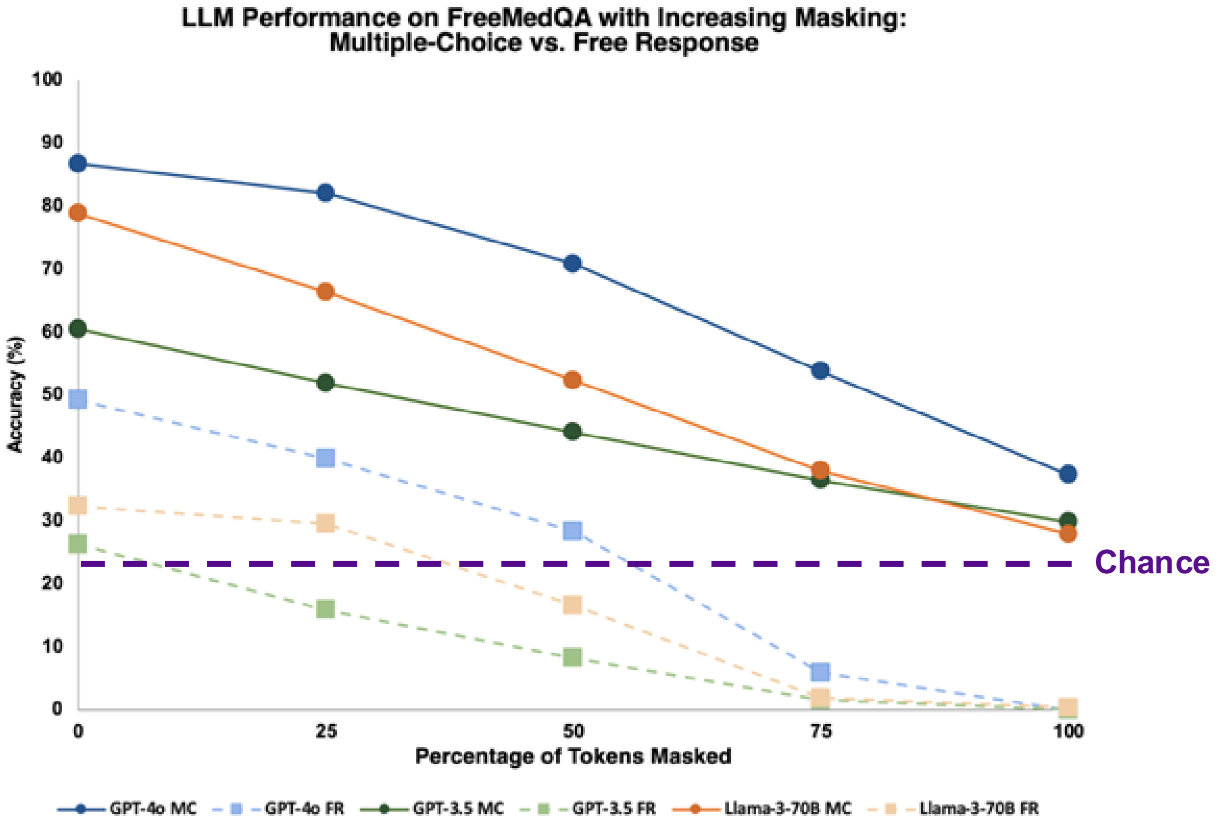
Does it even make sense to use medical exams in modern era?



Does it even make sense to use medical exams in modern era?

(1) Memorized in training data?

(2) Exploiting prior in answer distribution?



NYU Langone Health


Katie E. Link

nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature communications › articles › article

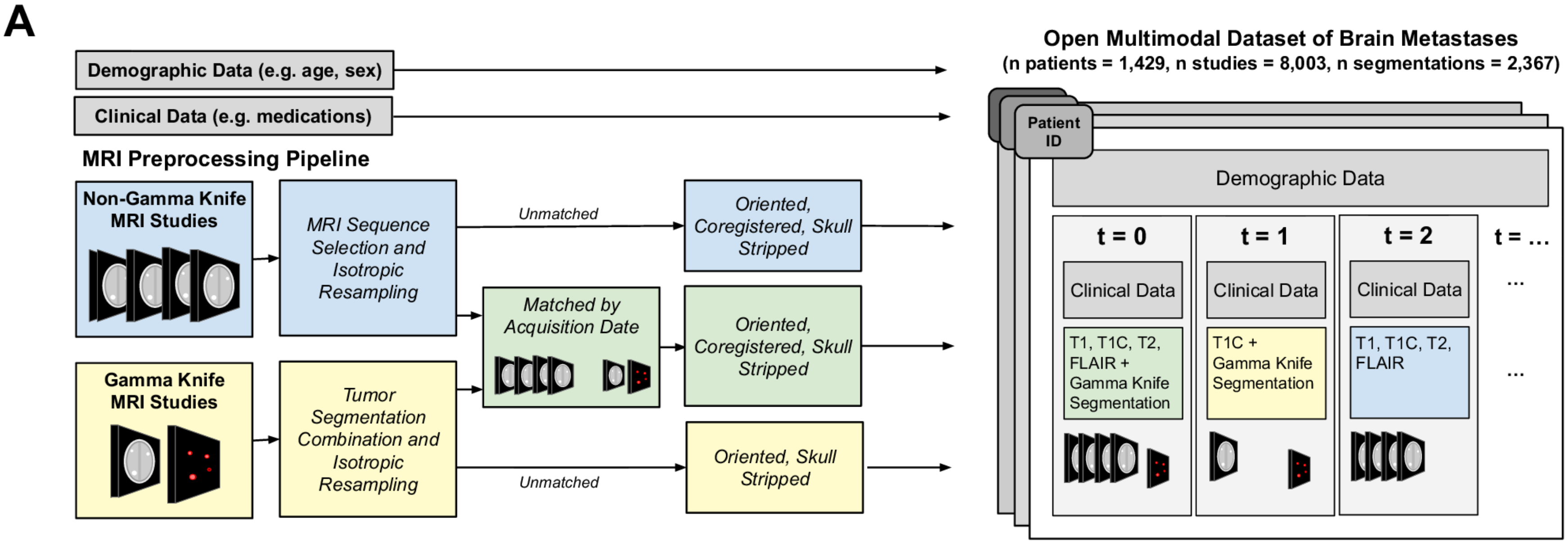
Article | Open access | Published: 17 September 2024

Longitudinal deep neural networks for assessing metastatic brain cancer on a large open benchmark

Katherine E. Link, Zane Schnurman, Chris Liu, Young Joon (Fred) Kwon, Lavender Yao Jiang, Mustafa Nasir-Moin, Sean Neifert, Juan Diego Alzate, Kenneth Bernstein, Tanxia Qu, Viola Chen, Eunice Yang, John G. Golfinos, Daniel Orringer, Douglas Kondziolka & Eric Karl Oermann

Nature Communications **15**, Article number: 8170 (2024) | Cite this article

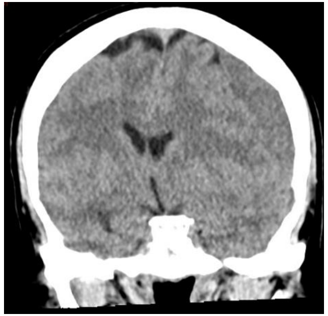
Mis-specification of inputs



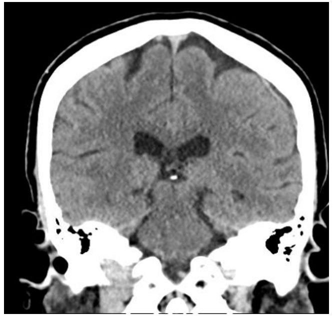
Mis-specification of inputs

Ignoring longitudinal data frequently masks ground truth conditional probabilities.

How we normally label our datasets



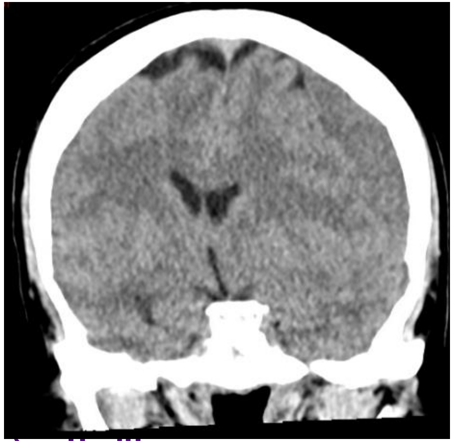
p(x_1) = surgery



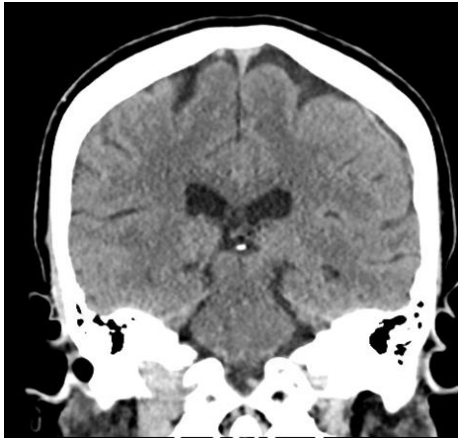
p(x_2) = home

Mis-specification of inputs

How these images are *actually* interpreted as ground truth



4 week
follow-up

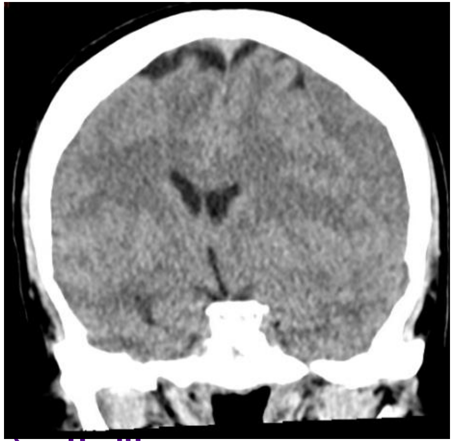


p(x_2|x_1) = home

Mis-specification of inputs

Health is dynamic – health data should be dynamic.

We *think* about patients over time.



4 week
follow-up

p(x_1|x_1) = home

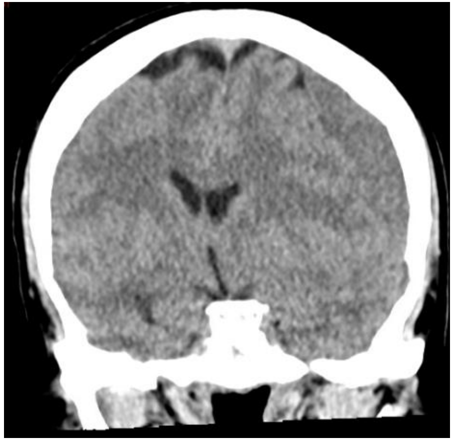
Mis-specification of inputs

Health is dynamic – health data should be dynamic.

We *think* about patients over time.

Conditional probability of clinical event is determined by sequence order rather than image contents

4 week follow-up



$$= \quad p(x\_1|x\_2) = \text{surgery}$$

Base Rate Fallacy in Prior Distribution

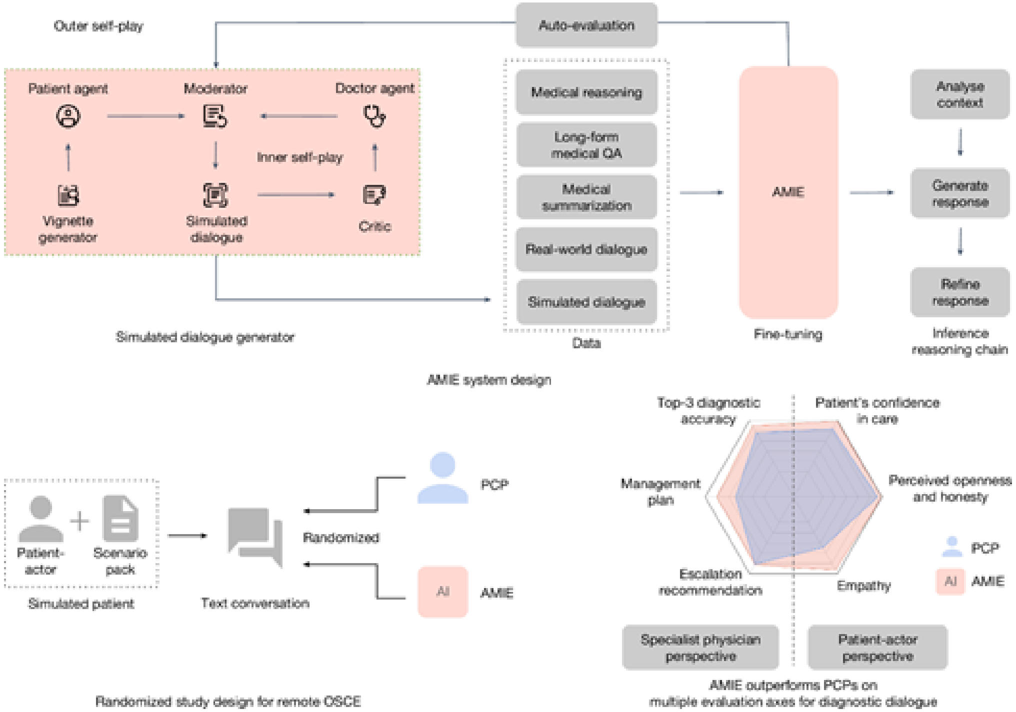
Article

Towards conversational diagnostic artificial intelligence

https://doi.org/10.1038/s41586-025-08866-7
Received: 18 January 2024
Accepted: 5 March 2025
Published online: 9 April 2025
Open access

Tao Tu[1,3] ✉, Mike Schaekermann[1,3] ✉, Anil Palepu[1,3], Khaled Saab[1], Jan Freyberg[1], Ryutaro Tanno[2], Amy Wang[1], Brenna Li[1], Mohamed Amin[1], Yong Cheng[2], Elahe Vedadi[1], Nenad Tomasev[2], Shekoofeh Azizi[2], Karan Singhal[1], Le Hou[1], Albert Webson[2], Kavita Kulkarni[1], S. Sara Mahdavi[2], Christopher Semturs[1], Juraj Gottweis[1], Joelle Barral[2], Katherine Chou[1], Greg S. Corrado[1], Yossi Matias[1], Alan Karthikesalingam[1,4] ✉ & Vivek Natarajan[1,4] ✉

- **More real world – FRQ with multi-turn dialogue…**
- **Is this an AI physician?**
  - Chatbot actually *chats with patients* to arrive at a diagnosis!
- **Compared to *real doctors* using *professional actors*.**



NYU Langone Health

Base Rate Fallacy in Prior Distribution

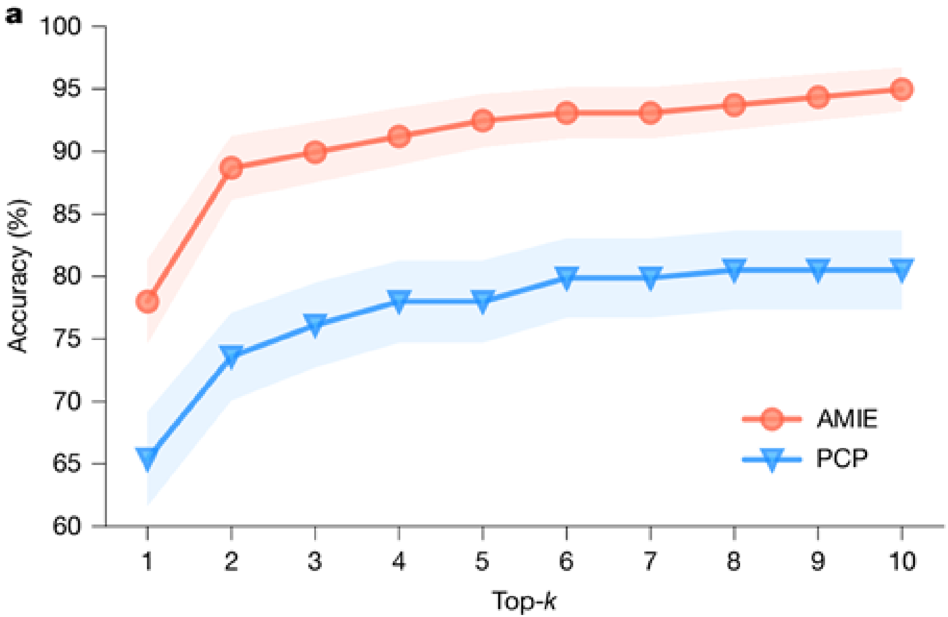
But what about the distribution of priors?

- **Better than doctors…**
- **But really?**
- **Massive prior shift b/c *not really being used in production* and the *synthetic problem setup is wrong*.**

– **159 simulated case scenarios (OSCE)**
  - **149 positive cases**
  - **10 negative cases**

p(disease_test) = 0.94 !!!!

NYU Langone Health



Base Rate Fallacy in Prior Distribution

- **Neurological: most common dx in primary care are headaches, back pain, and sleep disorders.**
  - Eval dataset distributions:
    - Low back pain: 0%
    - Headache disorders: 13%
    - GCA: 6.6%
  - Real world distributions:
    - Low back pain: 28%
    - Headache disorders: 16%
    - GCA: 0.01%

Is this a model that does great on rare diseases?

Or a model that will have an unacceptable false positive rate in the real world?

Table SI.10 | Neurology scenarios (N=30).

Location	Ground truth
Canada	Alzheimer's dementia
	Cluster headache
	Delirium tremans
	Essential Tremor
	Giant cell arteritis on background of polymyalgia rheumatica
	Hemorrhagic stroke
	Meningitis
	Migraine
	Migraine
	Multiple sclerosis with optic neuritis
	Optic neuritis which can be associated with multiple sclerosis
	Parkinson's disease
	Space occupying lesion (brain tumour)
	Temporal arteritis +/- Polymyalgia rheumatica
	Trigeminal neuralgia
India	Carpal tunnel syndrome
	Cavernous sinus thrombosis
	Dermatomyositis
	Dystonia
	Lead neuropathy
	Migraine
	Myasthenia gravis
	Neurocysticercosis
	Optic neuritis
	Parkinson's disease
	Posterior reversible encephalopathy syndrome (PRES)
	Tabes dorsalis
	Transient ischemic attack
	Trigeminal Neuralgia
	Wernicke's encephalopathy

Take-Home Points

1

Ischemic stroke care has fundamentally changed.

2026 AHA/ASA: tenecteplase as preferred lytic; thrombectomy extended to large cores and basilar occlusion.

2

We need real clinical research with AI technologies

RCTs, Cohort and Case Control Studies, etc…

3

AI is already in the workflow — evaluate it like any other device.

Platforms differ by 20+ points of sensitivity head-to-head. Base-rate awareness is essential.

4

The next decade is real ROI which will require physician leadership

Clinical LLMs on free text + streaming updates + federated merge — built for a field whose guidelines keep moving.

Eric K. Oermann, MD | *OLAB, NYU Langone Health* | *@ekoermann*

Thank you to all of our wonderful team and collaborators!

OLAB

Eric Karl Oermann, M.D.
Associate Professor of Neurosurgery, Radiology, and Data Science

Lavender Yao Jiang, B.S.
PhD Student
NYU Center for Data Science
(Prev. CMU)

Chris Xujin Liu, B.S.
PhD Student
NYU Tandon School of Engineering
(Prev. CMU)

Radhika Dua, B.S.
PhD Student
NYU Center for Data Science
(Prev. Google)

Ioana Marinescu, B.S.
PhD Student
NYU Center for Data Science
(Prev. Princeton)

Qingyang Zhu, B.S.
PhD Student
NYU Center for Data Science
(Prev. Shanghai Tech and Alibaba)

Frederick Kwon, M.D., Ph.D.
Resident in Radiology
(Prev. ISMMS)

STUDENTS

Anton Alyakin (Visiting)
Mustafa Nasir-Moin
Karl Sangwon Lee
Daniel Alber
Grace Von Oiste
Ashley Amend Thompson
Eunice Yang
Michelle Murata
Gavin Yang

Visiting Scholars

Sumedha Rai, M.S. (Kepos Capital)
Jan Niehues, Ph.D. (U Aachen)
Beakal Lemeneh, M.S. (Blackrock)
Paz Vives, M.S. (JP Morgan)

Perception and Brain Dynamics Laboratory
Biyu He, Ph.D.
Ayaka Hachisuka, B.S.
Jonathan Shor, M.S.

NYU Neurosurgery

Douglas Kondziolka, M.D.
John Golfinos, M.D.
Daniel Orringer, M.D.
Howard Riina, M.D.
Shannon Ciprut
Allison George Lucas
Qian Qian Du
Sean Neifert
Aly Valliani

Neurodata Hub

Xu Han, M.S
Alvin Pan, M.S.
Jaden Stryker, B.S.

NYU OSR
Dafna Bar-Sagi, Ph.D.

Kyunghyun Cho, Ph.D.
Professor of Data Science and Computer Science

NVIDIA

Anthony Costa, Ph.D.
Mona Flores, M.D.
Katherine E. Link, B.S.


NYU MCIT
Nader Mherabi, CIO NYULH
Yindalon Aphinyanaphongs, M.D.,Ph.D.
Jeff Zhang, M.S.

OTTO

NYU Langone Health
NYU Center for Data Science
NYU Neuroscience
Carl Feinberg
W. M. KECK FOUNDATION
NIH
NATIONAL CANCER INSTITUTE
NYU Langone Radiology—Center for Biomedical Imaging