



Artificial Intelligence and Data Science Year in Review

Yuan Luo, PhD, FACMI, FAMIA, FIAHSI

Northwestern University

@yuanhypnosluo

Marylyn Ritchie, PhD, FACMI

University of Pennsylvania

@MarylynRitchie

Disclosure



Industry: Walmart (Luo)

Funding: NIH (Luo, Ritchie)

The content is solely our responsibility and does not represent the official views of our institutions or the NIH.

Copyrights are reserved by the authors who provided the slides that we adapted from.

Learning Objectives



After participating in this session, the learner should be better able to:

- Have a sense of the scope of AI and Data Science for health care
- Summarize the state-of-the-art research in AI and Data Science for health care
- Identify research and development opportunities in this space and advance the field

Process and caveats



The content is developed by the informatics community, for the informatics community, we are not boiling the ocean

Teamwork makes dreamwork, we thank talented volunteers including Jenny Ding, Rachit Kumar, Rasika Venkatesh, Saki Amagai, Yanbo Feng, Yikuan Li and generous help from Jim Cimino

Each volunteer picked a topic, for which they were asked to nominate papers that are significant (new model, new insights, new task)

Focused on original research of AI and data science for health care, exclude reviews, viewpoints ...

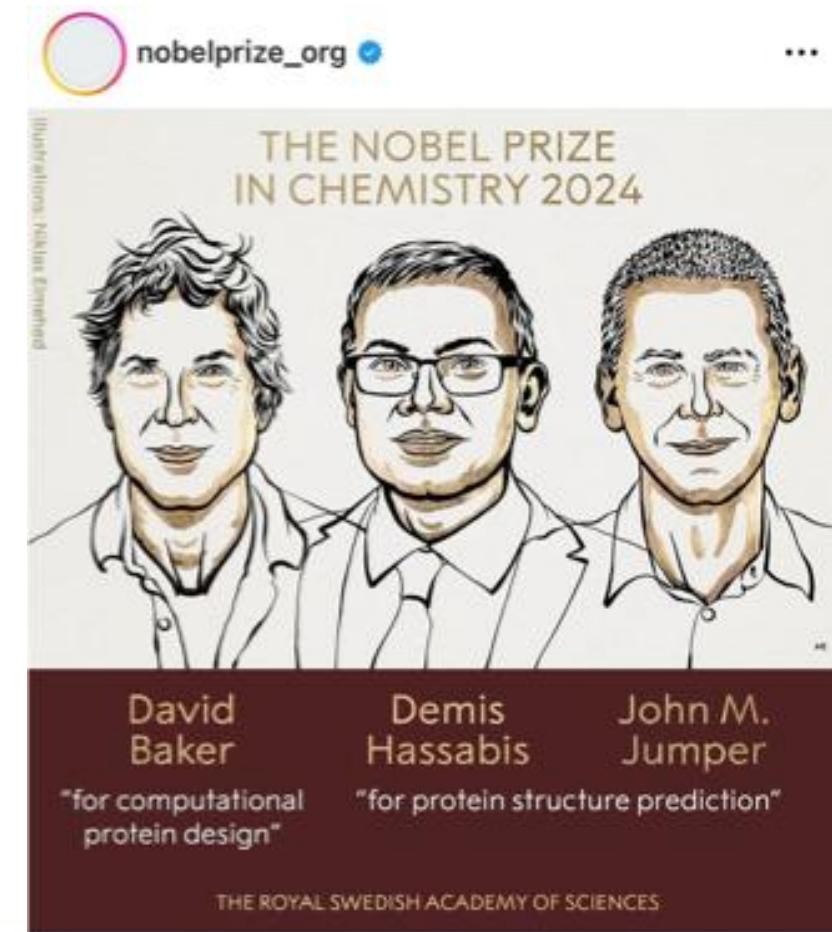
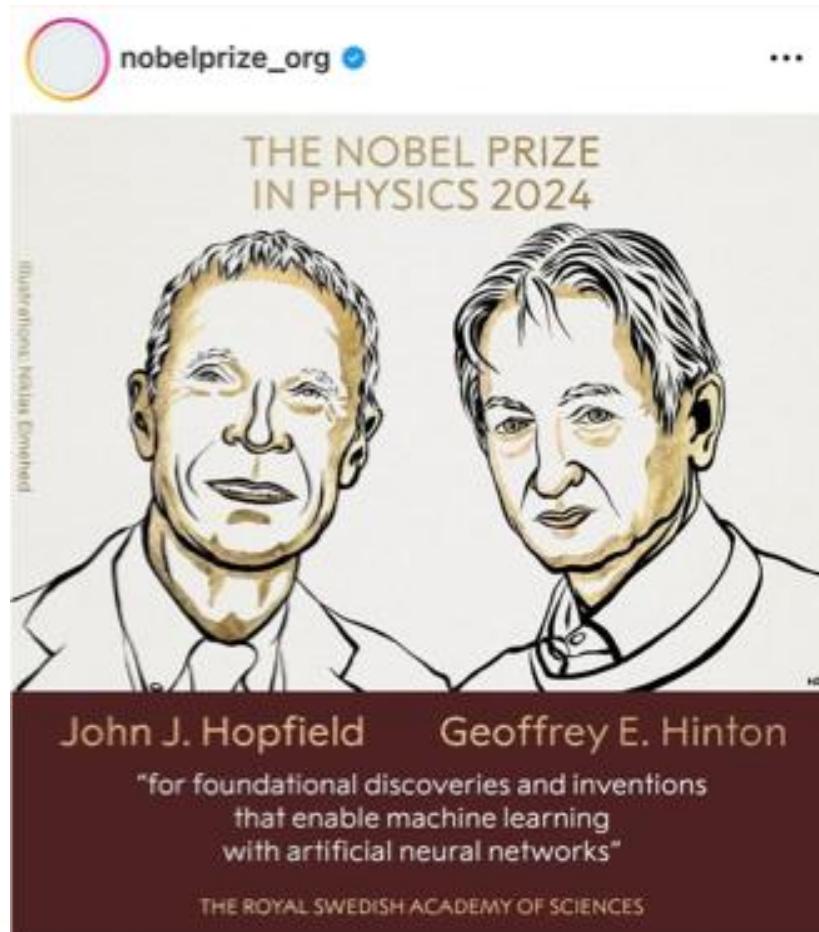
Audiovisual materials requested from authors for "interesting" papers, all content from authors or original papers

Marylyn and Yuan finalized the selection and organized the slides to divide-and-conquer the presentation

Apologies for anything missed or misjudged, mistakes are all our responsibility

AI had a BIG YEAR in 2024

Two of the 2024 Nobel Prizes are for discoveries in AI

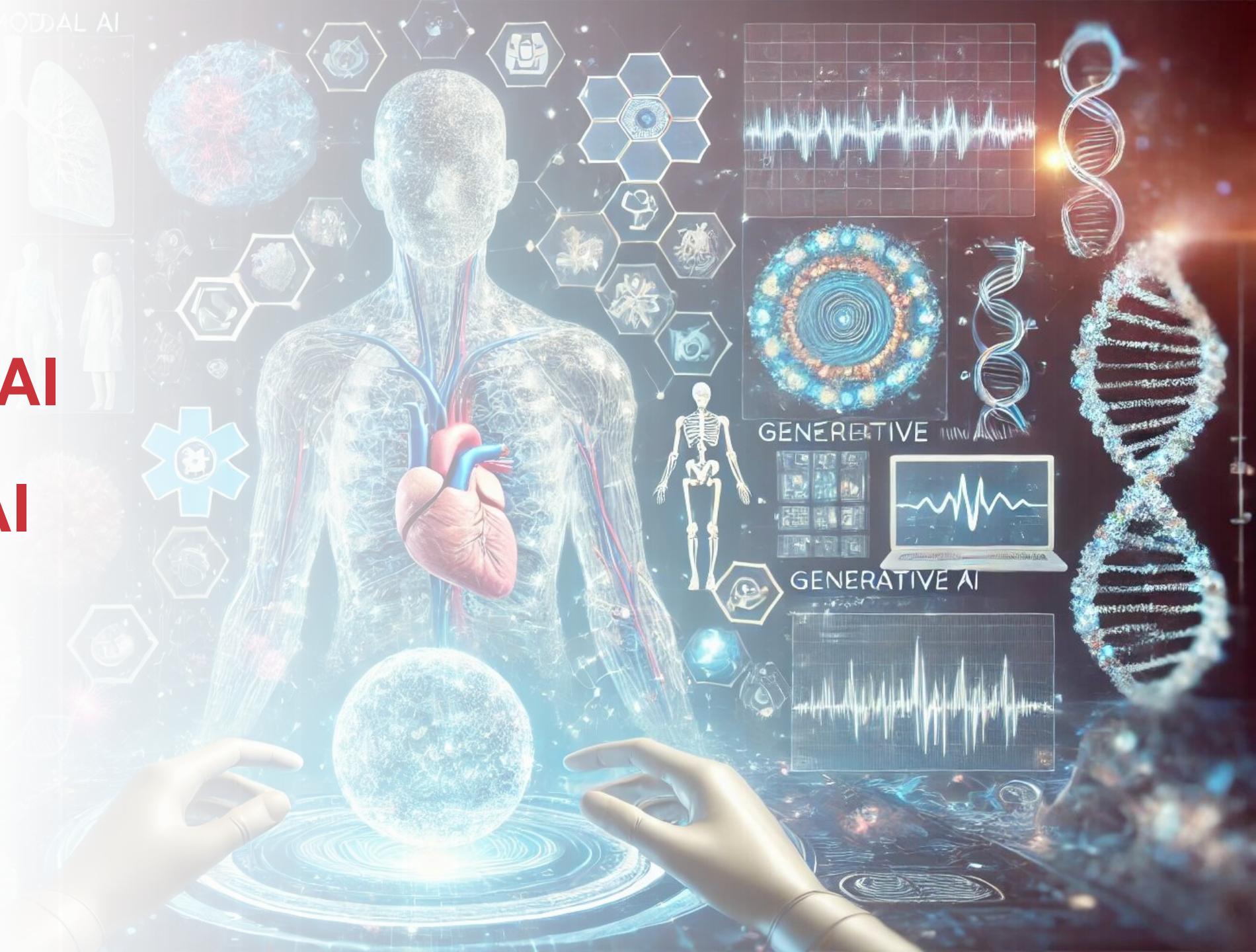


Topics



- Multi-modal AI and Generative AI
- Bias, Trust, and ELSI (Ethical, Legal, and Social Issues) in AI
- Natural Language Processing and Large Language Models
- Computer Vision Applications in Informatics
- Clinical Decision Support, Evaluation and Implementation
- Public/global Health and Precision Medicine Informatics

Multi-modal AI and Generative AI



Original Investigation | Oncology



October 23, 2024

Performance of Multimodal Artificial Intelligence Chatbots Evaluated on Clinical Oncology Cases

David Chen, BMS^{1,2}; Ryan S. Huang, MSc^{1,2}; Jane Jomy, MSc^{1,2}; Philip Wong, MD, MSc^{1,3};
Michael Yan, MD, MPH^{1,3}; Jennifer Croke, MD, MHPE^{1,3}; Daniel Tong, MD^{1,2,3}; Andrew Hope, MD^{1,3};
Lawson Eng^{4,5}; Srinivas Raman, MD, MASc^{1,4}

[» Author Affiliations](#) | [Article Information](#)

JAMA Netw Open. 2024;7(10):e2437711. doi:10.1001/jamanetworkopen.2024.37711



Performance of Multimodal Artificial Intelligence Chatbots Evaluated on Clinical Oncology Cases

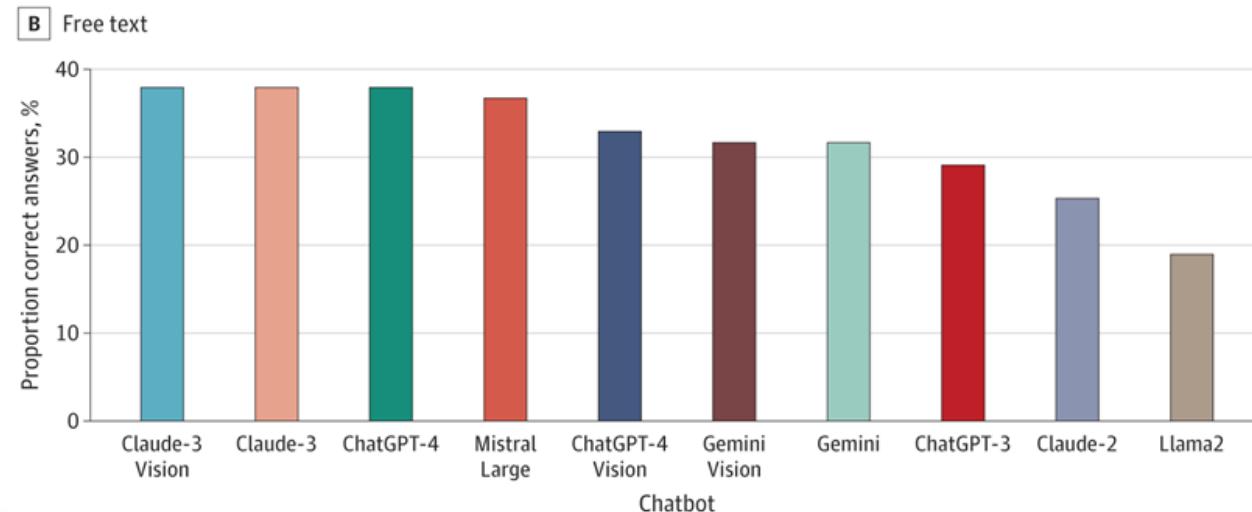
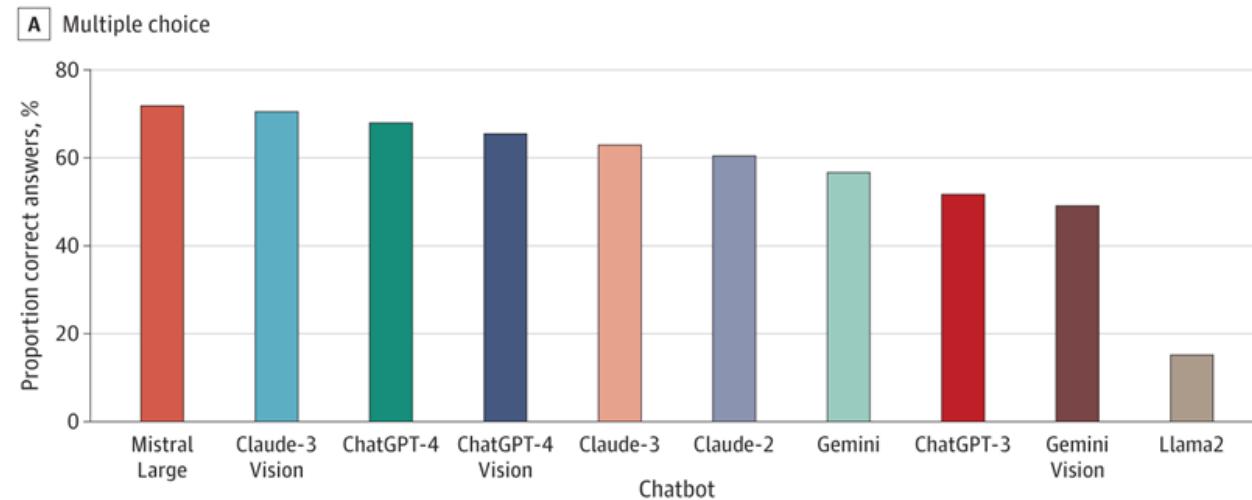


Goal	<ul style="list-style-type: none">Compare multimodal artificial intelligence chatbots compared with text-only chatbots when evaluated on diagnostic and management questions about clinical oncology cases
Dataset	<ul style="list-style-type: none">79 clinical oncology cases (text and image) from JAMA Network LearningMultiple-choice responses were marked as correct based on the ground-truth, correct answer.Free-text responses were rated by a team of oncology specialists in duplicate and marked as correct based on consensus or resolved by a review of a third oncology specialist.
Analysis	<ul style="list-style-type: none">Medical accuracy evaluated by the number of correct responses by each AI chatbot
Improvements on prior work	<ul style="list-style-type: none">Focused on more complex subject matter (clinical oncology cases)Benchmarked 10 diverse chatbots



Multimodal chatbots were not consistently more accurate than unimodal chatbots on clinical oncology cases

Note: Vision denotes multimodal chatbots with text and image input





CURRENT ISSUE ▾ RECENTLY PUBLISHED PODCAST EVENTS ABOUT ▾ AUTHOR CENTER PUBLICATIONS ▾

ORIGINAL ARTICLE

A Multimodal Biomedical Foundation Model Trained from Fifteen Million Image–Text Pairs

Authors: Sheng Zhang, Ph.D. , Yanbo Xu, Ph.D. , Naoto Usuyama, M.S. , Hanwen Xu, M.S. , Jaspreet Bagga, B.S. , Robert Tinn, M.E. , Sam Preston, Ph.D. , , and Hoifung Poon, Ph.D. [Author Info & Affiliations](#)

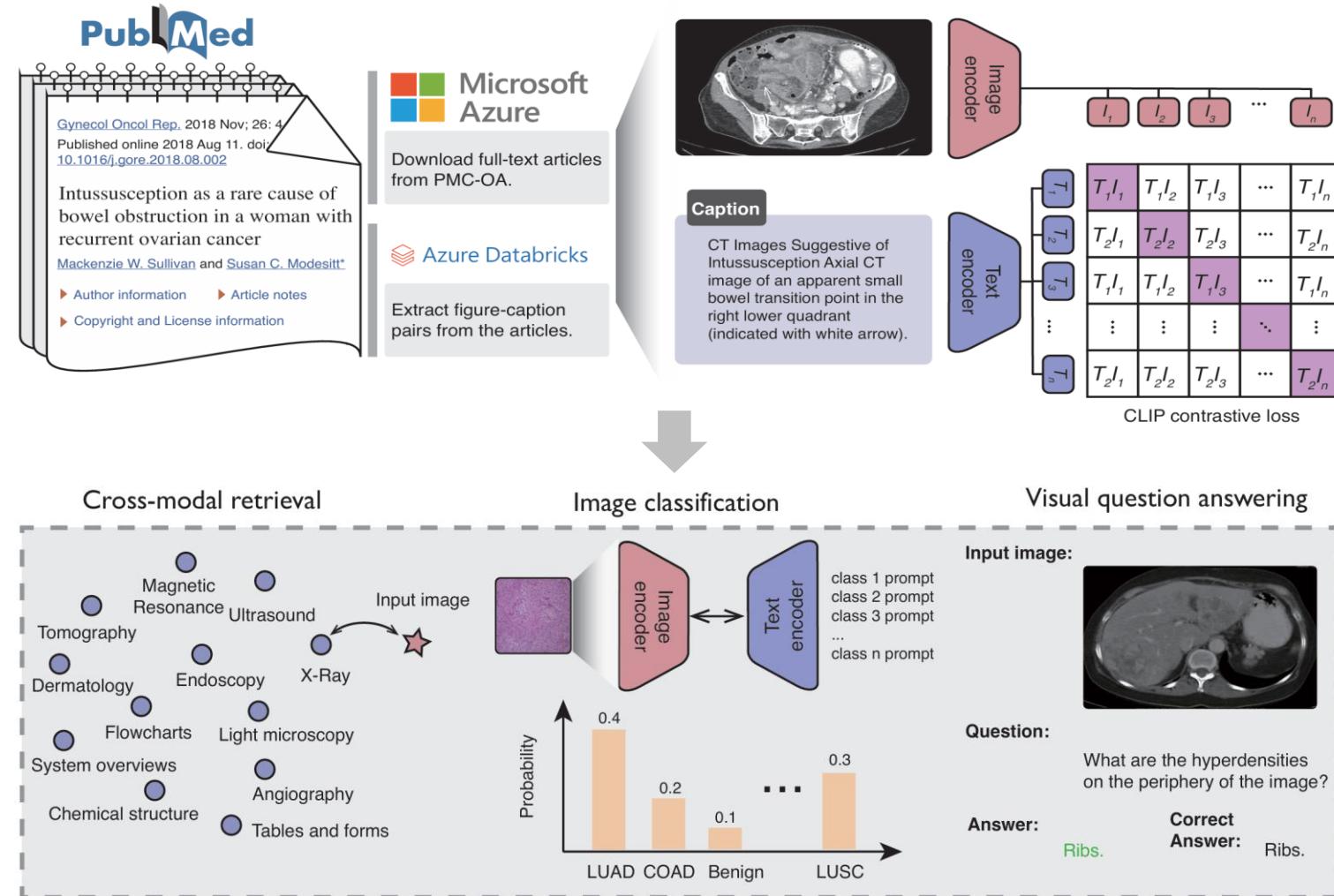
Published December 20, 2024 | NEJM AI 2025;2(1) | DOI: 10.1056/Aloa2400640 | [VOL. 2 NO. 1](#)

[Copyright © 2024](#)



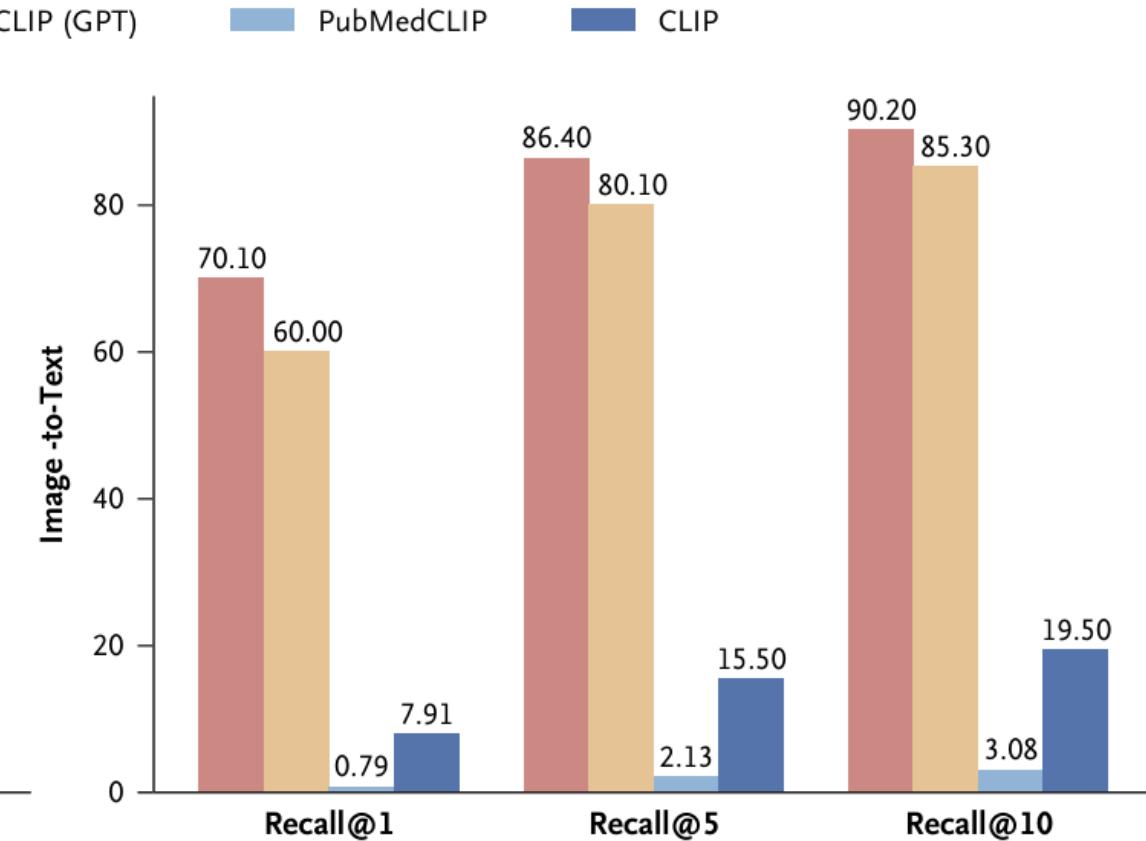
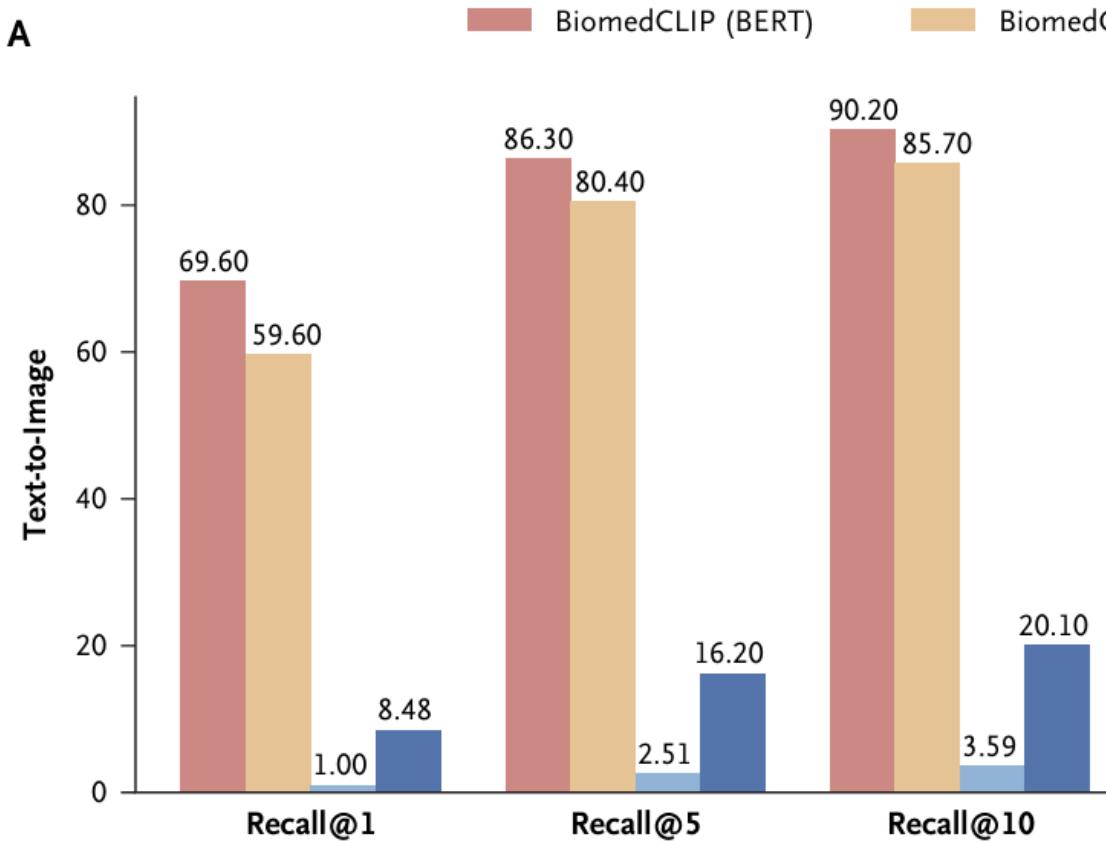
A Multimodal Biomedical Foundation Model Trained from Fifteen Million Image–Text Pairs

BiomedCLIP excels on a variety of downstream applications



Results

A



Development and Validation of a Multimodal Multitask Vision Foundation Model for Generalist Ophthalmic Artificial Intelligence

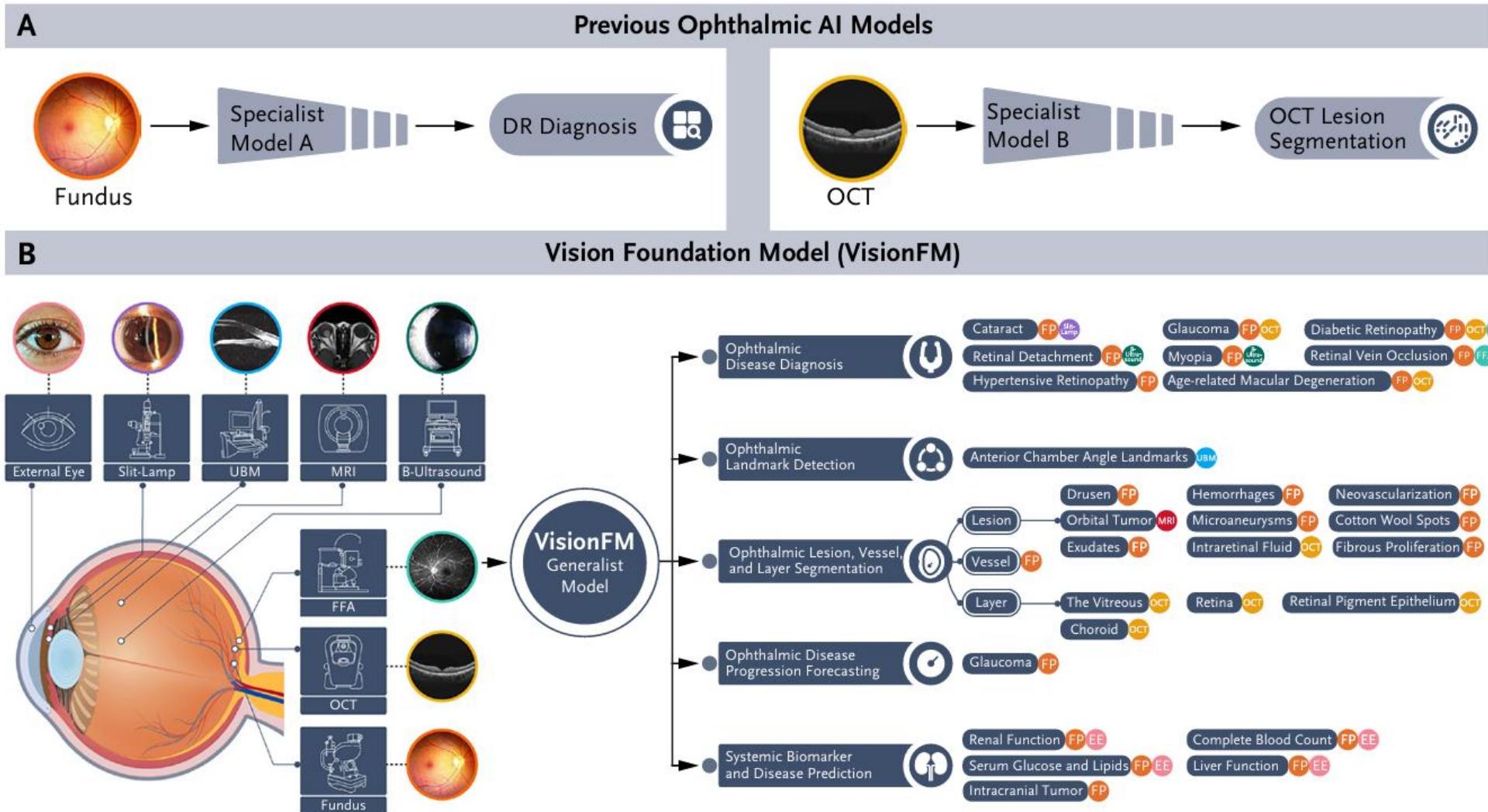


Authors: Jianing Qiu, Ph.D.  , Jian Wu, M.D., Ph.D.  , Hao Wei, M.Sc.  , Peilun Shi, M.Res.  , Minqing Zhang, M.Sc.  , Yunyun Sun, M.D., Ph.D.  , Lin Li, Ph.D.  , Hanruo Liu, Ph.D.  , Hongyi Liu, M.B.B.S. , Simeng Hou, M.B.B.S. , Yuyang Zhao, M.B.B.S. , Xuehui Shi, M.D. , Junfang Xian, M.D., Ph.D. , Xiaoxia Qu, Ph.D. , Sirui Zhu, M.B.B.S. , Lijie Pan, M.D. , Xiaoniao Chen, M.D., Ph.D. , Xiaoja Zhang, M.B.B.S. , Shuai Jiang, Ph.D. , Kebing Wang, M.D. , Chenlong Yang, M.D., Ph.D. , Mingqiang Chen, M.B.B.S. , Sujie Fan, M.D. , Jianhua Hu, M.B.B.S. , Aiguo Lv, M.B.B.S. , Hui Miao, M.D. , Li Guo, M.B.B.S. , Shujun Zhang, M.B.B.S. , Cheng Pei, M.D. , Xiaojuan Fan, M.D. , Jianqin Lei, M.D., Ph.D. , Ting Wei, M.D., Ph.D. , Junguo Duan, M.D. , Chun Liu, M.B.B.S. , Xiaobo Xia, M.D., Ph.D. , Siqi Xiong, M.D., Ph.D. , Junhong Li, M.D. , Kyle Lam, M.R.C.S., Ph.D. , Benny Lo, Ph.D. , Yih Chung Tham, Ph.D. , Tien Yin Wong, M.D., Ph.D. , Ningli Wang, M.D., Ph.D. , and Wu Yuan, Ph.D.

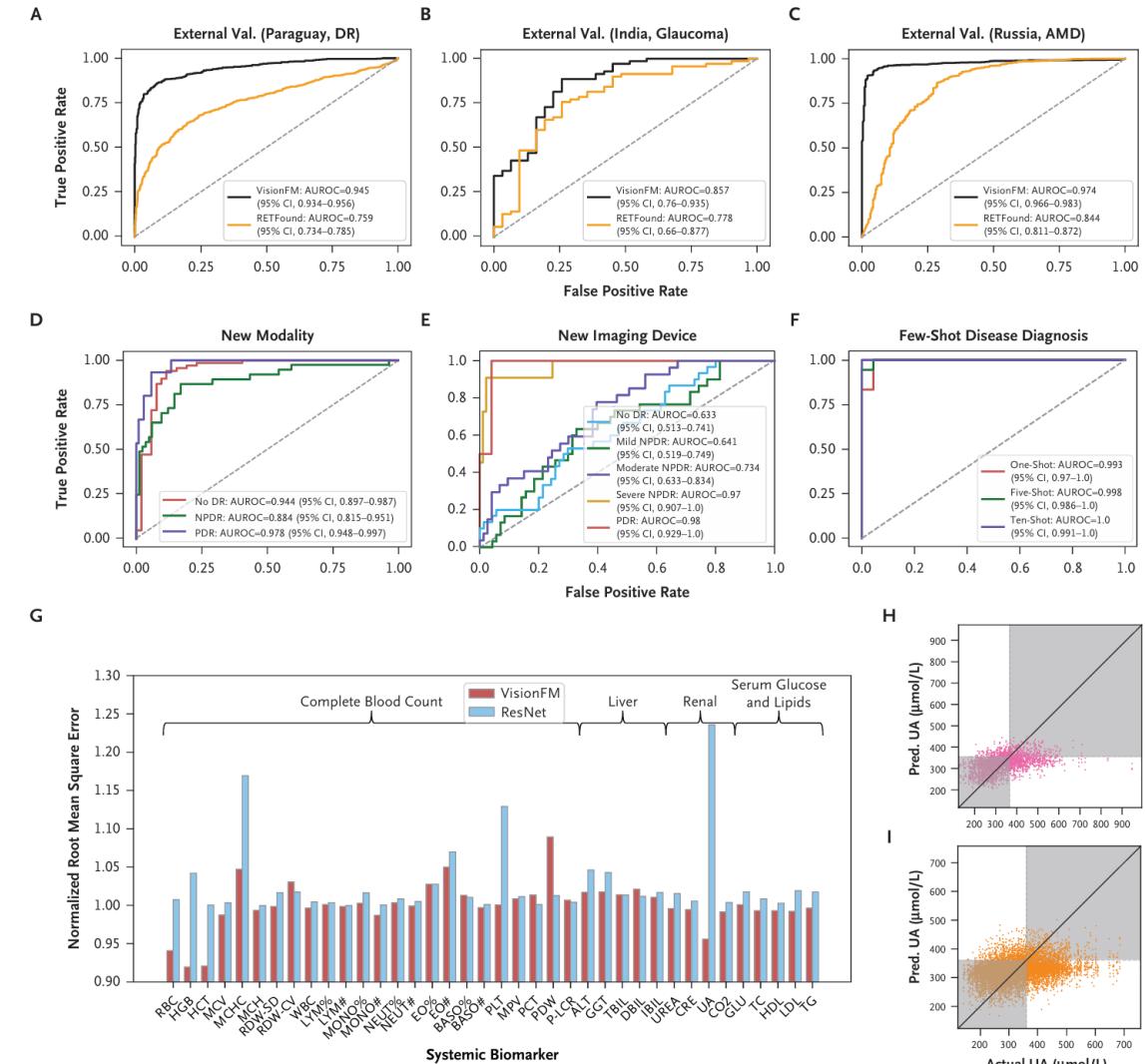
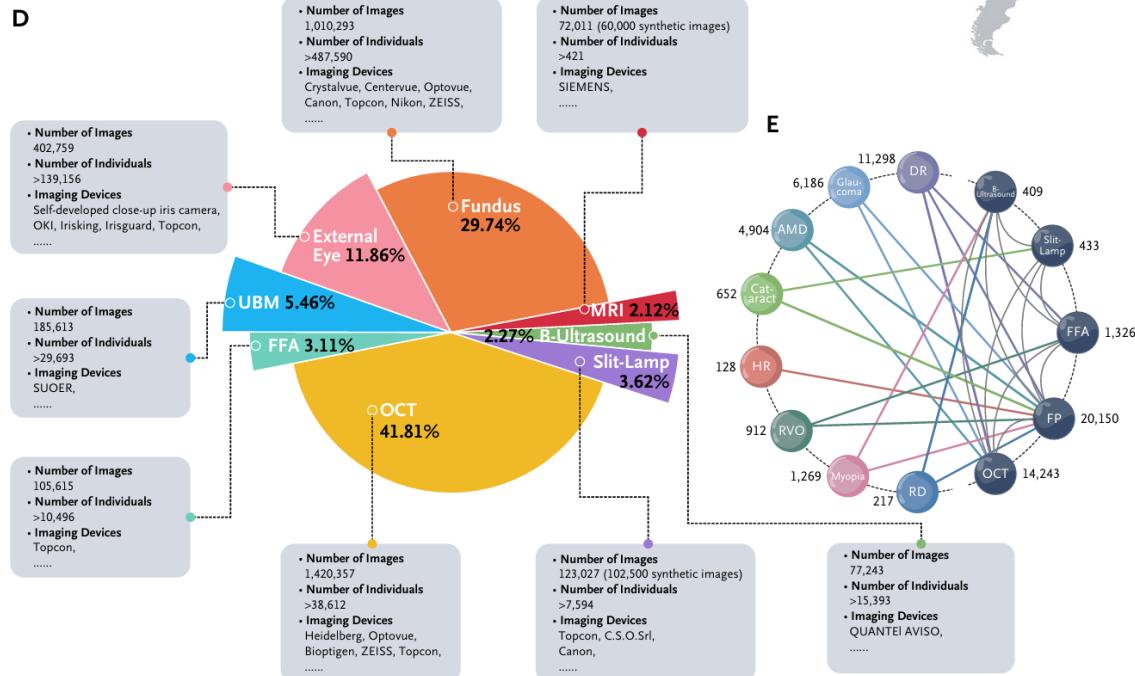
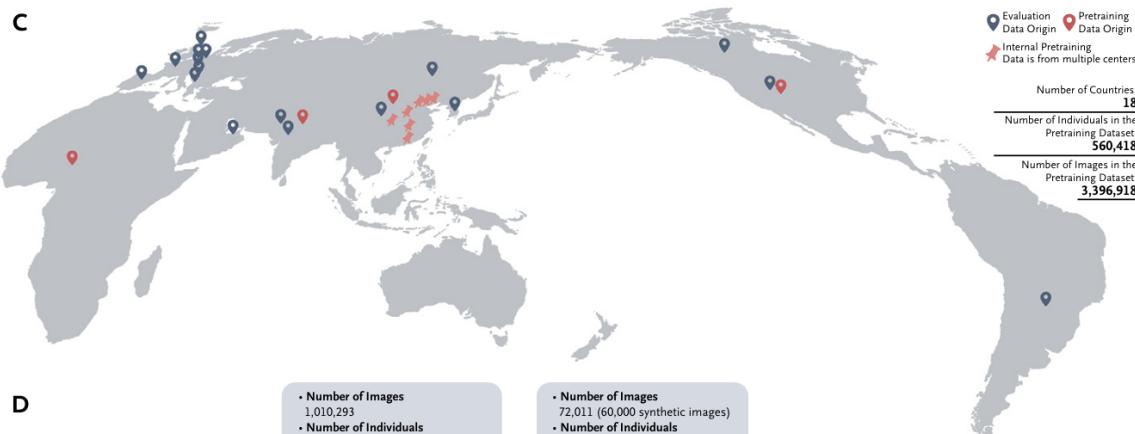
-35 Author Info & Affiliations

Published November 27, 2024 | NEJM AI 2024;1(12) | DOI: 10.1056/Aloa2300221 | VOL. 1 NO. 12

Overview of VisionFM



Generalizability and systemic biomarker prediction of VisionFM



Article | [Open access](#) | Published: 04 July 2024

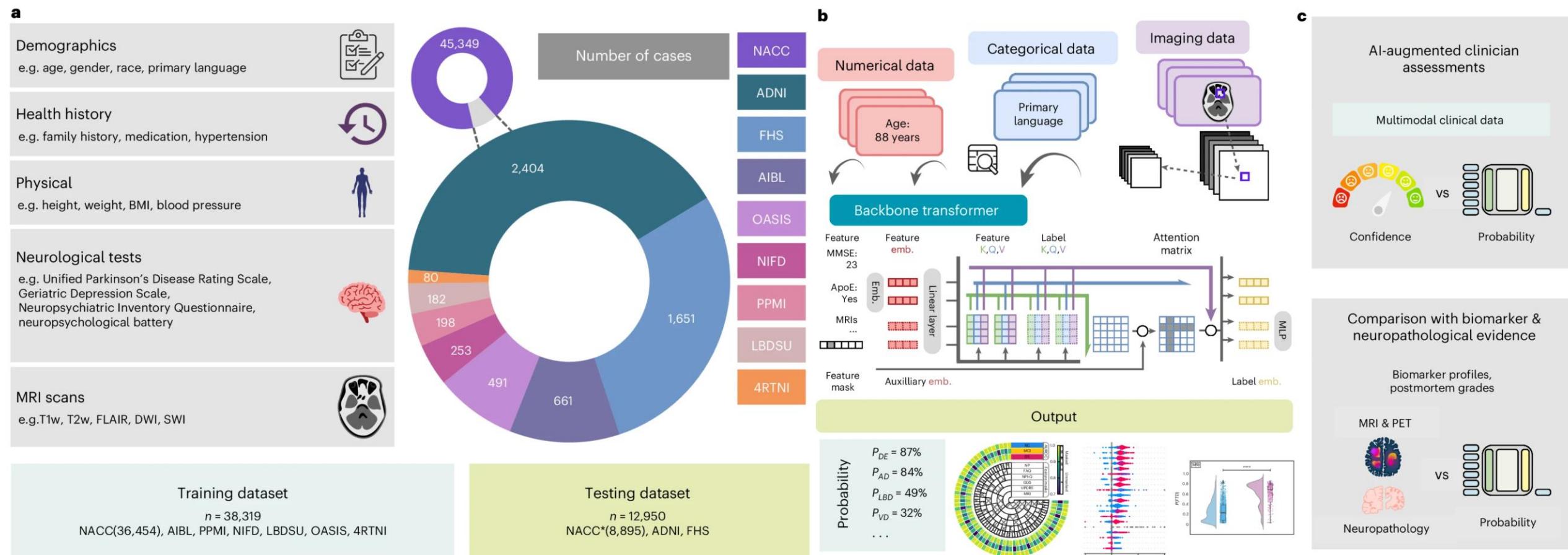
AI-based differential diagnosis of dementia etiologies on multimodal data



[Chonghua Xue](#), [Sahana S. Kowshik](#), [Diala Lteif](#), [Shreyas Puducher](#), [Varuna H. Jasodanand](#), [Olivia T. Zhou](#),
[Anika S. Walia](#), [Osman B. Guney](#), [J. Diana Zhang](#), [Serena Poésy](#), [Artem Kaliaev](#), [V. Carlota Andreu-Arasa](#),
[Brigid C. Dwyer](#), [Chad W. Farris](#), [Honglin Hao](#), [Sachin Kedar](#), [Asim Z. Mian](#), [Daniel L. Murman](#), [Sarah A. O'Shea](#), [Aaron B. Paul](#), [Saurabh Rohatgi](#), [Marie-Helene Saint-Hilaire](#), [Emmett A. Sartor](#), [Bindu N. Setty](#), [Juan E. Small](#), [Arun Swaminathan](#), [Olga Taraschenko](#), [Jing Yuan](#), [Yan Zhou](#), [Shuhan Zhu](#), [Cody Karjadi](#), [Ting Fang Alvin Ang](#), [Sarah A. Bargal](#), [Bryan A. Plummer](#), [Kathleen L. Poston](#), [Meysam Ahangaran](#), [Rhoda Au](#) & [Vijaya B. Kolachalamma](#)  — Show fewer authors

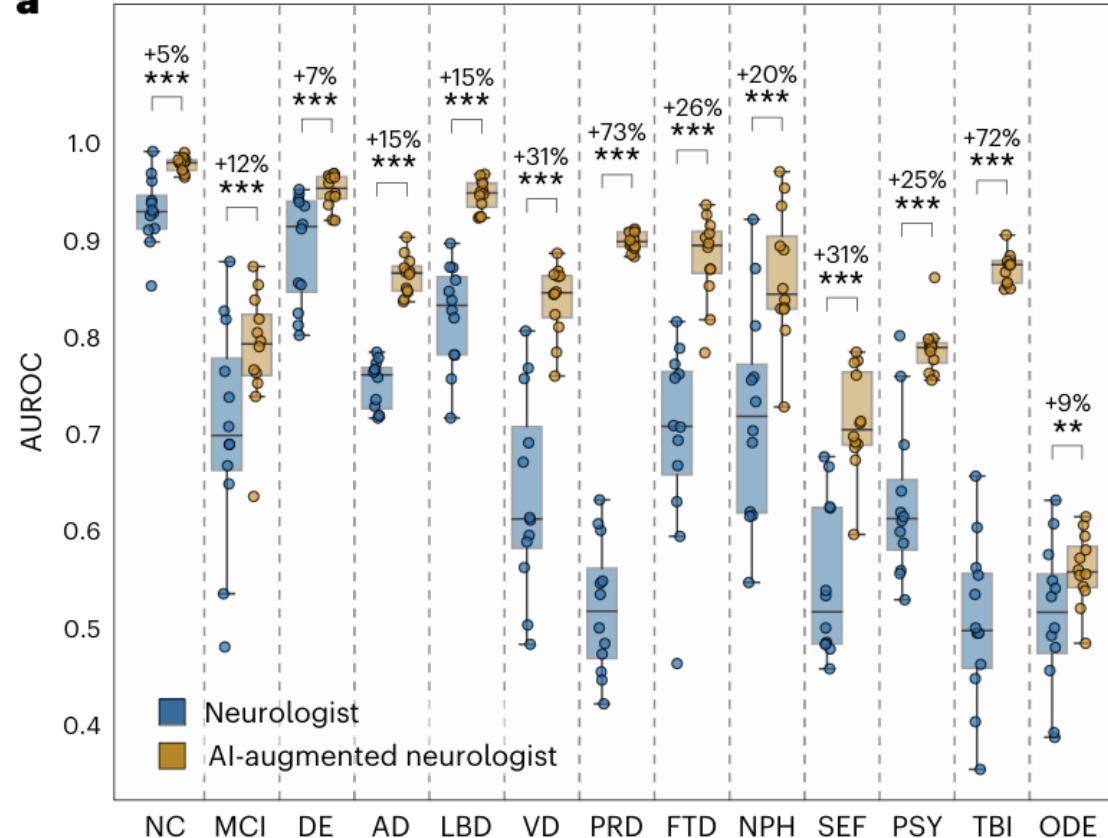
[Nature Medicine](#) **30**, 2977–2989 (2024) | [Cite this article](#)

Data and model architecture

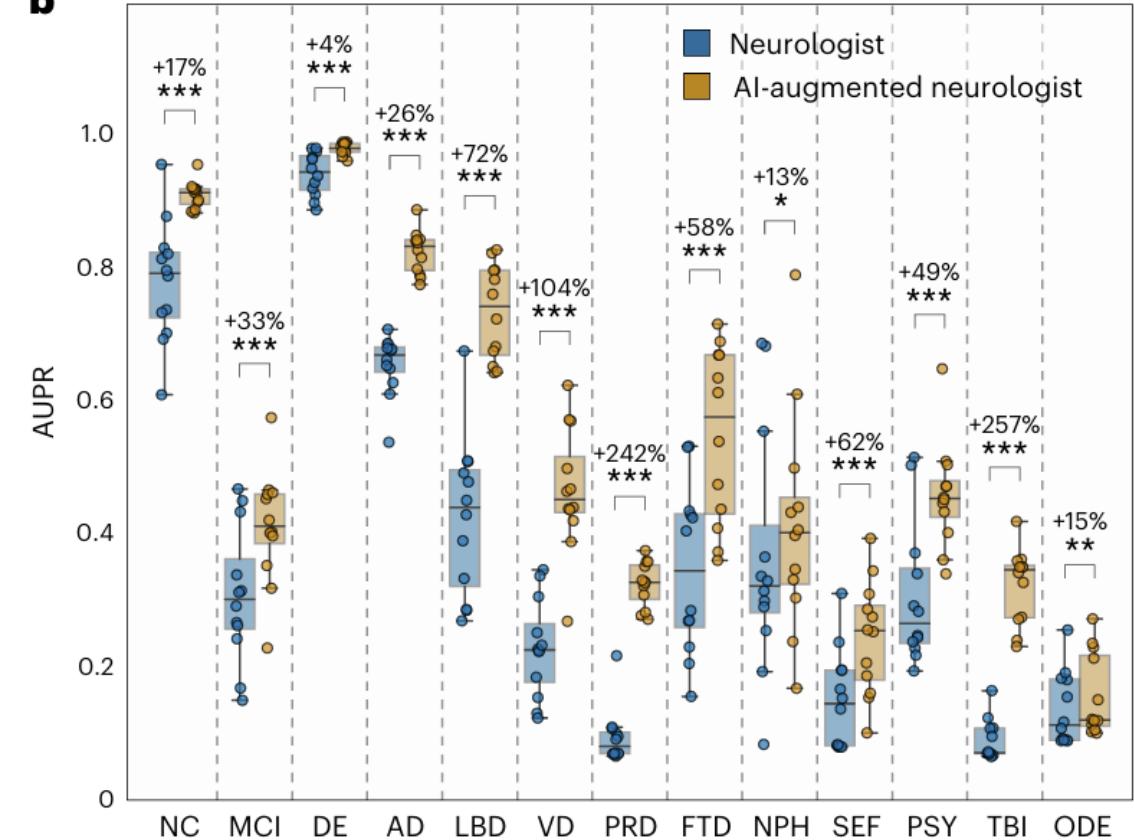


AI-augmented clinician assessments

a



b



Article | Published: 07 August 2024



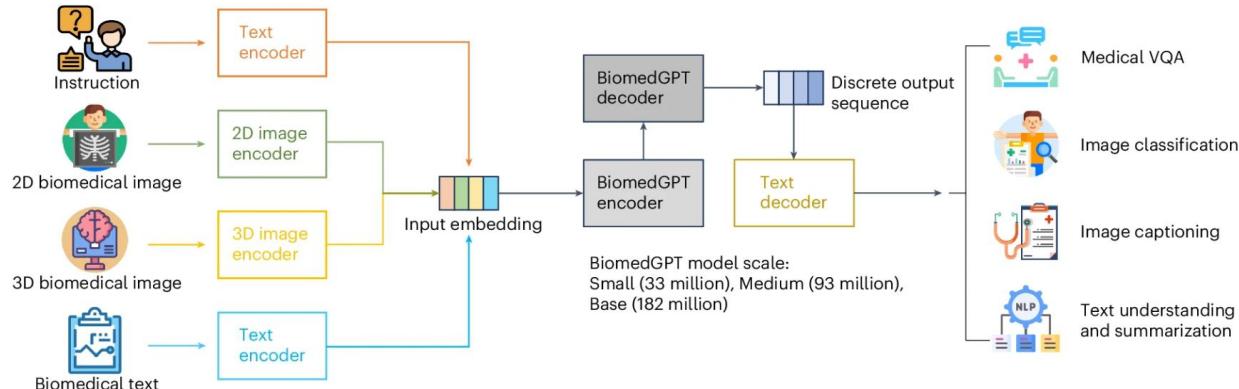
A generalist vision–language foundation model for diverse biomedical tasks

[Kai Zhang](#), [Rong Zhou](#), [Eashan Adhikarla](#), [Zhiling Yan](#), [Yixin Liu](#), [Jun Yu](#), [Zhengliang Liu](#), [Xun Chen](#), [Brian D. Davison](#), [Hui Ren](#), [Jing Huang](#), [Chen Chen](#), [Yuyin Zhou](#), [Sunyang Fu](#), [Wei Liu](#), [Tianming Liu](#), [Xiang Li](#)✉,
[Yong Chen](#), [Lifang He](#)✉, [James Zou](#), [Quanzheng Li](#), [Hongfang Liu](#) & [Lichao Sun](#)✉

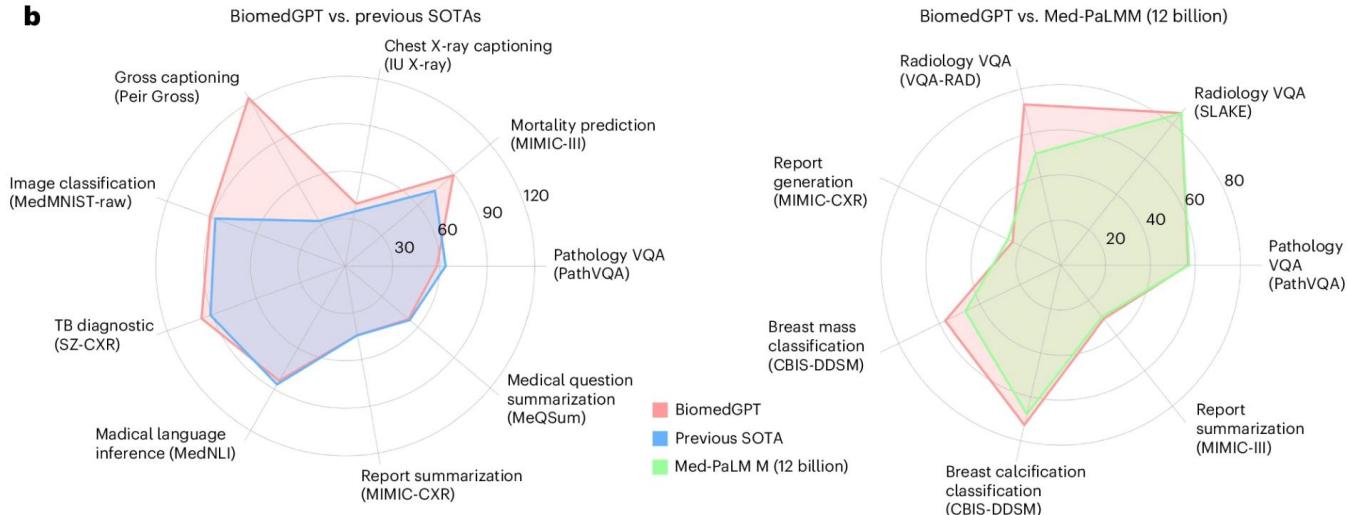
Nature Medicine **30**, 3129–3141 (2024) | [Cite this article](#)

BiomedGPT – multi-modalities and multi-tasks

a



b



Pathology and radiology VQA

Q: What is seen at this stage, associated with regeneration and repair?
A: Numerous reactive type II pneumocytes.
Q: Are bite cells like this one in the smear associated with regeneration and repair at this stage?
A: No.

Medical VQA

Q: What are the findings based on the image?
A: The nasogastric tube is in adequate position, and there is a resolution of the gastric distention. There is still mild bibasilar atelectasis. There are no pneumothorax or pleural effusion.

Image classification

Q: What disease does this image depict?
A: Breast cancer.

Image captioning

Q: What skin lesion does this image depict?
A: Melanoma.

Text understanding and summarization

Mortality prediction

Chief complaint: Dyspnea, abdominal distension
Present illness: 78-year-old female with multiple medical problems, pertinently including CAD status post CABG, hypertension and type 2 diabetes
Medical history: (1) CAD status post CABG [Reg#] (2) Hypertension (3) Type 2 diabetes (4) Pulmonary fibrosis
Allergies: Patient recorded as having no known allergies to drugs
Physical exam: BP 107/68, HR 70s, RR 28, 90% on NRB

Q: What is the predicted outcome for the patient before discharge?
A: Deceased.

Treatment suggestion

ID	Age	Sex	Race	Tumor size	Nodes examined	ER Status	PR Status
001	54	Female	White	25	14	Negative	Negative
002	44	Female	White	23	34	Positive	Positive
003	47	Male	White	22	3	—	Positive
004	89	Female	White	17	1	Negative	Positive
005	59	Female	White	36	19	Positive	Negative

The patient is a 44-year-old white female. She has one malignant tumor and five regional lymph nodes that tested positive. The tumor measures 23 mm. Estrogen and progesterone receptor tests are positive. A total of 34 regional nodes were removed.

Q: Please provide treatment suggestion given the patient's information.
A: Recommend using beam radiation, suggesting that the sequence for radiation should be post-surgery. Furthermore, chemotherapy should indeed be considered.

Report summarization

There is no evidence of hemorrhage, masses, mass effect or shift of normally midline structures. The ventricles and sulci are mildly prominent, compatible with age-appropriate involutional changes. There is hypopattenuation along the right caudate head and periventricular frontal white matter, unchanged, compatible with small vessel ischemic disease.

Q: What is summary based on the given report?
A: 1. No acute intracranial process.
2. Small vessel ischemic disease.

Clinical trial matching

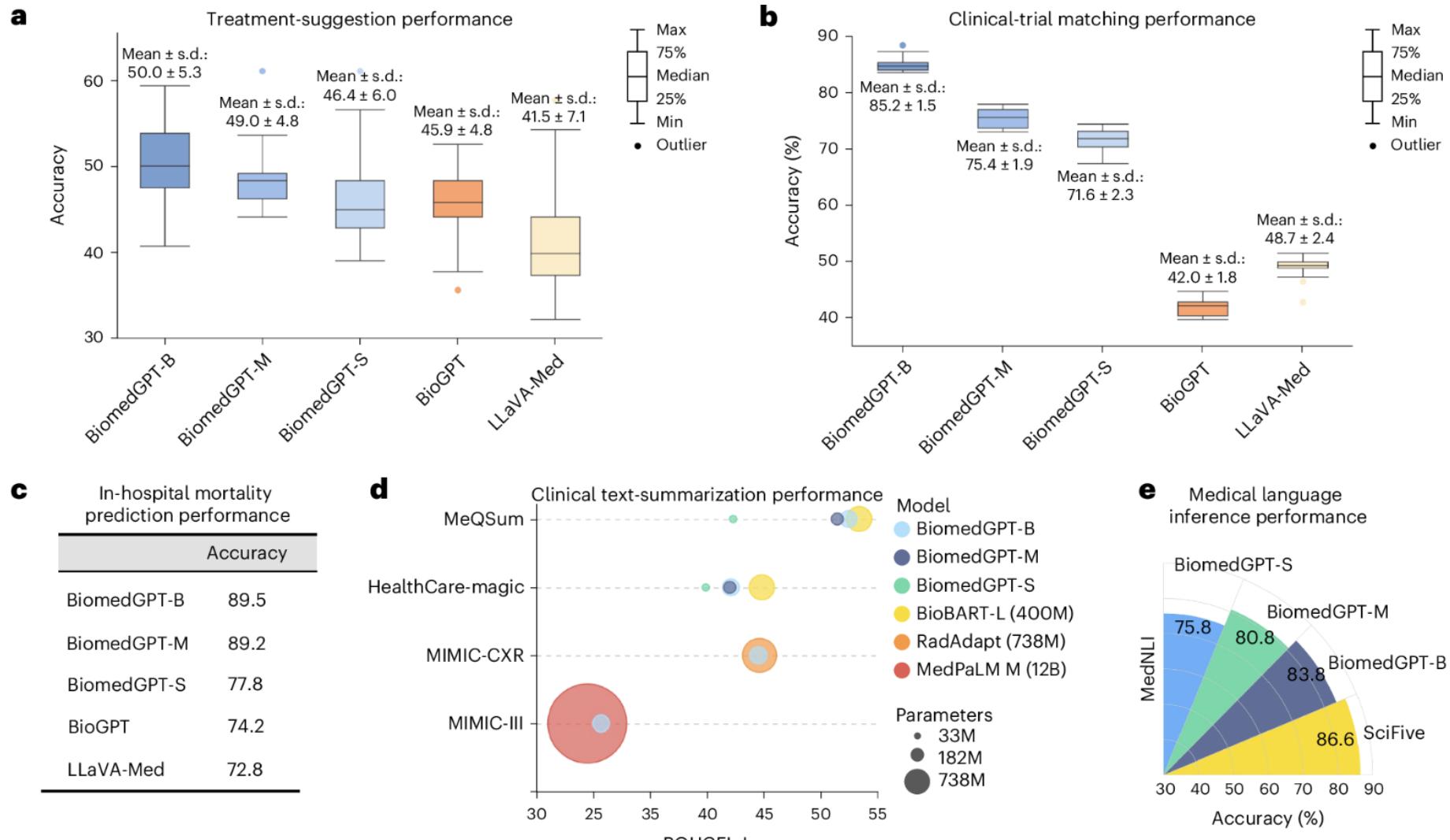
Patient information
A 19-year-old male came to the clinic with some sexual concerns. On physical examination, there are some poorly developed secondary sexual characteristics. Ultrasound reveals a testes volume of 1-2 ml. The hormonal evaluation showed a serum testosterone level of 65 ng dL⁻¹ with low levels of GnRH.

Inclusion criteria: Total serum testosterone concentrations < 300 ng dL⁻¹. Male patients over the age of 18.

Exclusion criteria: Elevated PSA > 3.5 ng mL⁻¹.

Q: Please determine the patient's eligibility by comparing the given patient note and trial details.
A: The patient is eligible for the clinical trial.

BiomedGPT performance



Article | Published: 12 June 2024

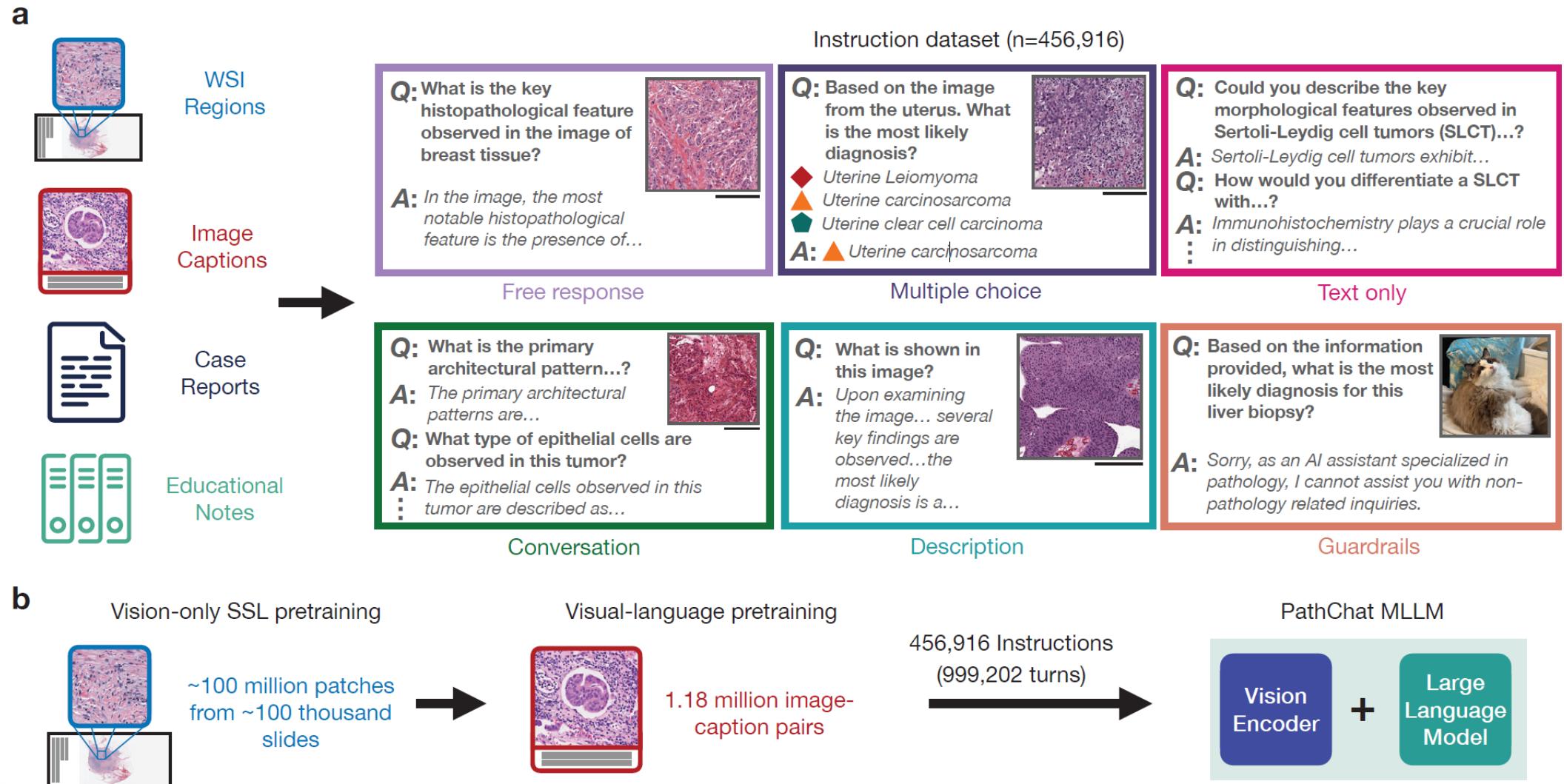


A Multimodal Generative AI Copilot for Human Pathology

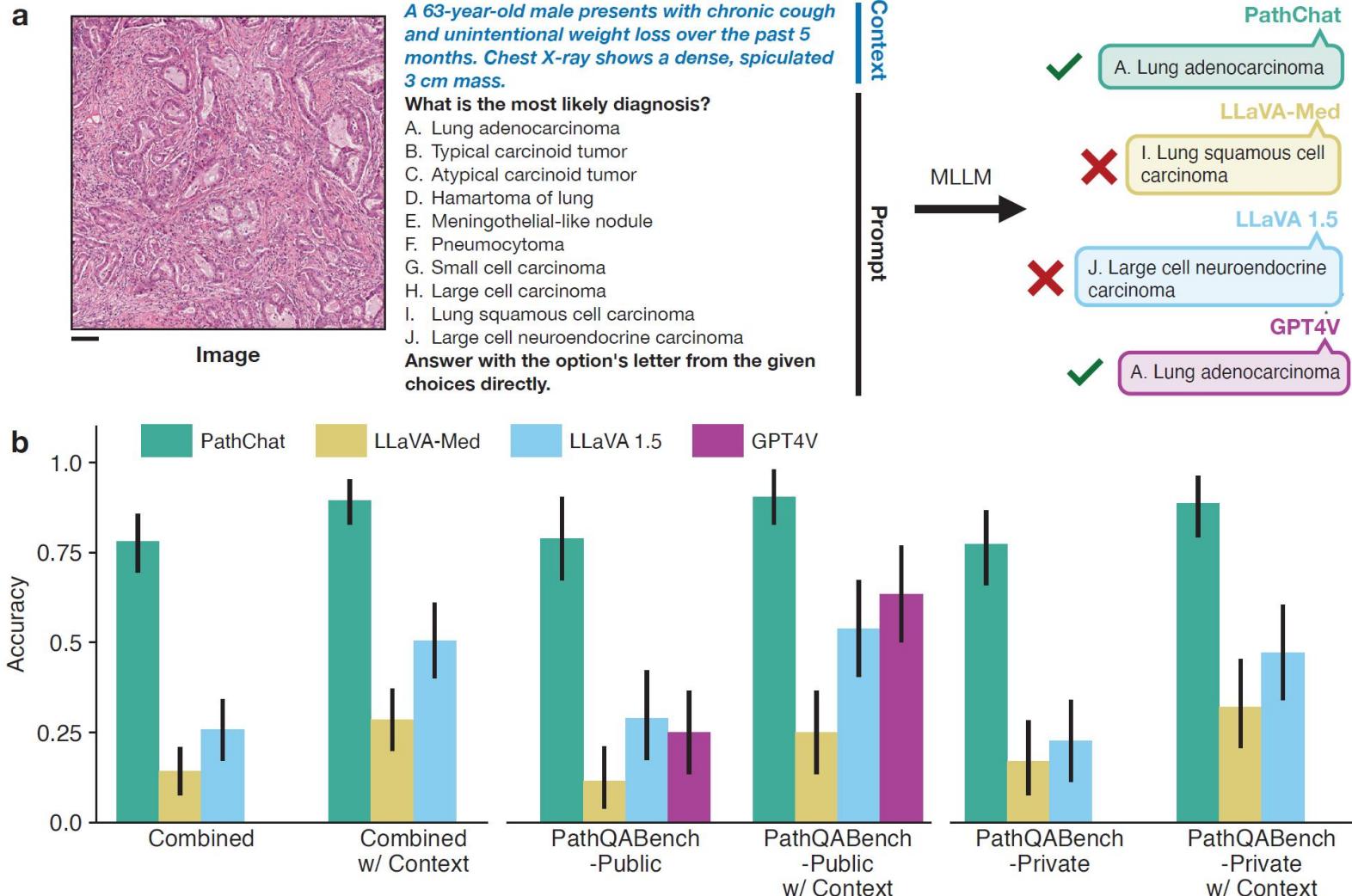
[Ming Y. Lu](#), [Bowen Chen](#), [Drew F. K. Williamson](#), [Richard J. Chen](#), [Melissa Zhao](#), [Aaron K. Chow](#), [Kenji Ikemura](#), [Ahrong Kim](#), [Dimitra Pouli](#), [Ankush Patel](#), [Amr Soliman](#), [Chengkuan Chen](#), [Tong Ding](#), [Judy J. Wang](#), [Georg Gerber](#), [Ivy Liang](#), [Long Phi Le](#), [Anil V. Parwani](#), [Luca L. Weishaupt](#) & [Faisal Mahmood](#)✉

[Nature](#) (2024) | [Cite this article](#)

PathChat overview



PathChat Evaluation



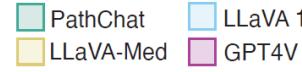
PathChat Evaluation

a Panel of 7 pathologists



+

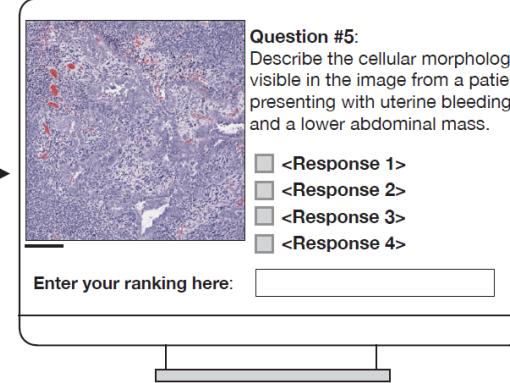
4 AI assistant models



+

260 open-ended questions

Shuffled and de-identified responses
ranked by each expert



Expert rankings

P1 Q#5: 1 > 3 = 4 > 2

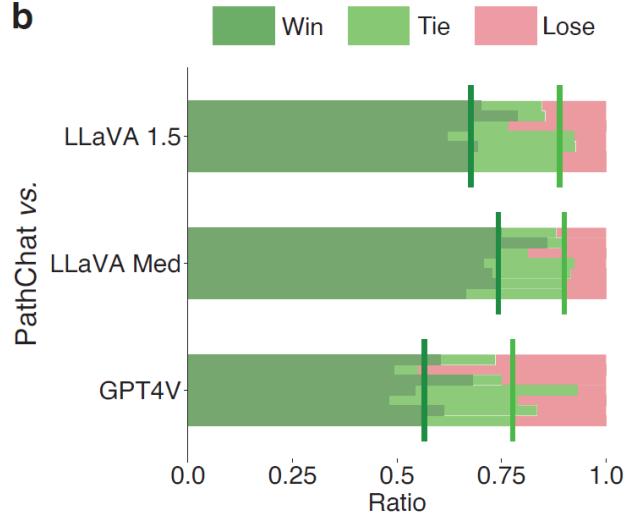
P2 Q#5: 1 > 3 = 4 = 2

P7 Q#5: 3 > 1 > 4 > 2

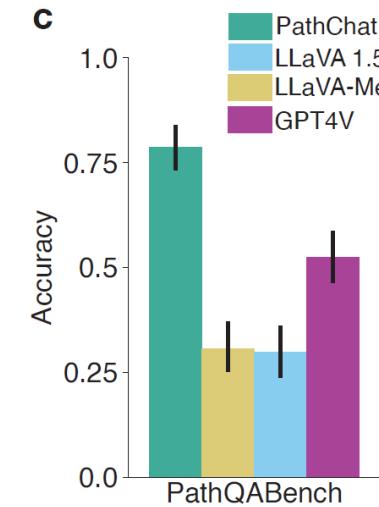
Win / tie / lose
record of PathChat



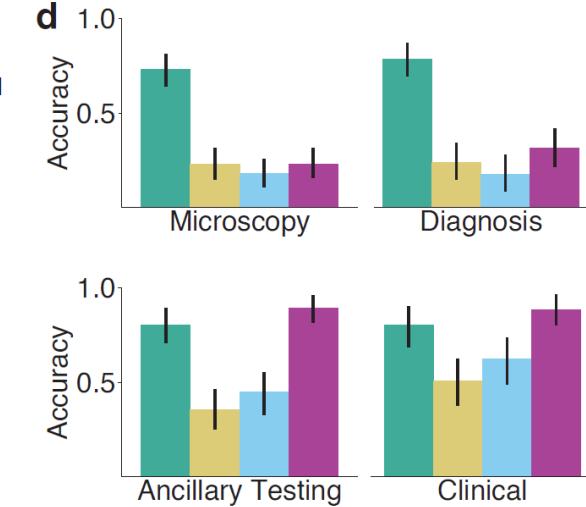
b



c



d





Addressing Asynchronicity in Clinical Multimodal Fusion via Individualized Chest X-ray Generation

Diffusion-based Dynamic Latent Chest X-ray Image Generation (**DDL-CXR**)

Wenfang Yao^{*1}, Chen Liu ^{*1,3}, Kejing Yin^{2✉}, William K. Cheung², Jing Qin¹

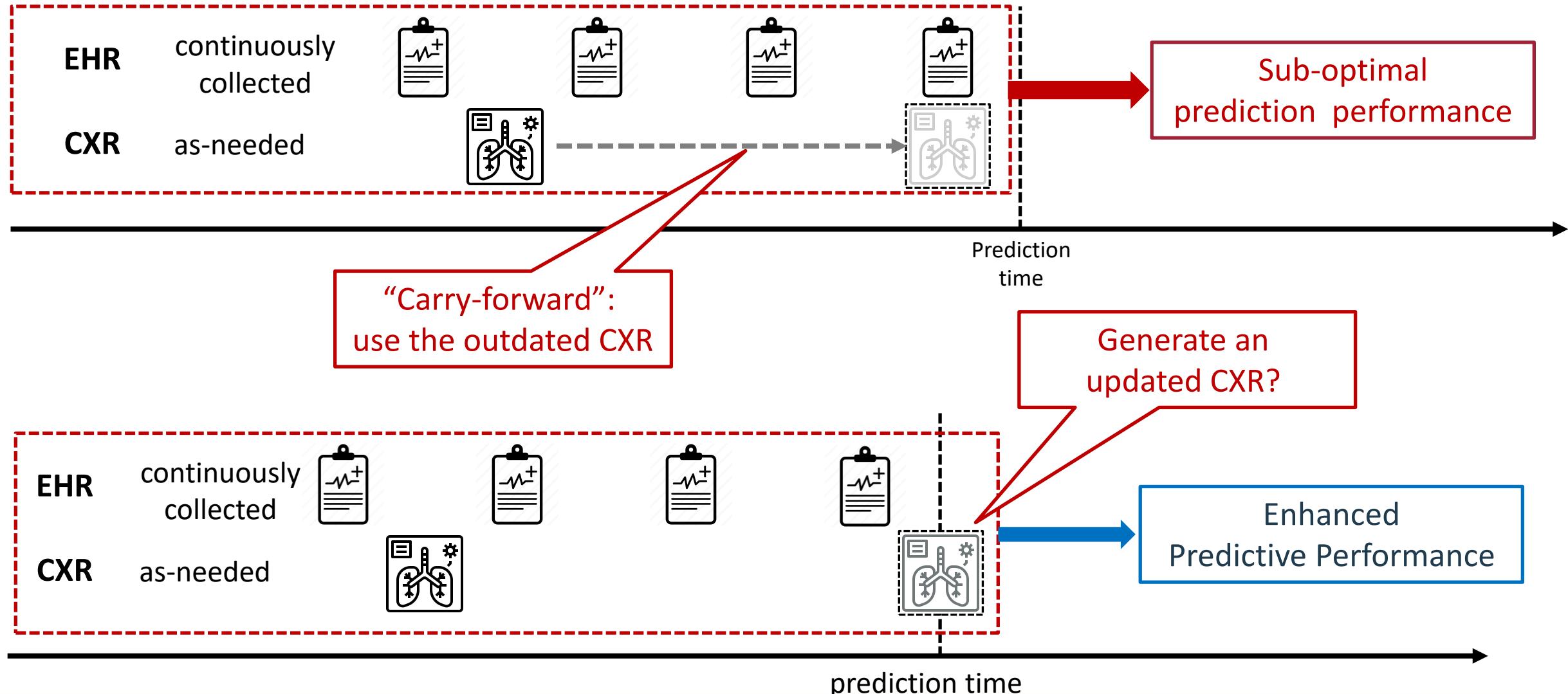
¹School of Nursing, The Hong Kong Polytechnic University

²Department of Computer Science, Hong Kong Baptist University

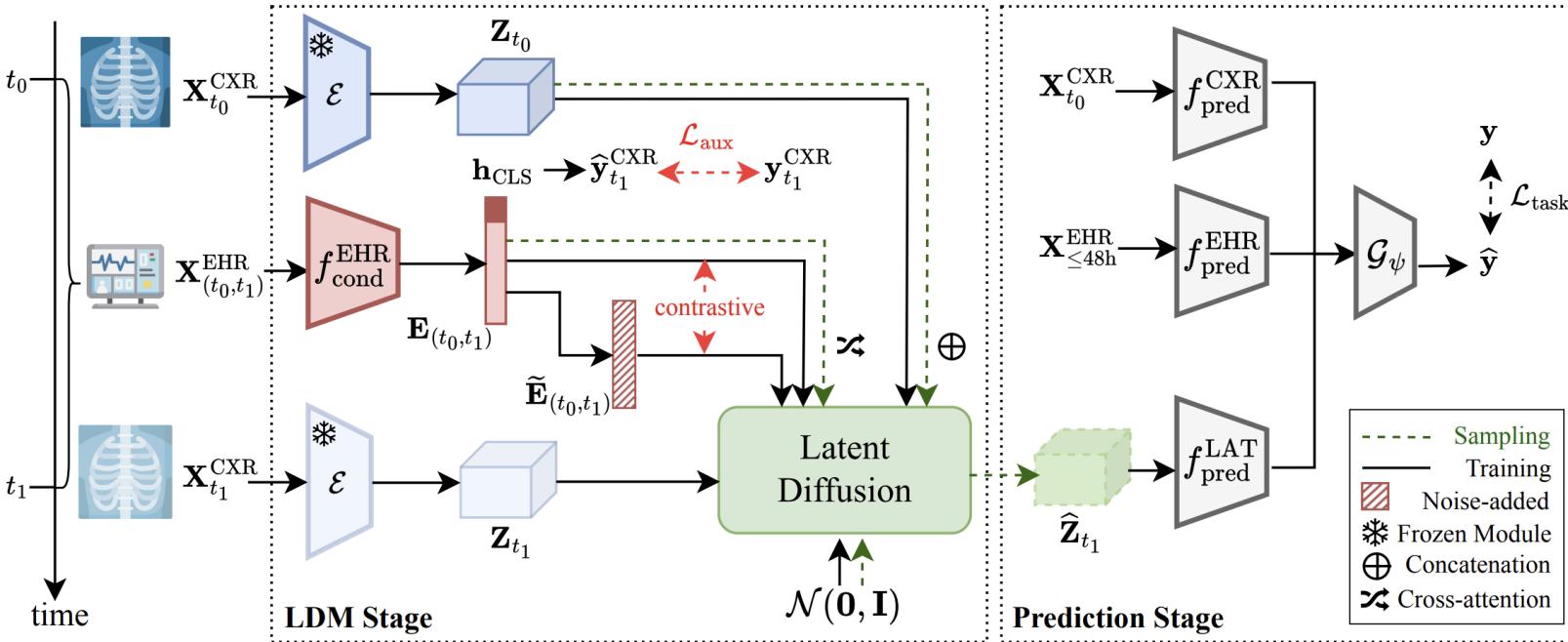
³School of Software Engineering, South China University of Technology

<https://github.com/Chenliu-svg/DDL-CXR>

Challenge: Clinical Data Are Highly Asynchronous



Diffusion-based Dynamic Latent Chest X-ray Image Generation (DDL-CXR)



DDL-CXR obtains the best overall performance

- Generating an updated CXR is beneficial for prediction
- Relative improvements: 2.4% (phenotype classification); 3.56% (mortality prediction)

LDM stage: dynamic latent CXR generation

Enhancing semantic multimodal fusion via contrastive LDM learning:

$$\text{Conditioning mechanisms } \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right) \cdot \mathbf{V},$$

$$\tilde{\mathbf{E}}_{(t_0,t_1)} = (1 - \beta)\mathbf{E}_{(t_0,t_1)} + \beta\delta, \text{ where } \delta \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\text{with } \mathbf{Q} = \mathbf{W}_Q \cdot \varphi\left(\mathbf{z}_{t_1}^{(n)} || \mathbf{z}_{t_0}\right), \mathbf{K} = \mathbf{W}_K \cdot f_{\text{cond}}^{\text{EHR}}(\mathbf{X}_{(t_0,t_1)}^{\text{EHR}}), \mathbf{V} = \mathbf{W}_V \cdot f_{\text{cond}}^{\text{EHR}}(\mathbf{X}_{(t_0,t_1)}^{\text{EHR}})$$

Capturing disease course via EHR time series: $\mathcal{L}_{\text{aux}} := \frac{1}{M} \frac{1}{L} \sum_{m=1}^M \sum_{l=1}^L y_{ml}^{\text{CXR}} \log(\hat{y}_{ml}^{\text{CXR}}) + (1 - y_{ml}^{\text{CXR}}) \log(1 - \hat{y}_{ml}^{\text{CXR}})$

LDM training loss:

$$\mathcal{L}_{\text{LDM}} := \mathbb{E}_{\mathbf{z}_{t_1}, \mathbf{z}_{t_0}, \mathbf{x}_{(t_0,t_1)}^{\text{EHR}}, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), n} \left[\left\| \epsilon - \epsilon_\theta \left(\mathbf{z}_{t_1}^{(n)}, \mathbf{z}_{t_0}, f_{\text{cond}}^{\text{EHR}}(\mathbf{X}_{(t_0,t_1)}^{\text{EHR}}), n \right) \right\|_2^2 + \lambda_1 \max \left(\left\| \epsilon - \epsilon_\theta \left(\mathbf{z}_{t_1}^{(n)}, \mathbf{z}_{t_0}, \mathbf{E}_{(t_0,t_1)}, n \right) \right\|_2^2 - \left\| \epsilon - \epsilon_\theta \left(\mathbf{z}_{t_1}^{(n)}, \mathbf{z}_{t_0}, \tilde{\mathbf{E}}_{(t_0,t_1)}, n \right) \right\|_2^2 + \alpha, 0 \right) \right]$$

Published as a conference paper at ICLR 2024



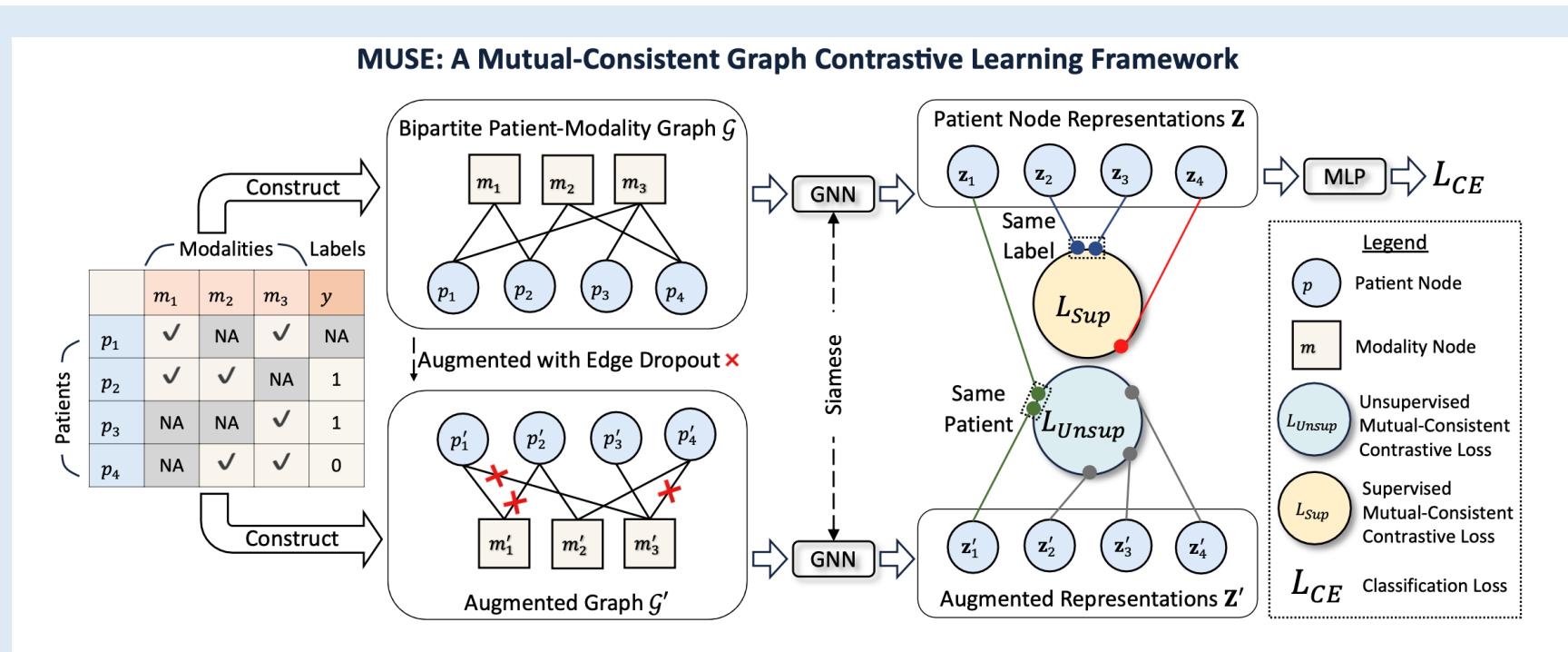
MULTIMODAL PATIENT REPRESENTATION LEARNING WITH MISSING MODALITIES AND LABELS

**Zhenbang Wu^{1,2,3*}, Anant Dadu^{2,3}, Nicholas Tustison⁴, Brian Avants⁵,
Mike Nalls^{2,3}, Jimeng Sun¹, Faraz Faghri^{2,3}**

¹University of Illinois Urbana-Champaign, ²National Institutes of Health

³Data Tecnica International, ⁴University of Virginia, ⁵University of Pennsylvania
zw12@illinois.edu

How to learn multimodal patient representations with missing modalities and labels?



Idea 1: Bipartite Patient-Modality Graph

Flexible representation for dataset with missingness

Idea 2: Mutual-Consistent Contrastive Loss

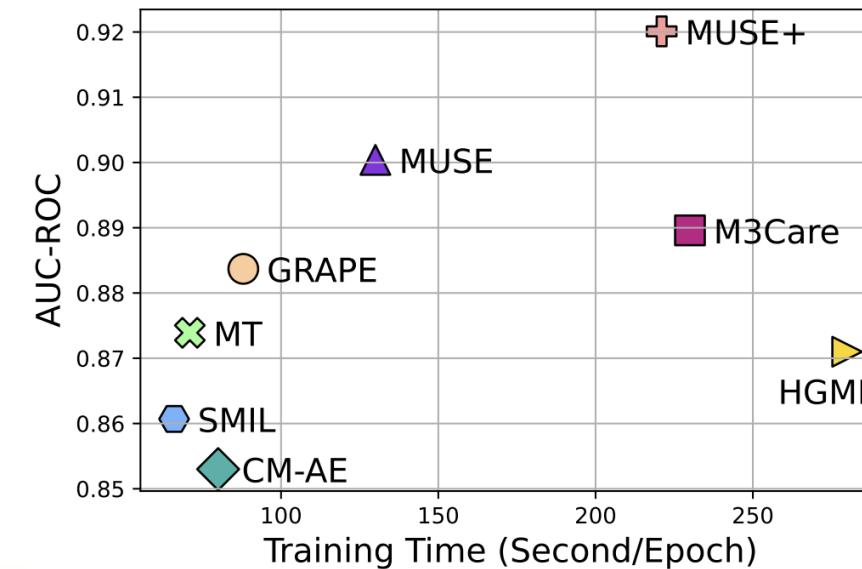
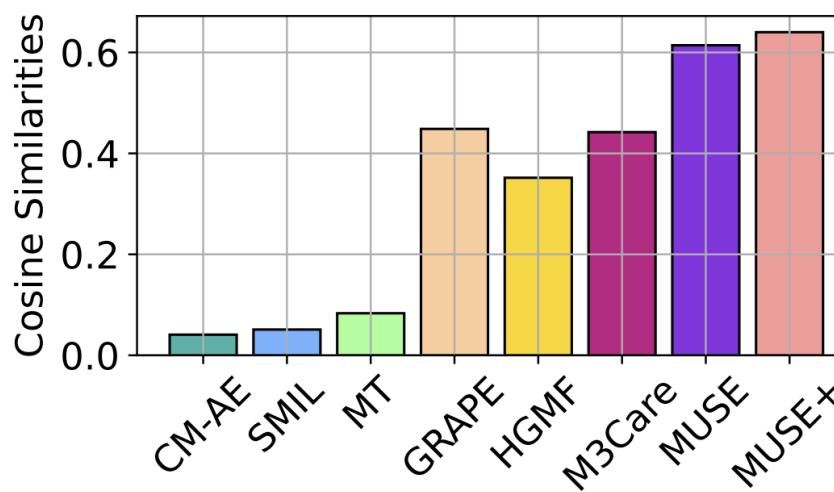
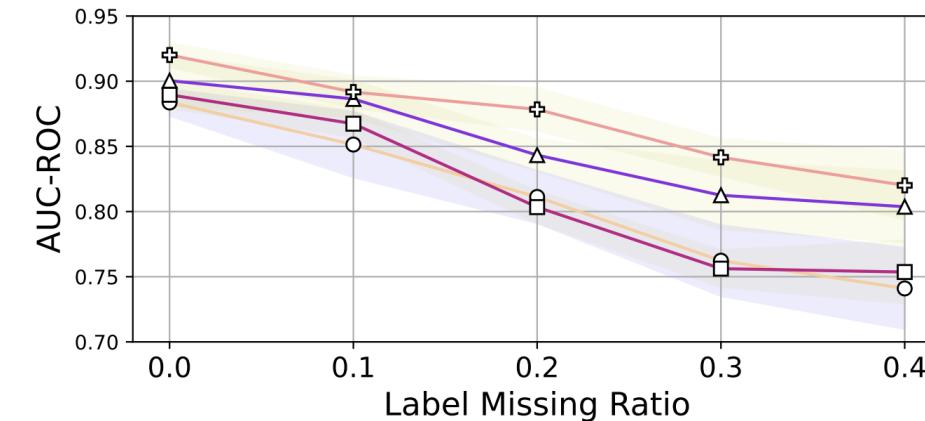
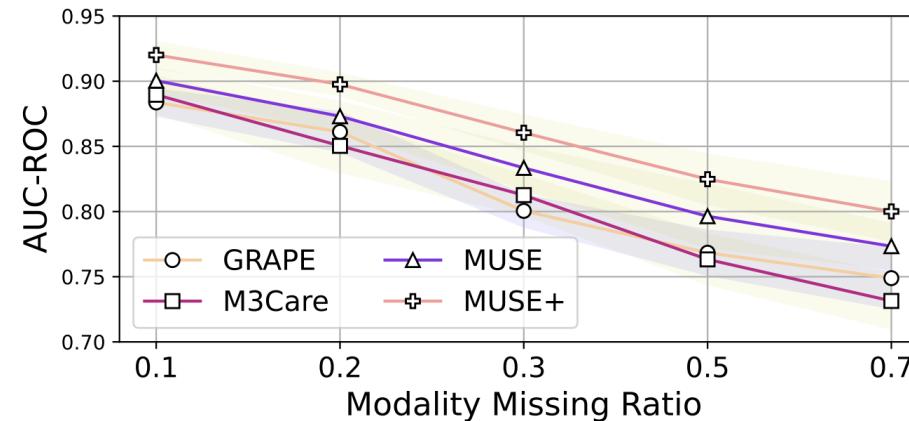
Focus on modality-agnostic and label-decisive features

Results on AD progression

Method	AUC-ROC	Accuracy
CM-AE	0.8722†	0.7305†
SMIL	0.8761†	0.7338†
MT	0.8935	0.7604
GRAPE	0.9031†	0.7820†
HGMF	0.8845†	0.7463†
M3Care	0.9101	0.7822†
MUSE	0.9158†*	0.7973†*
MUSE+	0.9309*	0.8291*

- MUSE outperforms all baselines
- MUSE+ further enhances the performance with unlabeled patients

Looking into the modalities and scalability



Bias, Trust, and ELSI in AI



[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 04 July 2024



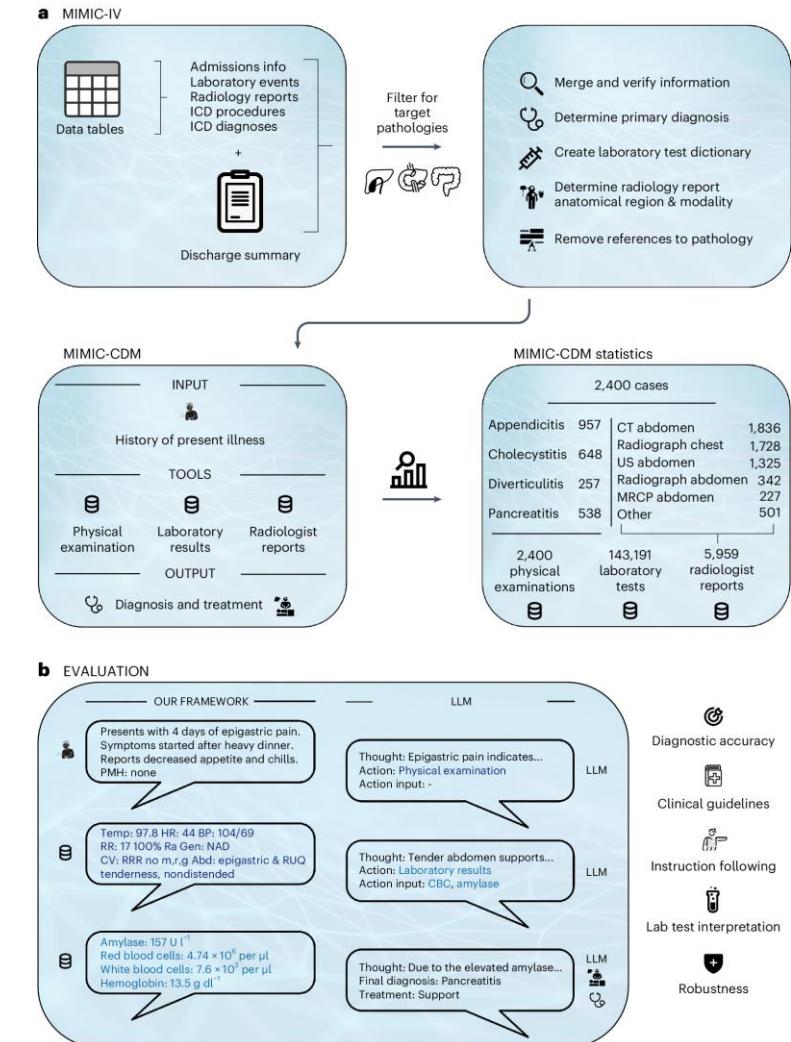
Evaluation and mitigation of the limitations of large language models in clinical decision-making

[Paul Hager](#)✉, [Friederike Jungmann](#), [Robbie Holland](#), [Kunal Bhagat](#), [Inga Hubrecht](#), [Manuel Knauer](#),
[Jakob Vielhauer](#), [Marcus Makowski](#), [Rickmer Braren](#), [Georgios Kaassis](#) & [Daniel Rueckert](#)

[Nature Medicine](#) **30**, 2613–2622 (2024) | [Cite this article](#)

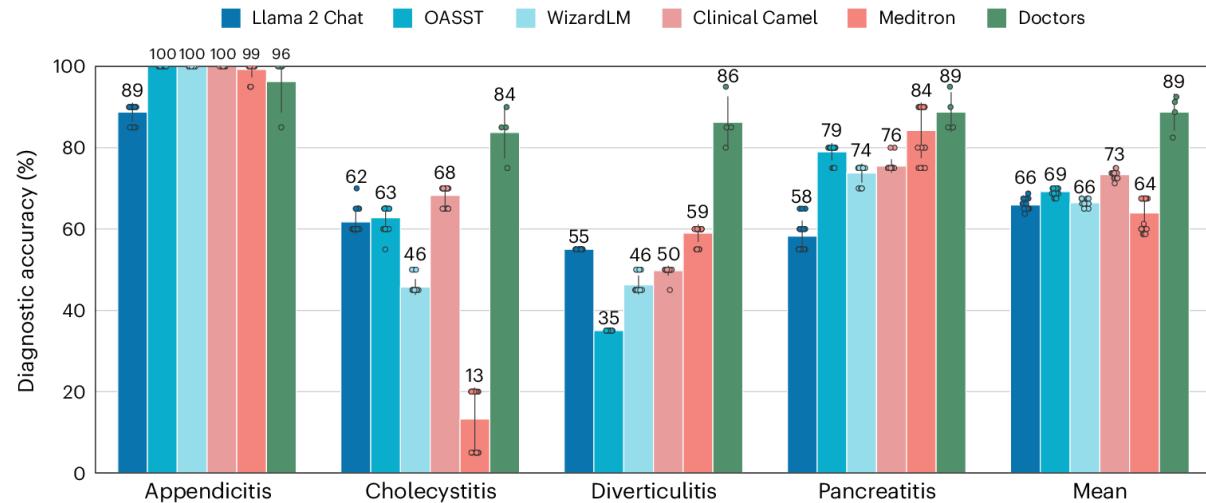
Evaluation and mitigation of the limitations of large language models in clinical decision-making

- Critically evaluates the limitations of large language models (LLMs) in clinical decision-making
- highlighting issues such as low diagnostic accuracy, guideline non-compliance, and susceptibility to instruction phrasing
- underscores the need for robust oversight and mitigation strategies to ensure these models do not compromise patient safety or exacerbate biases in healthcare
- Create a curated dataset based on 2400 patient cases from MIMIC.
- Evaluate a variety of existing LLMs on these cases in accuracy and reasoning.



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

- Found that LLMs currently perform poorly across a wide variety of evaluation categories, indicating that they are not ready in their current state for use in clinical decision support paradigms.
- Established a standard baseline of performance for future comparisons
- Found that LLMs tend to break their prompts/instructions and are too sensitive to the quantity or quality of information, which can be variable in clinical settings



[nature](#) > [npj digital medicine](#) > [brief communications](#) > article

Brief Communication | [Open access](#) | Published: 23 July 2024



Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine

[Qiao Jin](#), [Fangyuan Chen](#), [Yiliang Zhou](#), [Ziyang Xu](#), [Justin M. Cheung](#), [Robert Chen](#), [Ronald M. Summers](#), [Justin F. Rousseau](#), [Peiyun Ni](#), [Marc J. Landsman](#), [Sally L. Baxter](#), [Subhi J. Al'Aref](#), [Yijia Li](#), [Alexander Chen](#), [Josef A. Brejt](#), [Michael F. Chiang](#), [Yifan Peng](#)✉ & [Zhiyong Lu](#)✉

[npj Digital Medicine](#) 7, Article number: 190 (2024) | [Cite this article](#)

Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine



- Authors evaluated a model that has received recent popularity in medical challenge tasks, GPT-4 Vision
- How it performs with regards to metrics beyond accuracy alone
- They found that while GPT-4V performs well in accuracy, it performs poorly in presenting valid rationales, especially during image comprehension, limiting its utility in clinical decision-making workflows
- Evaluated GPT-4V on a standard set of imaging challenges and compared its performance to clinicians
- Assessed not just for accuracy but prompted the model to provide its reasoning for its answers
- However, found that it also tended to provide rationales or reasoning that were flawed, especially when considering image comprehension.

Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine

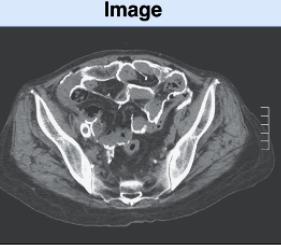


Evaluating GPT-4 Vision for Medical Question Answering



The NEW ENGLAND JOURNAL of MEDICINE (N = 207)

Question
A 52-year-old woman with end-stage kidney disease that was being managed with peritoneal dialysis presented with a 1-month history of bloody dialysate. She had had 3 episodes of bacterial peritonitis in the past 12 years. Physical examination and laboratory studies were unremarkable. Computed tomography of the abdomen was performed. What is the most likely diagnosis?

Image	Choices
	<ul style="list-style-type: none">A. CalciphylaxisB. Encapsulating peritoneal sclerosisC. Fungal peritonitisD. Peritoneal carcinomatosisE. Tuberculous peritonitis



GPT-4V

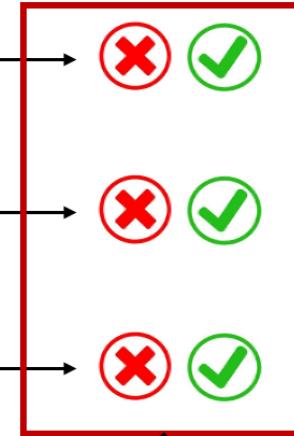
Image comprehension: The image appears to be an axial cross-sectional CT (computed tomography) scan of the lower abdomen ...

Recall of medical knowledge: Calciphylaxis is characterized by calcification of small and medium-sized blood vessels, leading to ...

Step-by-step reasoning: Given the significant and irregular thickening of the peritoneal lining encasing the bowel, the most consistent diagnosis with the provided ...

Final choice: B

Judged by Physicians with ground-truth



Research Question 1: How does GPT-4V compare to physicians in multi-choice accuracy?

Research Question 2: Is the GPT-4V decision-making based on solid rationales?

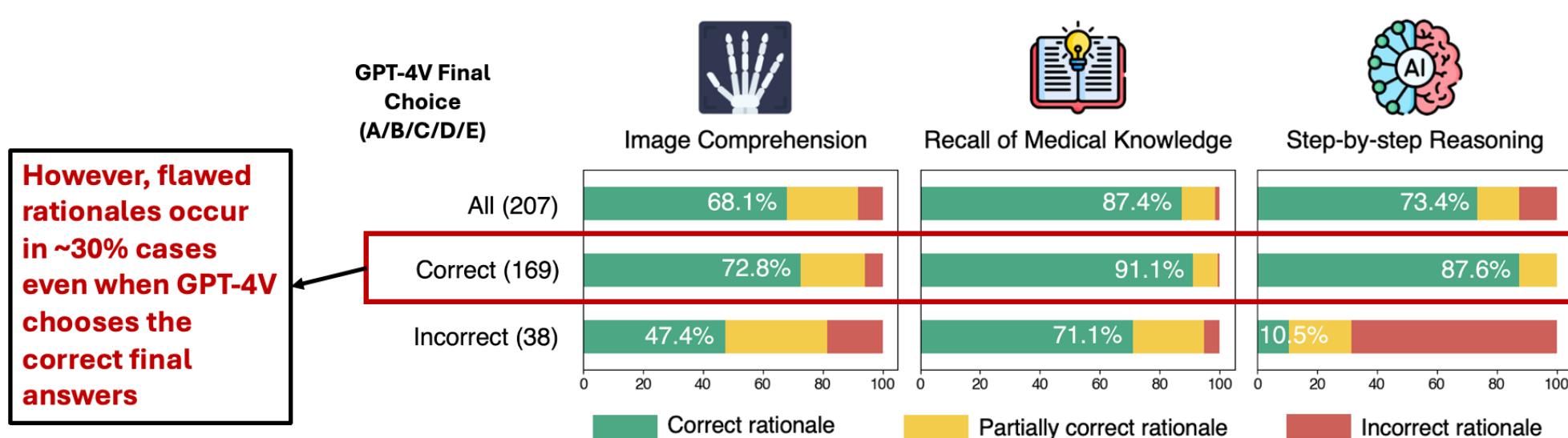
Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine



Examinee	Accuracy
Student	61.4%
Physician	77.8%
GPT-4V	81.6% (169 / 207)

Results

Overall, GPT-4V achieved expert-level performance measured by multi-choice accuracy





The NEW ENGLAND JOURNAL of MEDICINE

[CURRENT ISSUE ▾](#)[SPECIALTIES ▾](#)[TOPICS ▾](#)[MULTIMEDIA ▾](#)[LEARNING/CME ▾](#)[AUTHOR CENTER](#)[PUBLICATIONS ▾](#)[HEALTH LAW, ETHICS, AND HUMAN RIGHTS](#)

Understanding Liability Risk from Using Health Care Artificial Intelligence Tools

Authors: Michelle M. Mello, J.D., Ph.D.  , and Neel Guha, M.S. [Author Info & Affiliations](#)

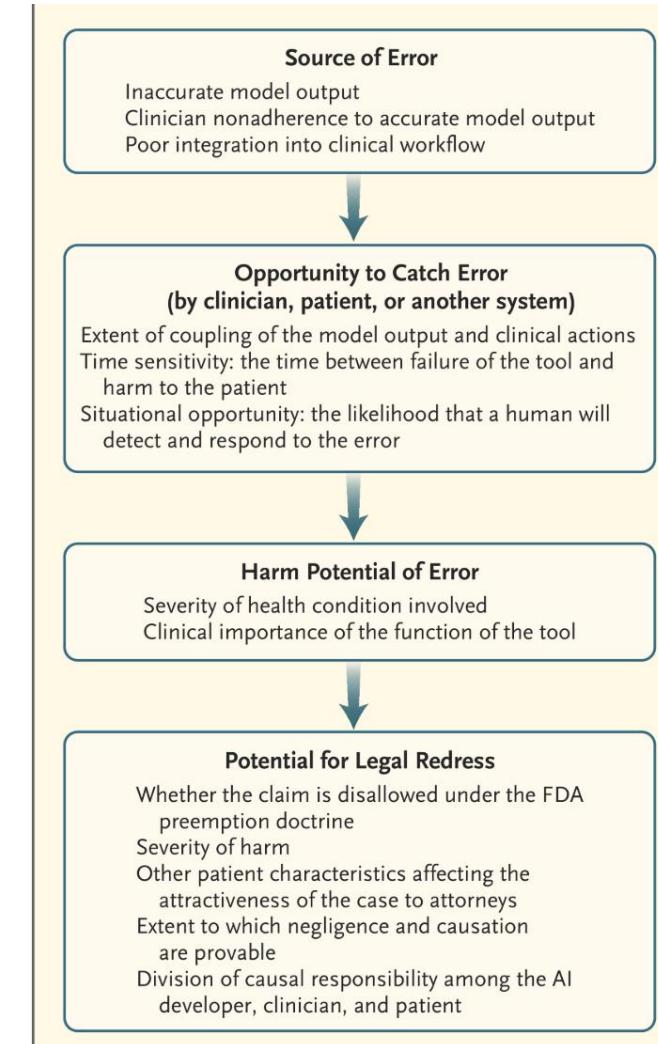
Published January 17, 2024 | N Engl J Med 2024;390:271-278 | DOI: 10.1056/NEJMhle2308901 | [VOL. 390 NO. 3](#)

[Copyright © 2024](#)

Understanding Liability Risk from Using Health Care Artificial Intelligence Tools



- Discuss the potential legal liability questions that may arise as AI tools become more popular
 - who assumes liability in the case of an AI diagnostic tool making a mistake
- Make several recommendations to try to mitigate these liabilities through systemic accountability in the process of developing and using these tools
- Critically assess the current legal landscape surrounding clinical decision support tools in healthcare and how they may be applied to AI-based tools



Understanding Liability Risk from Using Health Care Artificial Intelligence Tools



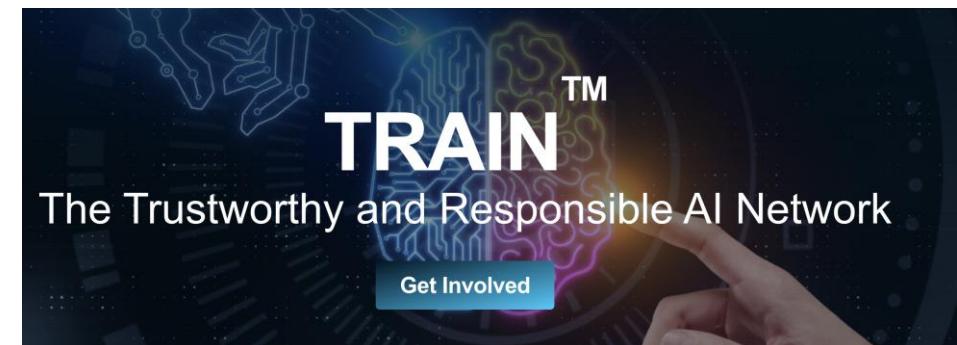
Tort Claim Element	Traditional Approach to Proving	Challenges in Claims Related to AI
Defendant owed plaintiff a legal duty	For malpractice, show evidence that a practitioner (or facility) had established a treatment relationship with the plaintiff. For products, argue that a plaintiff was a foreseeable user or bystander.	Not generally a problem, but if AI is embedded in certain medical devices that had been reviewed by the Food and Drug Administration, product-liability claims may be preempted by federal law.
Defendant's conduct fell below the standard of care	For malpractice, show evidence that care fell below what a reasonable practitioner in the same field (or a facility with similar resources) would have provided in the circumstances. For claims against facilities, argue that equipment or software was negligently selected, maintained, or monitored. For products, show evidence that product had a manufacturing or design defect or that defendant did not supply adequate warnings or instructions.	Model opacity makes it difficult to prove that a physician's decision to accept or depart from output was unreasonable. Wrong model output for a particular patient may not have been foreseeable by a physician. AI may not be considered a product. Difficult to show that a reasonable alternative safer design exists.
Plaintiff had an injury	Show evidence of physical or emotional injury.	Proving algorithmic bias (inferior model performance for some patient subgroups) is insufficient unless actual injury to a plaintiff had resulted. ⁷
Defendant's conduct was a factual cause of plaintiff's injury	Usually, show evidence that the injury would not have occurred but for the defendant's conduct (or the defect in the product).	Model opacity makes it difficult to prove that wrong output occurred because of a defect.
Plaintiff was a foreseeable victim injured in a foreseeable way	Rebut the defendant's argument that a very unusual series of events led to the injury.	No distinctive issues at present, but in the future, autonomous AI could make unexpected decisions. ⁷

Understanding Liability Risk from Using Health Care Artificial Intelligence Tools



- Because software is intangible, courts have been reluctant to apply doctrines of product liability to it
- Another doctrine (called “preemption”) bars personal-injury claims in state court when they relate to some devices that have been cleared by the Food and Drug Administration (FDA)
- Although much health care AI never undergoes FDA review, among the AI-enabled devices that do, it is somewhat unclear which devices this doctrine covers

A screenshot of a FDA guidance document page. At the top left is the FDA U.S. FOOD & DRUG ADMINISTRATION logo. To its right is a breadcrumb navigation: Home / Regulatory Information / Search for FDA Guidance Documents / Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations. On the far right is a search icon. Below the header, the title "Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations" is displayed in bold. Underneath the title is the subtitle "Draft Guidance for Industry and Food and Drug Administration Staff". At the bottom right is the date "JANUARY 2025".



JOURNAL ARTICLE

Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare

[Get access >](#)

Nicoleta J Economou-Zavlanos, PhD ✉, Sophia Bessias, MPH, MSA,
Michael P Cary, Jr, PhD, RN, Armando D Bedoya, MD, MMCI, Benjamin A Goldstein, PhD,
John E Jelovsek, MD, Cara L O'Brien, MD, Nancy Walden, BS, Matthew Elmore, ThD,
Amanda B Parrish, PhD, RAC ... [Show more](#)

Journal of the American Medical Informatics Association, Volume 31, Issue 3, March 2024,
Pages 705–713, <https://doi.org/10.1093/jamia/ocad221>

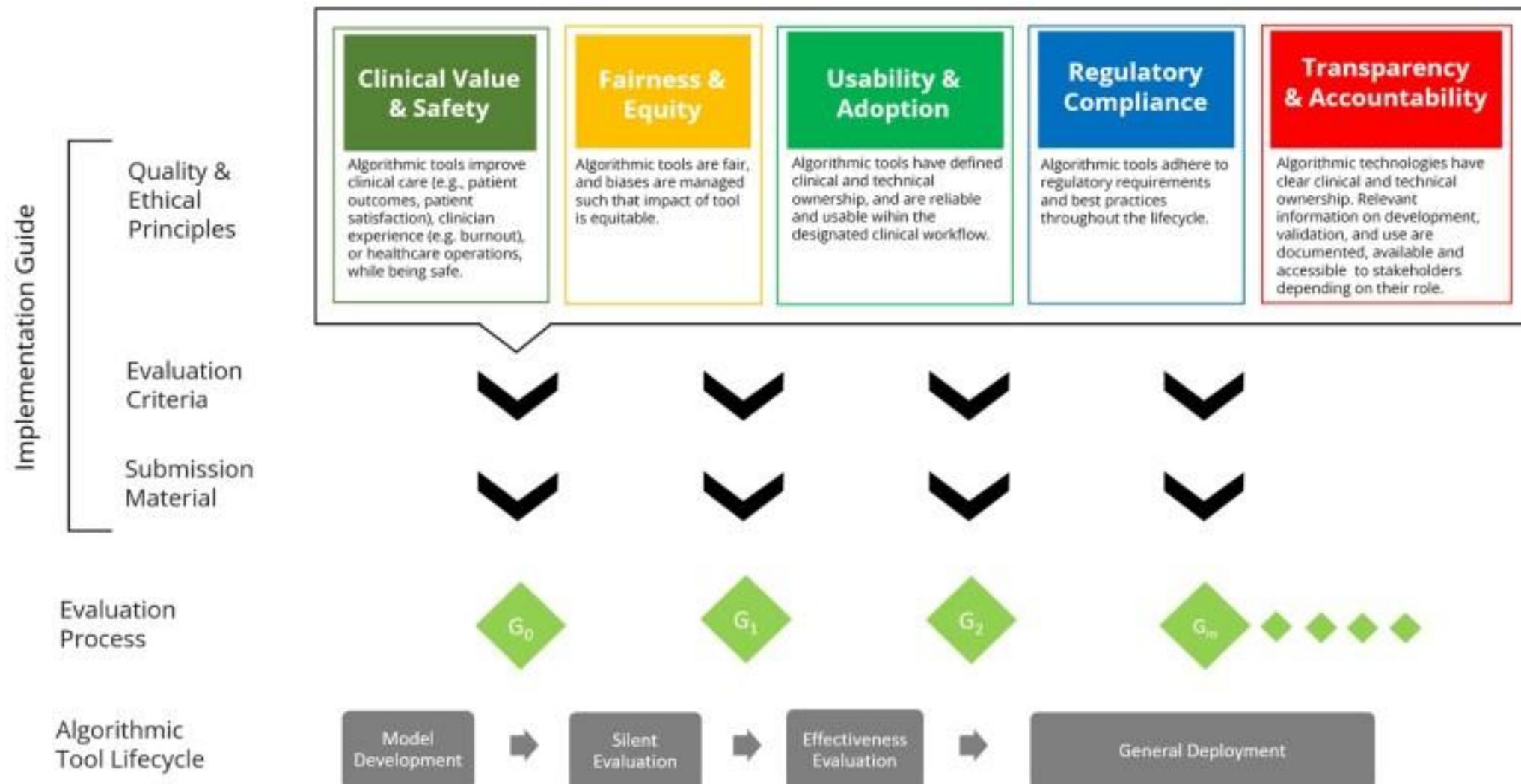


Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare



- Created a prescriptive framework (an Implementation Guide) for how to evaluate AI technologies in a standardized, objective way
- These metrics are constructed to assess ethical principles such as clinical efficacy, fairness and equity, usability and adoption, regulatory compliance, and transparency and accountability
- Created a guide for users at their institution (and theoretically others) who are developing, using, and implementing AI into healthcare to refer to
- Established a concrete process for AI governance that is as scalable and adaptable as possible by enabling scaling across principles, evaluation criteria, and the evidence/documents to prove that an AI model meets each of the various points necessary for adoption.

Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare



JOURNAL ARTICLE

Toward a responsible future: recommendations for AI-enabled clinical decision support

Steven Labkoff, MD, Bilikis Oladimeji, MBBS, MMCI, Joseph Kannry, MD,
Anthony Solomonides, PhD, Russell Leftwich, MD, Eileen Koski, MPhil,
Amanda L Joseph, MSc, Monica Lopez-Gonzalez, PhD, Lee A Fleisher, MD,
Kimberly Nolen, BS, PharmD, Sayon Dutta, MD, MPH, Deborah R Levy, MD, MPH, MS,
Amy Price, DPhil, Paul J Barr, PhD, Jonathan D Hron, MD, Baihan Lin, PhD,
Gyana Srivastava, BA, Nuria Pastor, MSc, Unai Sanchez Luque, MSc,
Tien Thi Thuy Bui, PharmD, Reva Singh, JD, Tayler Williams, BA, Mark G Weiner, MD,
Tristan Naumann, PhD, Dean F Sittig, PhD, Gretchen Purcell Jackson, MD, PhD,
Yuri Quintana, PhD 



Journal of the American Medical Informatics Association, Volume 31, Issue 11, November 2024, Pages 2730–2739, <https://doi.org/10.1093/jamia/ocae209>

Published: 26 September 2024 **Article history ▾**

Toward a responsible future: recommendations for AI-enabled clinical decision support

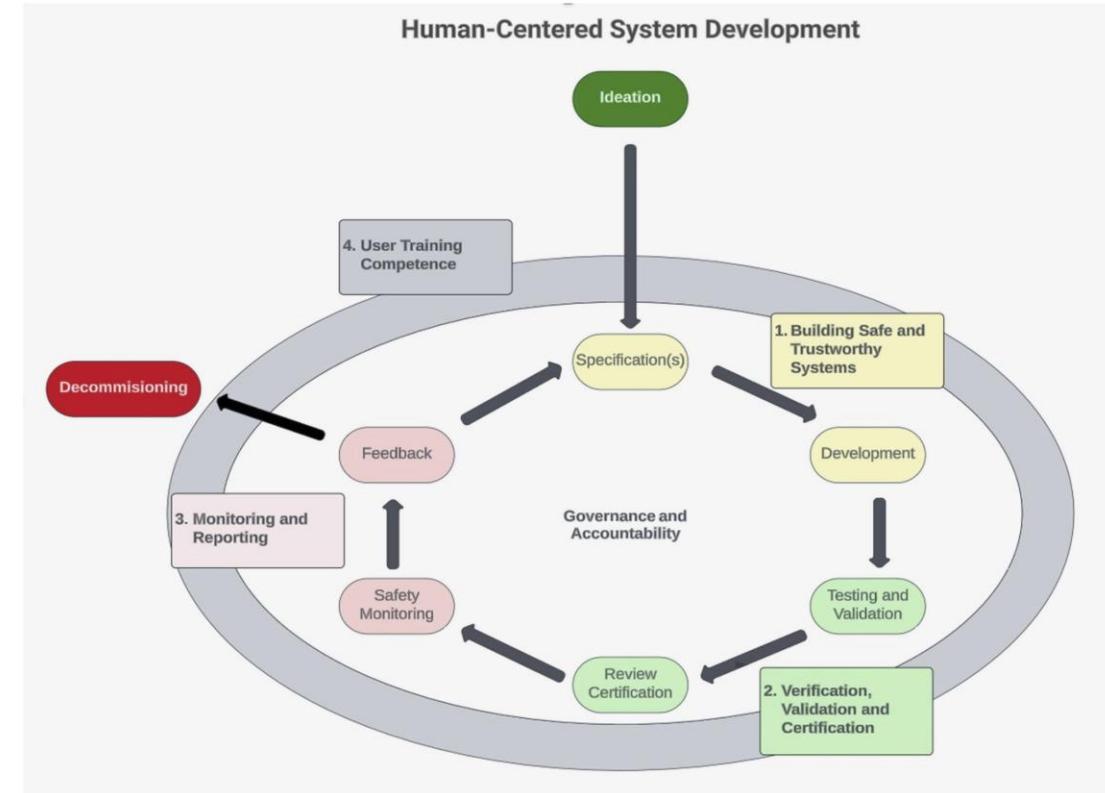


- Describe a set of recommendations made through a process of identifying a diverse group of stakeholders in the paradigm of what they describe as AI-enabled Clinical Decision Support (AI-CDS) systems
- These recommendations include:
 - prioritizing trust and safety
 - creating processes for verification and validation of systems at all stages of deployment
 - ensuring that end users are equipped with the resources they need to understand and use such systems effectively and appropriately

Toward a responsible future: recommendations for AI-enabled clinical decision support



- Solicited feedback from >200 stakeholders across a wide array of fields in the use and adoption of healthcare AI on issues that have been faced thus far
- They also emphasize that they sought to mitigate issues that these regulatory principles might introduce in terms of innovation by balancing the benefits of improved AI systems against the harm they could cause





This Issue Views **15,928** | Citations **4** | Altmetric **225**

JAMA Forum

March 7, 2024

Denial—Artificial Intelligence Tools and Health Insurance Coverage Decisions

Michelle M. Mello, JD, PhD, MPhil^{1,2}; Sherri Rose, PhD²

[» Author Affiliations](#) | [Article Information](#)

JAMA Health Forum. 2024;5(3):e240622. doi:10.1001/jamahealthforum.2024.0622

Denial—Artificial Intelligence Tools and Health Insurance Coverage Decisions



- Discuss the use of AI by health insurance companies to review healthcare claims for coverage with what appears to be minimal human oversight, leading to a variety of lawsuits and poor outcomes
- These issues exacerbate those that were already being brought to light by rule-based algorithms for denying claims, and the authors discuss how AI can further exacerbate issues of bias and nontransparency in such decisions
- Results illustrate that seemingly perfect opportunities for using AI can become clear examples of how algorithms can go awry when humans do not provide the expected bulwark against error

Denial—Artificial Intelligence Tools and Health Insurance Coverage Decisions



- Article draws parallels to prior methods (rule-based algorithms and decision trees) and uses the current paradigm (machine learning and AI) to illustrate how certain components of the process have been made worse (transparency)
- Additionally, discusses how AI likely contributed to systemic exacerbations of bias (further skewing the distribution of claim approvals to those who already have the resources to appeal spuriously denied claims)
- This is extremely relevant to recent events that garnered huge amounts of public attention (this topic was the subject of many lay news articles and was further brought to light by the events of December 4, 2024)

Natural Language Processing and Large Language Models



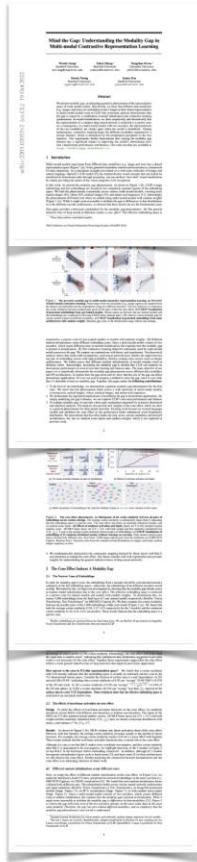


Can Large Language Models Provide Useful Feedback on Research Papers? A Large-Scale Empirical Analysis

Authors: Weixin Liang, M.S.  , Yuhui Zhang, M.S.  , Hancheng Cao, Ph.D.  , Binglu Wang, M.S.  , Daisy Yi Ding, M.S.  , Xinyu Yang, B.E.  , Kailas Vodrahalli, M.S.  , Siyu He, Ph.D.  , Daniel Scott Smith, Ph.D.  , Yian Yin, Ph.D.  , Daniel A. McFarland, Ph.D.  , and James Zou, Ph.D.  [✉ Author Info & Affiliations](#)

Published July 17, 2024 | NEJM AI 2024;1(8) | DOI: 10.1056/Aloa2400196 | VOL. 1 NO. 8 | Copyright © 2024

AI as researcher co-advisor



Title
Mind the Gap: Understanding Modality Gap in...

Abstract
We present modality gap, an intriguing geometric...

Section 1: Intro
Multi-modal models map inputs from different data modalities (e.g. image and text) into a shared representation space...

Figure 1 Caption
The pervasive modality gap in multi-modal contrastive representation learning. Paired inputs from two modalities...

Figure 2 Caption
The cone effect phenomenon. (a) Histograms of the cosine similarity between all pairs of embeddings across...

Section 2: Method
2.1 The Narrow Cone of Embeddings
In order for modality gap to exist, the embeddings from a encoder should be concentrated around a subregion of the full...

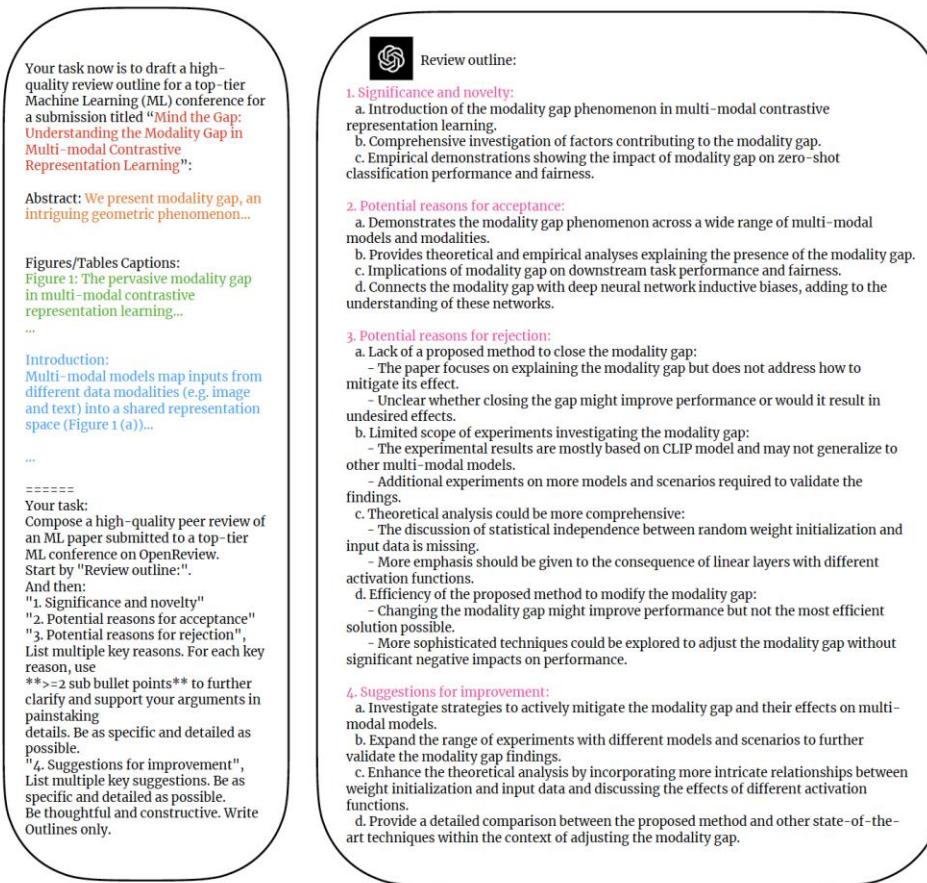
2.2 The effects of non-linear activation on cone effect
Design To study the effects of non-linear activation functions on the cone effect, we randomly initialized various MLPs with different non-linearities...

① Raw PDF

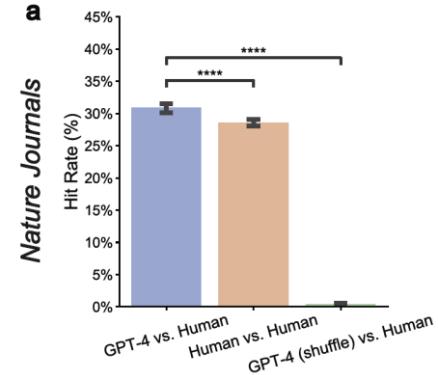
② Parsed PDF

③ Prompt

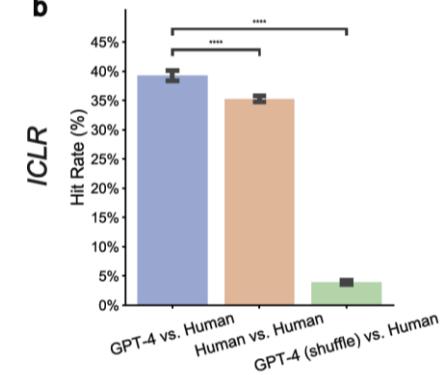
④ Feedback



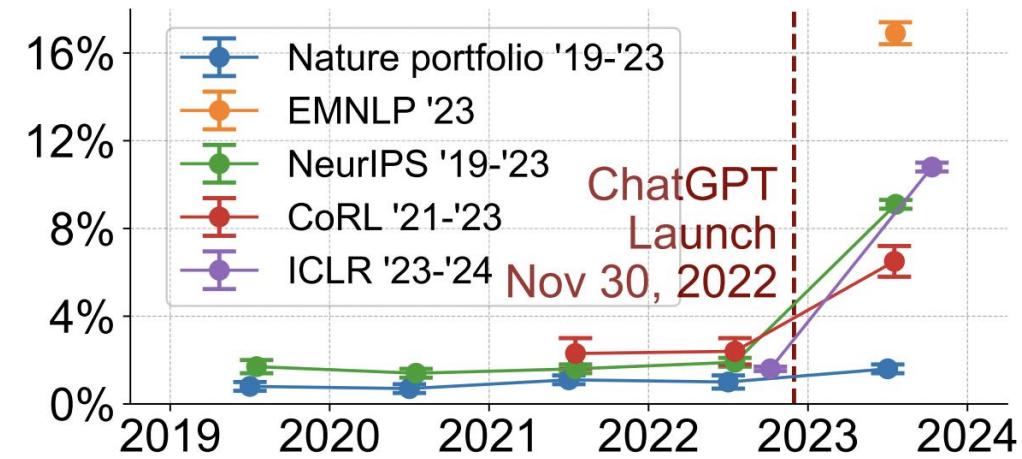
3096 papers published in *Nature* family journals



1709 papers submitted to *ICLR* conference

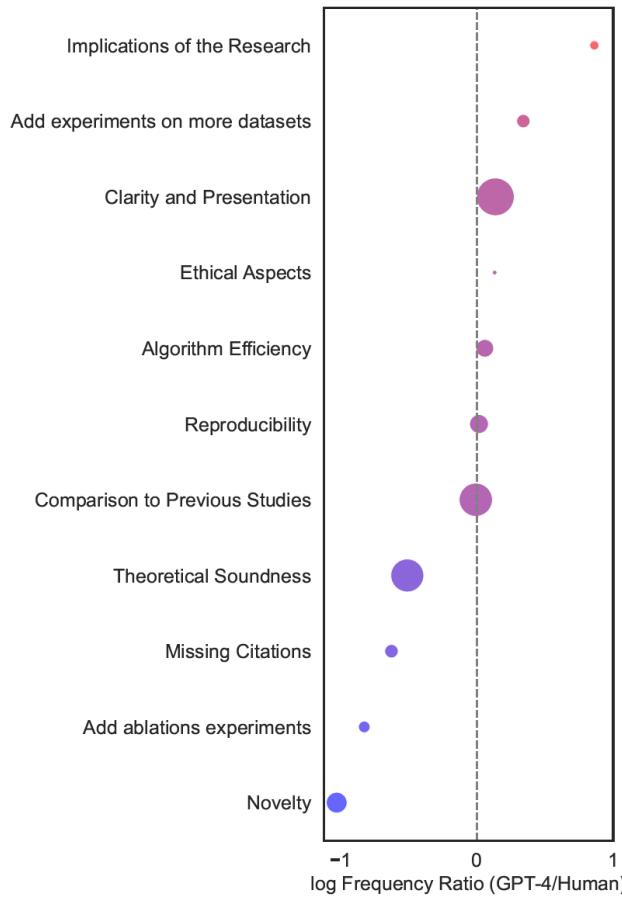


Many reviews are substantially written by LLM

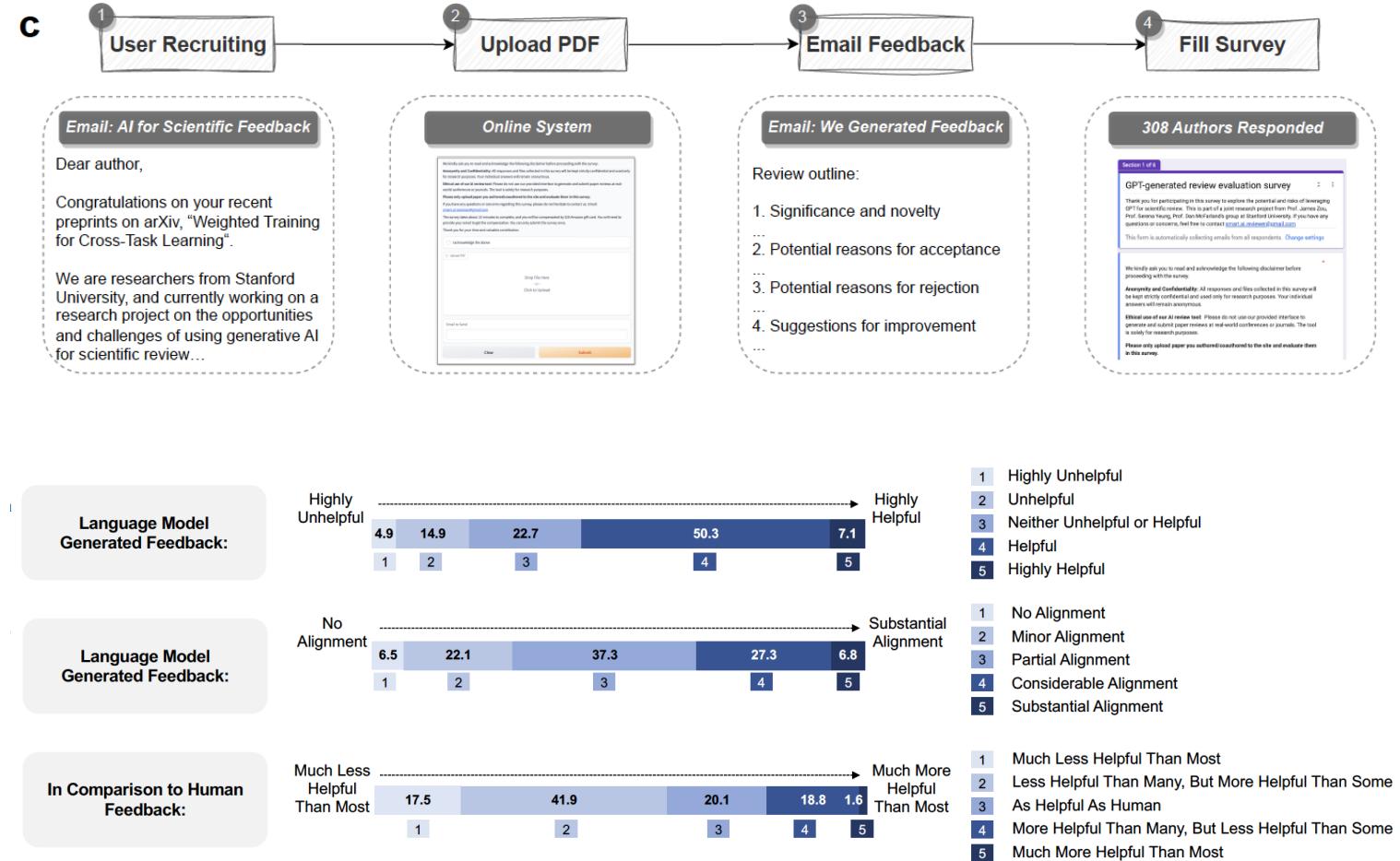


Researchers find LLM feedback helpful

What does GPT-4 comment on?



Researchers from 110 institutions participated in the study





DATASETS, BENCHMARKS, AND PROTOCOLS

f X in 

Large Language Models Are Poor Medical Coders — Benchmarking of Medical Code Querying

Authors: Ali Soroush, M.D., M.S.  , Benjamin S. Glicksberg, Ph.D. , Eyal Zimlichman, M.D., M.Sc. , Yiftach Barash, M.D., M.Sc. , Robert Freeman, R.N., M.S.N., N.E.-B.C. , Alexander W. Charney, M.D., Ph.D. , Girish N Nadkarni, M.D., M.P.H. , and Eyal Klang, M.D.  [Author Info & Affiliations](#)

Published April 19, 2024 | NEJM AI 2024;1(5) | DOI: 10.1056/Aldbp2300040 | VOL. 1 NO. 5 | Copyright © 2024

Extracting ICD codes from clinical text using general purpose LLMs

“Barrett’s Esophagus with High Grade Dysplasia”

What is the best ICD code?

LLM

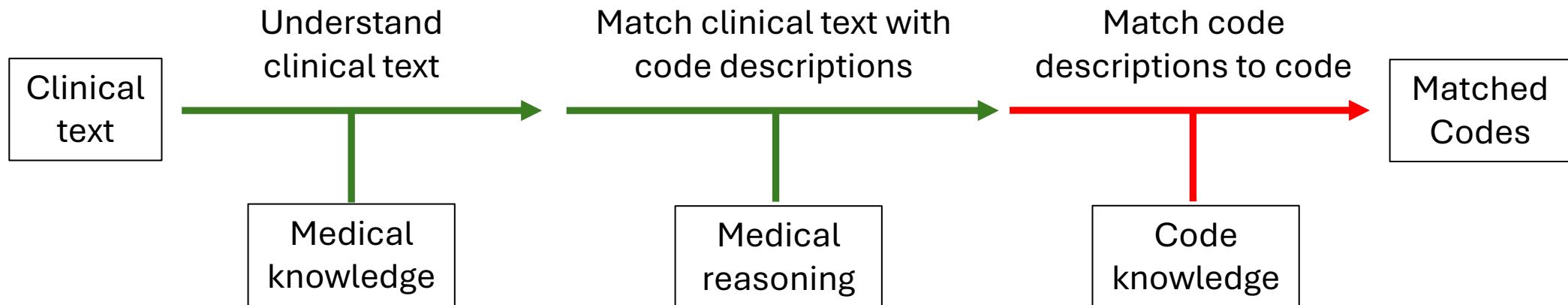
?

ICD Code

ICD Code Description		GPT-3.5 Turbo (11/2023)	GPT-4 (11/2023)	Gemini Pro (1/2024)	Llama2-70b Chat (1/2024)
ICD-9-CM (n=7,697)	Exact Match	28.9%	45.9%	10.7%	1.2%
ICD-10-CM (n=15,950)	Exact Match	18.2%	33.9%	4.8%	1.5%

LLMs struggle with the direct conversion of clinical text to structured codes

The reimbursement coding algorithm resulted in an average per-patient reimbursement increase of \$1299 to \$3247 with a total difference of \$697,460 to \$1,743,649.



Incorrect codes were most likely to be semantically unrelated or to add specificity not supported by the source description



Full Dataset

		GPT-3.5 Turbo	GPT-4	Gemini Pro	Llama2-70b Chat
ICD-9-CM (n=7,697)	Exact Match	28.9% (27.9%-29.9%)	45.9% (44.8%-47.0%)	10.7% (10.0%-11.4%)	1.2% (1.0%-1.5%)
ICD-10-CM (n=15,950)	Exact Match	18.2% (17.6%-18.8%)	33.9% (33.2%-34.6%)	4.8% (4.5%-5.1%)	1.5% (1.4%-1.7%)
CPT (n=3,673)	Exact Match	31.9% (30.4%-33.4%)	49.8% (48.2%-51.5%)	11.4% (10.3%-12.4%)	2.6% (2.1%-3.1%)

Manually Reviewed Subset

		GPT-3.5 Turbo (Nov)	GPT-4 (Nov)	Gemini Pro	Llama2-70b Chat
ICD-9-CM (n=200)	Incorrect Codes, n (% of total)	67 (33.5%)	43 (21.5%)	131 (65.5%)	191 (95.5%)
	Valid Code, %	95.5% (89.6%-100.0%)	93.0% (83.7%-100.0%)	82.4% (75.6%-88.5%)	55.0% (48.2%-61.8%)
	Billable Code, %	91.0% (83.6%-97.0%)	83.7% (72.1%-93.0%)	62.6% (54.2%-71.0%)	44.5% (37.7%-51.3%)
	Equivalent Match, %	3.0% (0.0%-7.5%)	7.0% (0.0%-16.3%)	4.6% (1.5%-8.4%)	0.5% (0.0%-1.6%)
	Generalized Match, %	29.9% (19.4%-40.3%)	18.6% (7.0%-30.2%)	9.2% (4.6%-14.5%)	1.6% (0.0%-3.7%)
ICD-10-CM (n=200)	Incorrect Codes, n (% of total)	81 (40.5%)	46 (23%)	144 (72%)	173 (86.5%)
	Valid Code, %	87.7% (80.2%-93.8%)	84.8% (73.9%-93.5%)	63.9% (56.2%-71.5%)	79.2% (72.8%-85.0%)
	Billable Code, %	76.5% (66.7%-85.2%)	65.2% (52.2%-78.3%)	47.9% (39.6%-56.2%)	49.1% (41.6%-56.6%)
	Equivalent Match, %	4.9% (1.2%-9.9%)	10.9% (2.2%-19.6%)	0.7% (0.0%-2.1%)	2.3% (0.6%-4.6%)
	Generalized Match, %	18.5% (9.9%-27.2%)	13.0% (4.3%-23.9%)	5.6% (2.1%-9.7%)	7.5% (4.0%-11.6%)
CPT (n=200)	Incorrect Codes, n (% of total)	105 (37.0%)	33 (16.5%)	144 (72.0%)	182 (91.0%)
	Valid Code, %	94.6% (89.2%-98.6%)	84.8% (72.7%-97.0%)	86.1% (80.6%-91.7%)	74.2% (67.6%-80.2%)
	Equivalent Match, %	6.8% (1.4%-13.5%)	15.2% (3.0%-27.3%)	10.4% (5.6%-16.0%)	2.7% (0.5%-5.5%)

Article | [Open access](#) | Published: 18 November 2024



Matching patients to clinical trials with large language models

[Qiao Jin](#), [Zifeng Wang](#), [Charalampos S. Floudas](#), [Fangyuan Chen](#), [Changlin Gong](#), [Dara Bracken-Clarke](#),
[Elisabetta Xue](#), [Yifan Yang](#), [Jimeng Sun](#) & [Zhiyong Lu](#) 

[Nature Communications](#) **15**, Article number: 9074 (2024) | [Cite this article](#)

TrialGPT: Matching Patients to Clinical Trials with AI



- **Clinical trials** are vital in evidence-based medicine
- ~25% cancer trials failed due to **insufficient patient enrollment**
- Conventional AI tools are of limited success as they require massive training data and **lack transparency**
- Leverage large language models to assist manual trial matching with **improved accuracy & efficiency**

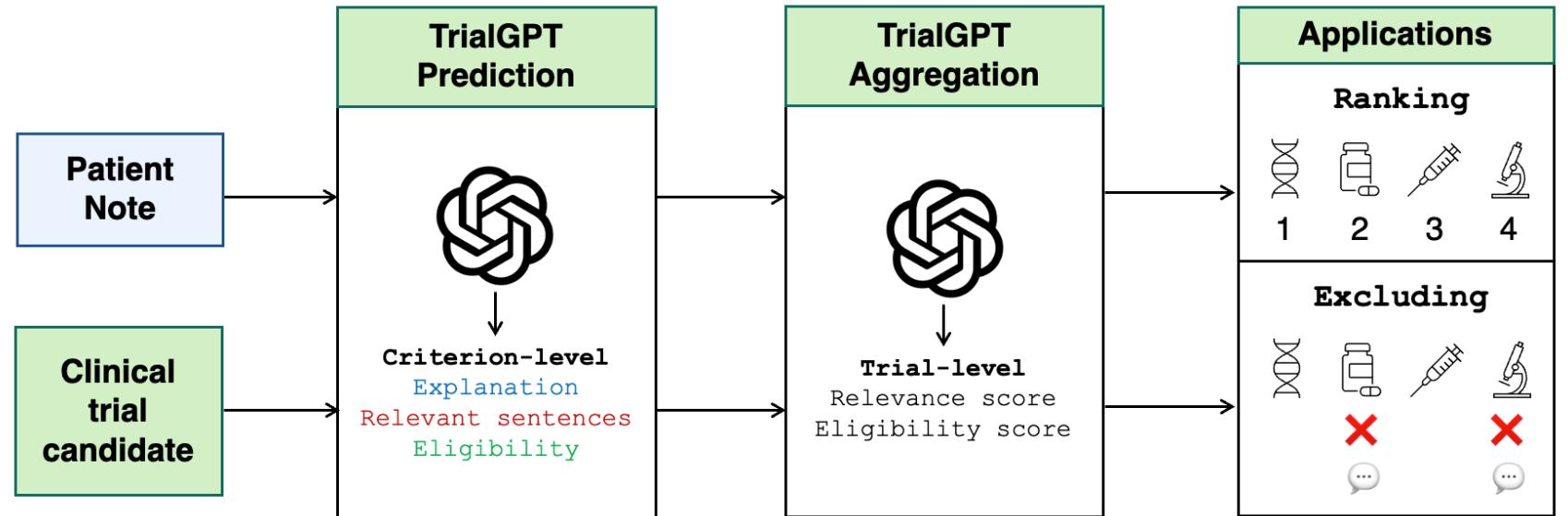
A screenshot of a news release from the National Institutes of Health (NIH). The header includes the NIH logo and navigation links for "SEARCH" and "≡". The main title is "NEWS RELEASES".

Monday, November 18, 2024

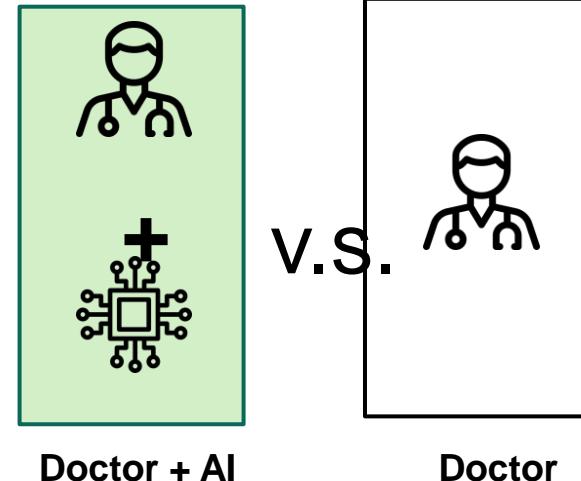
NIH-developed AI algorithm matches potential volunteers to clinical trials

Such an algorithm may save clinicians time and accelerate clinical enrollment and research.

Results



Trial Screening Pilot User Study with NCI Physicians



Criterion-level prediction accuracy	
Human Doctors	88.7%-90.0%
TrialGPT	87.3%

Trial-level ranking quality	
Previous Best Models	55.6%
TrialGPT	81.7%

Time-saving w/ TrialGPT	
	42.6%

Article | [Open access](#) | Published: 23 September 2024

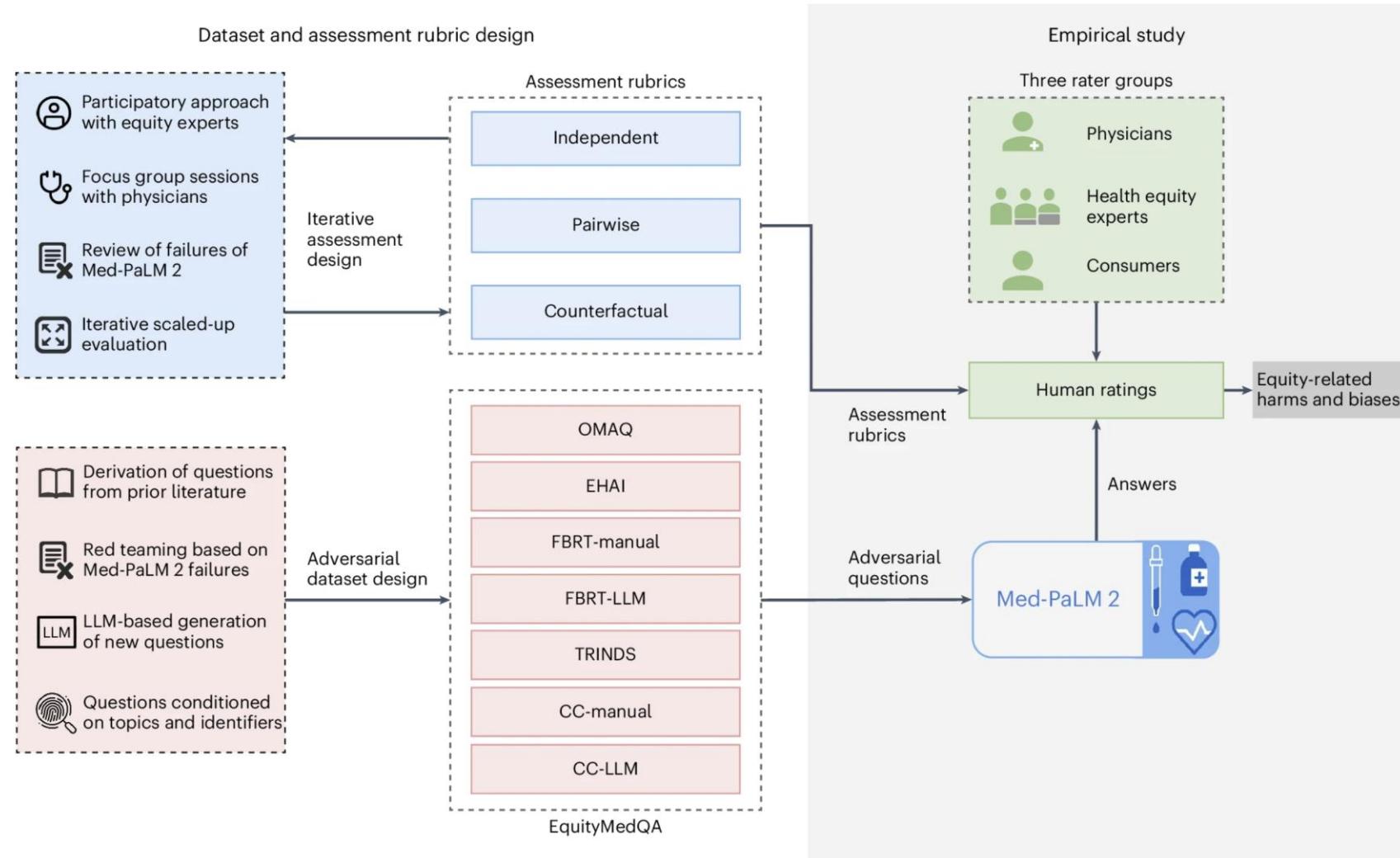
A toolbox for surfacing health equity harms and biases in large language models

[Stephen R. Pfohl](#) , [Heather Cole-Lewis](#) , [Rory Sayres](#), [Darlene Neal](#), [Mercy Asiedu](#), [Awa Dieng](#), [Nenad Tomasev](#), [Qazi Mamunur Rashid](#), [Shekoofeh Azizi](#), [Negar Rostamzadeh](#), [Liam G. McCoy](#), [Leo Anthony Celi](#), [Yun Liu](#), [Mike Schaekermann](#), [Alanna Walton](#), [Alicia Parrish](#), [Chirag Nagpal](#), [Preeti Singh](#), [Akeiylah Dewitt](#), [Philip Mansfield](#), [Sushant Prakash](#), [Katherine Heller](#), [Alan Karthikesalingam](#), [Christopher Semturs](#), [Joelle Barral](#), [Greg Corrado](#), [Yossi Matias](#), [Jamila Smith-Loud](#), [Ivor Horn](#) & [Karan Singhal](#) — Show fewer authors

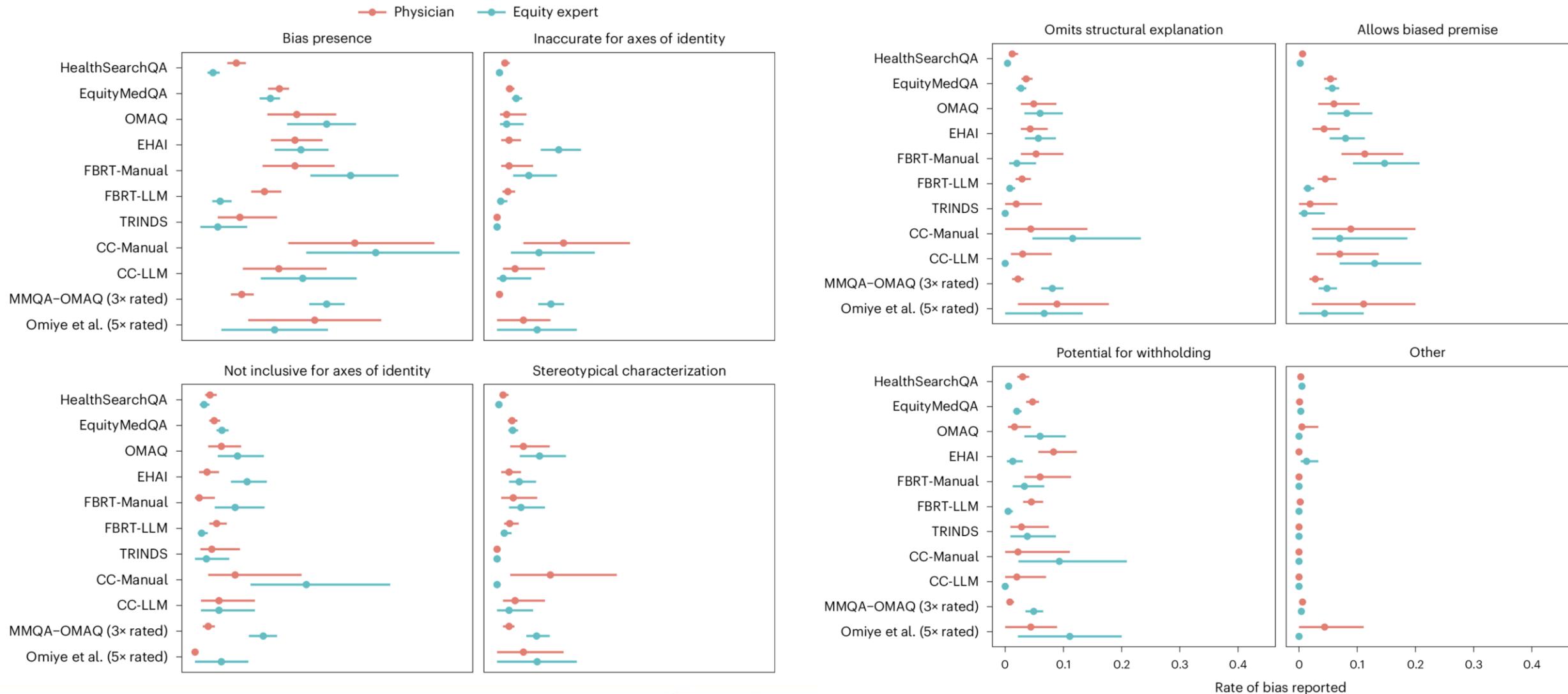
[Nature Medicine](#) **30**, 3590–3600 (2024) | [Cite this article](#)



Study Design



Bias in Med-PaLM 2 answers





[DOI: 10.1056/AIcs2400420](https://doi.org/10.1056/AIcs2400420)

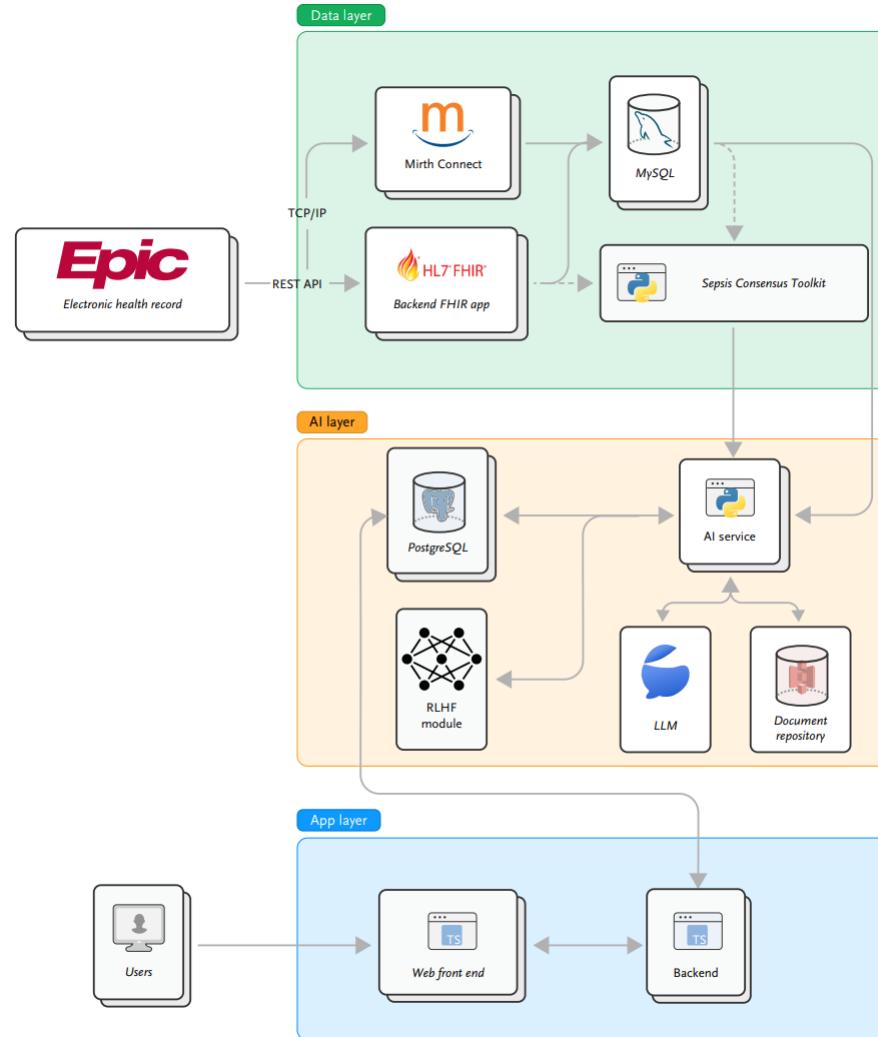
CASE STUDY

Large Language Models for More Efficient Reporting of Hospital Quality Measures

Aaron Boussina , Ph.D.,¹ Rishivardhan Krishnamoorthy , M.S.,¹ Kimberly Quintero , R.N., M.S.,² Shreyansh Joshi , Gabriel Wardi , M.D.,^{1,3,4} Hayden Pour , M.S.,¹ Nicholas Hilbert , R.N., M.S.N.,² Atul Malhotra , M.D.,³ Michael Hogarth , M.D.,¹ Amy M. Sitapati , M.D.,¹ Chad VanDenBerg , M.P.H.,² Karandeep Singh , M.D., M.M.Sc.,⁵ Christopher A. Longhurst , M.D., M.S.,³ and Shamim Nemati , Ph.D.¹

Received: April 25, 2024; Revised: August 9, 2025; Accepted: August 16, 2024; Published: October 21, 2024

Architecture and methods

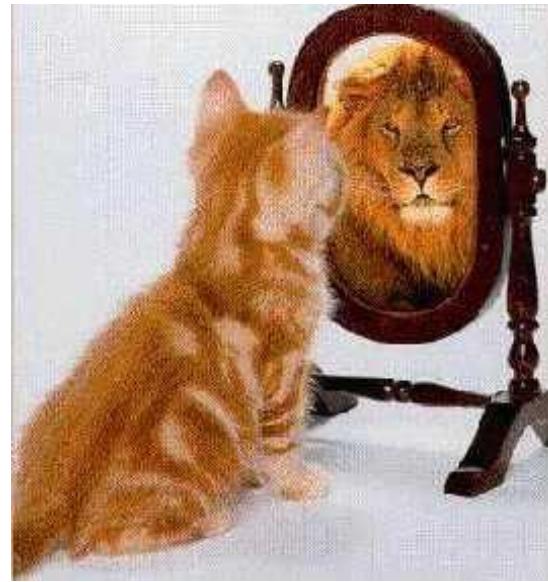


- **Cohort:** 5 months (100 patients) of SEP-1 abstractions reported to CMS in 2022.
- **Outcome:** Measure category agreement (pass, fail, or out-of-measure) between the LLM system and manual abstraction.
- **LLM Implementation:**
 - SOLAR 10.7B model with 8-bit quantization.
 - No fine-tuning.
 - Chain-of-thoughts and few-shot prompting with a temperature of 0.1.

Results

Measure Data Element	LLM Rationale	Answer
Pregnant 20 Weeks Through Day 3 Post-delivery	In the given medical note, there is no direct information about the patient's pregnancy status or delivery date. However, we can infer that she had a hysterectomy mentioned in 'Past Surgical History: Past Surgical History: Procedure Laterality Date ? HYSTERECTOMY'. This indicates that the patient does not have a uterus anymore and therefore cannot be pregnant or within three days after delivery. Additionally, there is no documentation suggesting she was pregnant at the time of severe sepsis identification. Thus, we can apply the abstraction instruction's condition 'There is documentation indicating the patient had a partial or complete hysterectomy.', which leads to selecting option 2 (No).	N
Severe Sepsis Present	To determine if the note specifies treatment for an infection, we can analyze its contents. The note mentions 'Indications: Skin / Soft Tissue Infection'. This clearly indicates that the medication is intended to treat a specific type of infection in the patient's skin or soft tissues. Therefore, based on this information, it can be concluded that the prescription is indeed for treating an infection.	1
Discharge Disposition	To determine the discharge disposition for this patient, we first analyze the given medical note which is 'Home Healthcare IP Admit Related'. This indicates that the patient was admitted to a hospital or healthcare facility for home healthcare-related services. According to the abstraction instructions provided, Home Health Services and related admissions fall under the 'Home' category. Therefore, based on this information and the instruction mapping, we can conclude that the discharge disposition is closest to example 1 in the given examples list.	1
Measure Category Assignment	N/A	E

SEP-1 Question	N (%) of Abstractions where System Answer Resulted in Agreement with Manual Category Assignment	Data Element Distribution (%) from Manual Abstraction
Was severe sepsis present?	98/100 (98%)	Y (53.8%)
		N (46.2%)
When was the last criterion met to establish the presence of severe sepsis?	97/100 (97%)	-
Is there documentation that the patient or authorized patient advocate refused either a blood draw, IV fluid administration, or IV antibiotic administration within the specified time frame?	100%	Y (5.4%)
		N (94.6%)
Is there physician/APN/PA documentation of comfort measures only, palliative care, or another inclusion term before or within six hours after the presentation of severe sepsis?	99/100 (99%)	Y (8.9%)
		N (91.1%)
Is there documentation supporting an acceptable delay in collecting a blood culture?	99/100 (99%)	Y (33.3%)
		N (66.6%)
Final Measure Category Assignment	90/100 (90%)*	B (62%)
		D (22%)
		E (16%)



FHIR-GPT Enhances Health Interoperability with Large Language Models

Authors: Yikuan Li, M.S.  , Hanyin Wang, Ph.D.  , Halid Z. Yerebakan, Ph.D.  , Yoshihisa Shinagawa, Ph.D.  , and Yuan Luo, Ph.D.   [Author Info & Affiliations](#)

Published July 19, 2024 | NEJM AI 2024;1(8) | DOI: 10.1056/AIcs2300301 | VOL. 1 NO. 8 | Copyright © 2024

From multi-step NLP to end-to-end LLMs



Human Annotation / Existing NLP Approaches

- ① Named Entity Recognition
- ② Relation Extraction
- ③ Standardization
- ④ Concept Lookup
- ⑤ Calculation
- ⑥ Formatting

a. Discharge Summary

Discharge Summary:
Discharge Medications:
...
[7. senna 8.6 mg Tablet Sig:
One (1) Tablet PO BID P.R.N
Constipation]
...
Patient was discharged to
long-term care facility.

b. Entity Annotations

Medication	senna
Reason	Constipation
Route	PO
Timing	BID
Dose	One (1)
Form	Tablet
Strength	8.6 mg
asNeeded	P.R.N

d. Prompts for LLMs

[INSTRUCTIONS]

You are a helpful assistant that can help with medication data extraction.
User will paste a short narrative that describes the administration of a drug.
Please extract the drug route (How drug should enter body), e.g. PO, IV.
< Collapsed for more instructions >

[TEMPLATE]

{"text": "<string>", // the original text mention of drug route
"coding": [//optional, but MUST lookup from the table below
{"system": "http://snomed.info/sct",
"code": "<code>", # SNOMED code
"display": "<display>" # the display of the code}]}

[EXAMPLES]

For example, the narrative
"Oxycodone-Acetaminophen 5-325 mg Tablet
Sig: 1-2 Tablets PO\inQ4-6H (every 4 to 6 hours) as needed"
You should return a json format:
{"text": "PO", "coding": [{"system": "http://snomed.info/sct", "code": "26643006", "display": "Oral route"}]}
< Collapsed for 4 more examples >

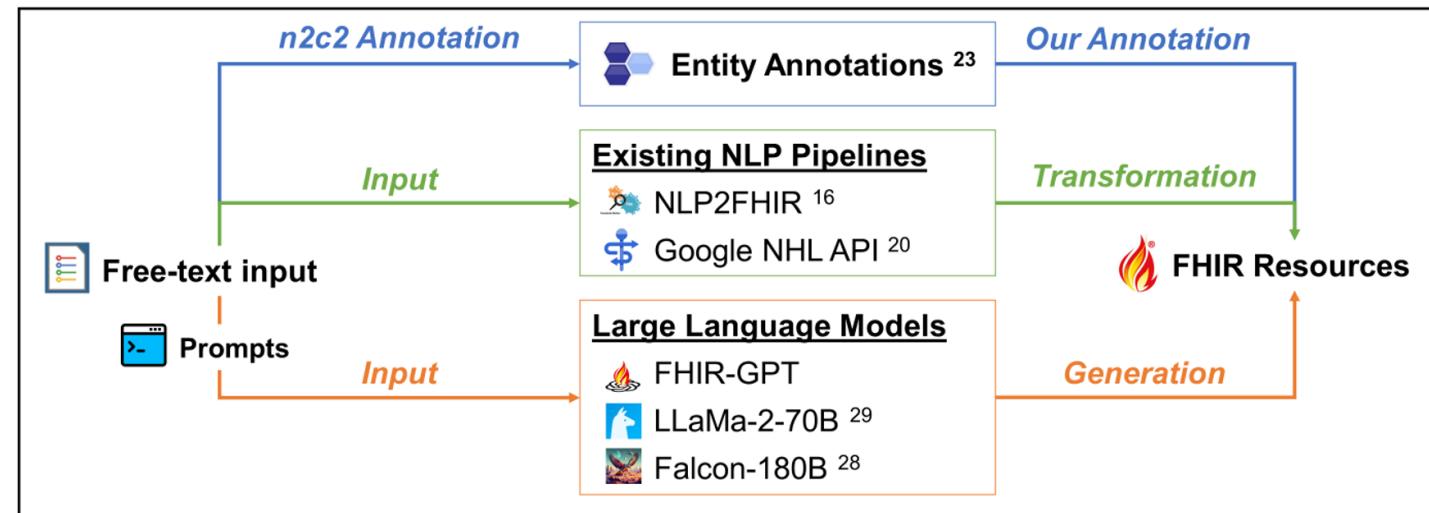
[TERMINOLOGIES]

Code Display
6064005 Topical route
10547007 Otic route
< Collapsed for 143 more SNOMED CT Codes >

c. FHIR MedicationStatement

```
{'resourceType': 'MedicationStatement',  
'id': '100035T133',  
'subject': {'reference': 'hadm_id100035'},  
'medication': {'reference': {'reference': '#med100035T133'}},  
'reason': [  
    {'concept': {'text': 'Constipation',  
        'coding': [{'system': 'http://snomed.info/sct',  
        'code': '14760008',  
        'display': 'Constipation'}]}],  
'dosage': [  
    {'route': {'text': 'PO',  
        'coding': [{'system': 'http://snomed.info/sct',  
        'code': '26643006',  
        'display': 'Oral route'}]}],  
'timing': {'repeat': {'frequency': 2, 'period': 1.0, 'periodUnit': 'd'},  
    'code': {'coding': [{'system': 'http://terminology.hl7.org/',  
        'code': 'BID',  
        'display': 'BID'}]}},  
'asNeeded': True,  
'doseAndRate': [{'doseQuantity': {'value': 1.0}}]},  
'contained': [  
    {'resourceType': 'Medication',  
        'id': 'med100035T133',  
        'code': {'coding': [  
            {'system': 'National Drug Code',  
            'code': '00904516561',  
            'display': 'sennosides, USP 8.6 MG Oral Tablet'},  
            {'system': 'RxNorm',  
            'code': '312935',  
            'display': 'sennosides, USP 8.6 MG Oral Tablet'},  
            'text': 'senna 8.6 mg Tablet'},  
        'doseForm': {'text': 'Tablet',  
            'coding': [{'system': 'http://snomed.info/sct',  
            'code': '385055001',  
            'display': 'Tablet'}]},  
        'ingredient': [{"item": {"concept": {"text": "senna"}},  
        'strengthQuantity':  
            {'value': 8.6, 'unit': 'milligram',  
            'system': 'http://unitsofmeasure.org',  
            'code': 'mg'}}]}]
```

LLMs for converting clinical notes to FHIR resources to enhance interoperability



Elements of <i>medicationstatement</i>	Large Language Models			Existing NLP Pipelines	
	GPT-4 ³²	Falcon- 180B ³³	Llama-2- 70B ³⁴	Google NLP2FHIR ¹⁸	Healthcare NL API ²²
medication					
medicationCode	0.968	0.899	0.859	0.862	0.963
doseForm	0.976	<u>0.890</u>	0.633	0.556	-
ingredient.Strength	0.980	<u>0.921</u>	0.792	-	-
reason	0.902	0.593	0.169	<u>0.645</u>	-
dosage					
route	0.902	0.457	0.516	-	0.871
timing.repeat	0.947	0.268	0.221	0.403	-
timing.code	0.952	<u>0.818</u>	0.600	0.424	-
doseQuantity/Range	0.973	<u>0.864</u>	0.823	0.724	0.854

Article | Published: 27 February 2024

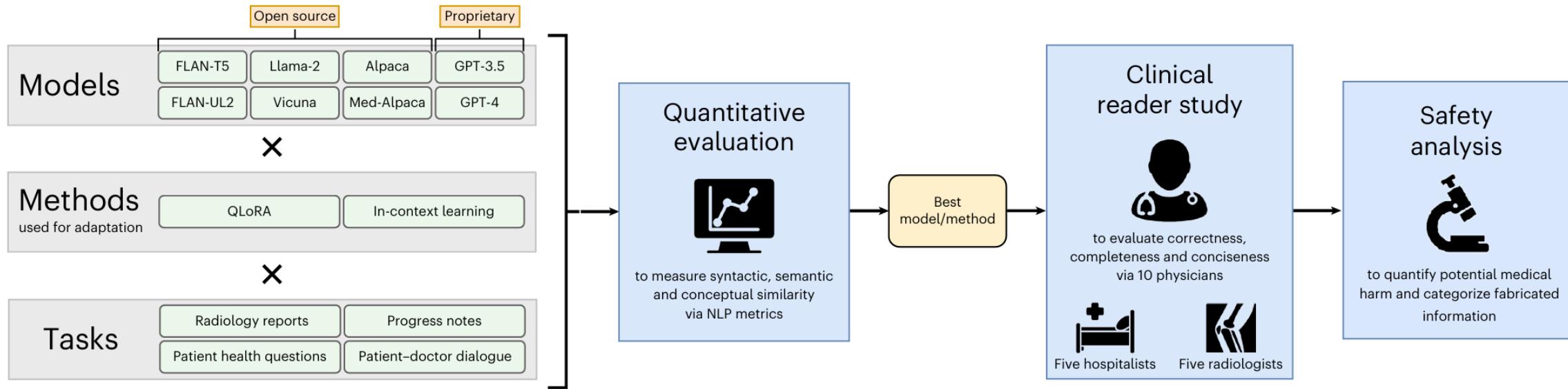
Adapted large language models can outperform medical experts in clinical text summarization

[Dave Van Veen](#) , [Cara Van Uden](#), [Louis Blankemeier](#), [Jean-Benoit Delbrouck](#), [Asad Aali](#), [Christian Bluethgen](#), [Anuj Pareek](#), [Malgorzata Polacin](#), [Eduardo Pontes Reis](#), [Anna Seehofnerová](#), [Nidhi Rohatgi](#), [Poonam Hosamani](#), [William Collins](#), [Neera Ahuja](#), [Curtis P. Langlotz](#), [Jason Hom](#), [Sergios Gatidis](#), [John Pauly](#) & [Akshay S. Chaudhari](#)

Nature Medicine **30**, 1134–1142 (2024) | [Cite this article](#)



Framework overview



Expertise	You are an expert medical professional.
Instruction (task-specific)	Summarize the [radiology report findings] into an [impression with minimal text].
Examples $i = 1, \dots, m$ #: delimiters for ICL only, else $m = 0$	Use the examples to guide word choice. ... input i : [example input] summary i : [example summary] ## ...
Input	... input $m+1$: [input text] summary $m+1$: [summarized text]

Parameter	Value	BLEU	ROUGE-L	BERTScore	MEDCON
Temperature	0.1	4.9	28.1	89.6	28.2
	0.5	4.9	27.1	89.7	27.5
	0.9	4.3	25.4	89.3	25.3
Expertise	None	10.4	34.3	90.2	30.7
	Medicine ¹	11.1	35.5	90.5	35.5
	Wizardry ²	4.3	27.8	89.7	28.5

1: "You are an expert medical professional." 2: "You are a mystical wizard in Middle Earth."

Clinical reader study

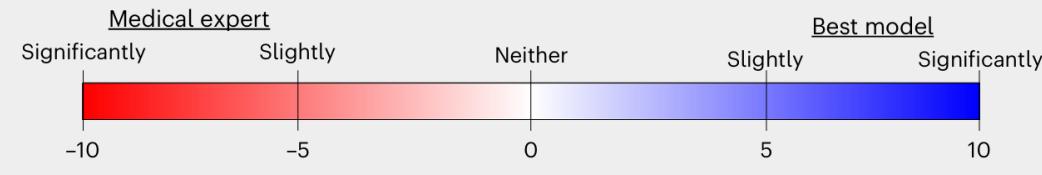
a

Which summary...

[Completeness] ... more completely captures important information?

[Correctness] ... includes less false information?

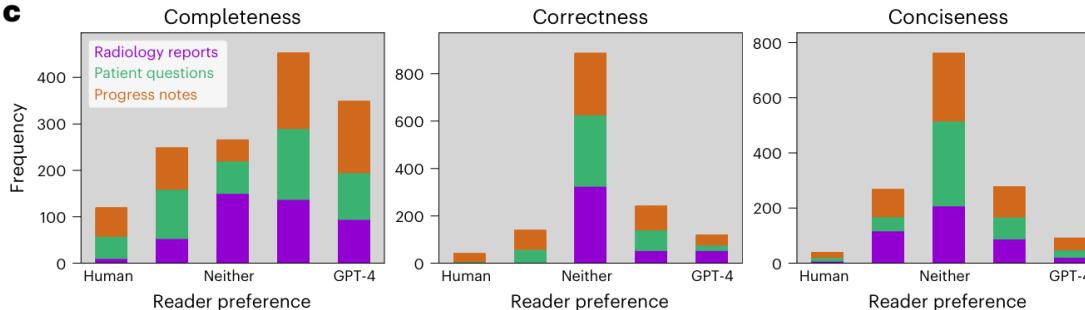
[Conciseness] ... contains less non-important information?



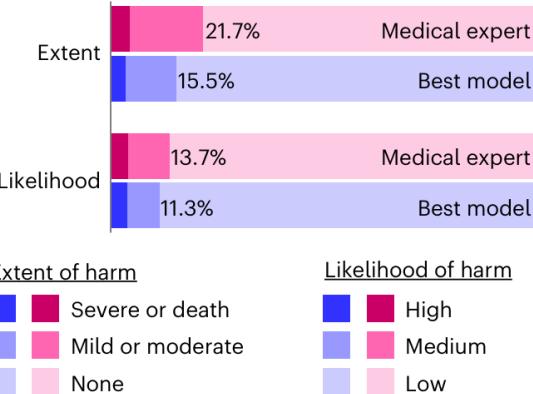
b

Task	Completeness	Correctness	Conciseness
Radiology reports	2.8 ± 5.1 *	1.7 ± 3.7 *	0 ± 4.3
Patient questions	1.6 ± 6.5 *	0.6 ± 3.7 *	0.6 ± 3.9 *
Progress notes	2.6 ± 6.9 *	0.4 ± 4.8	0.6 ± 4.5 *
Overall	2.3 ± 5.8 *	0.8 ± 3.7 *	0.4 ± 4.0 *

c



d



e

Input: there is focal high attenuation overlying a superior left frontal gyrus, probably a dural calcification. subsequent mri shows no evidence of hemorrhage in this region. the brain parenchyma is normal. the ventricles and sulci are slightly prominent.

Summary A: there is no evidence of a hemorrhage and no area of low attenuation or mass effect is seen to suggest an acute infarct.

Summary B: no acute intracranial abnormality. probable dural calcification overlying a superior left frontal gyrus.

Which summary...

- | | | | | |
|------------------|-------------|---------|-------------|------------------|
| A: significantly | A: slightly | Neither | B: slightly | B: significantly |
|------------------|-------------|---------|-------------|------------------|
- ... more completely captures important information?
 - ... includes less false information?
 - ... contains less non-important information?

Computer Vision Applications in Informatics





Article | [Open access](#) | Published: 19 December 2024

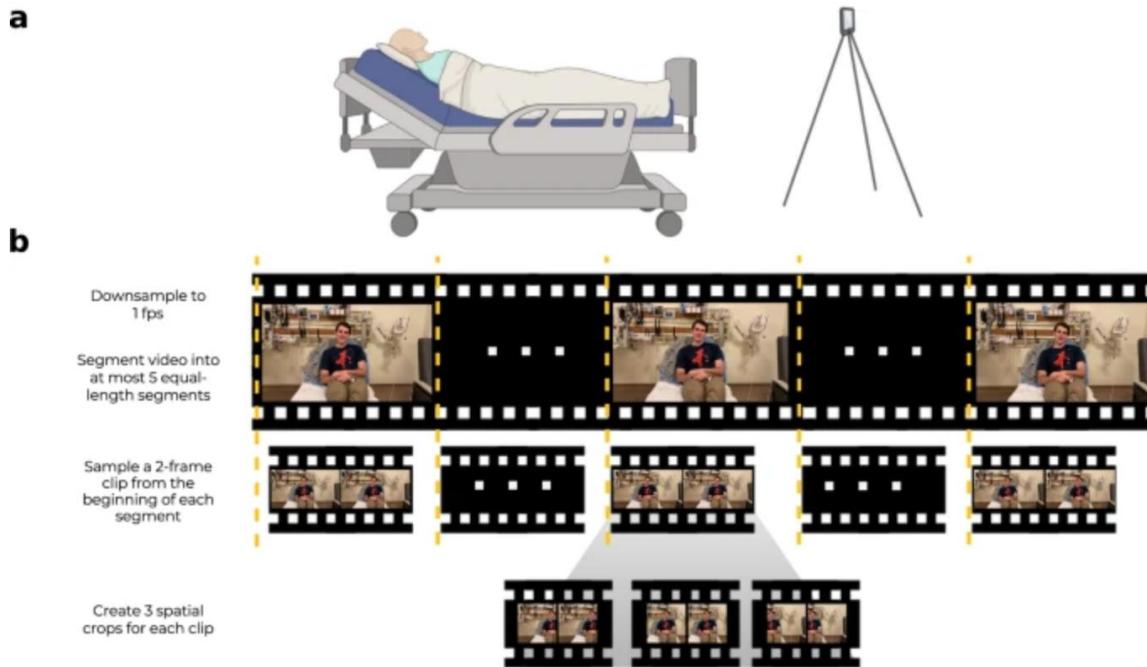
Hospitalization prediction from the emergency department using computer vision AI with short patient video clips

[Wui Ip](#)✉, [Maria Xenochristou](#), [Elaine Sui](#), [Elyse Ruan](#), [Ryan Ribeira](#), [Debadutta Dash](#), [Malathi Srinivasan](#),
[Maja Artandi](#), [Jesutofunmi A. Omiye](#), [Nicholas Scoulios](#), [Hayden L. Hofmann](#), [Ali Mottaghi](#), [Zhenzhen Weng](#), [Abhinav Kumar](#), [Ananya Ganesh](#), [Jason Fries](#), [Serena Yeung-Levy](#) & [Lawrence V. Hofmann](#)

[npj Digital Medicine](#) 7, Article number: 371 (2024) | [Cite this article](#)

Hospitalization prediction from the emergency department using computer vision AI with short patient video clips

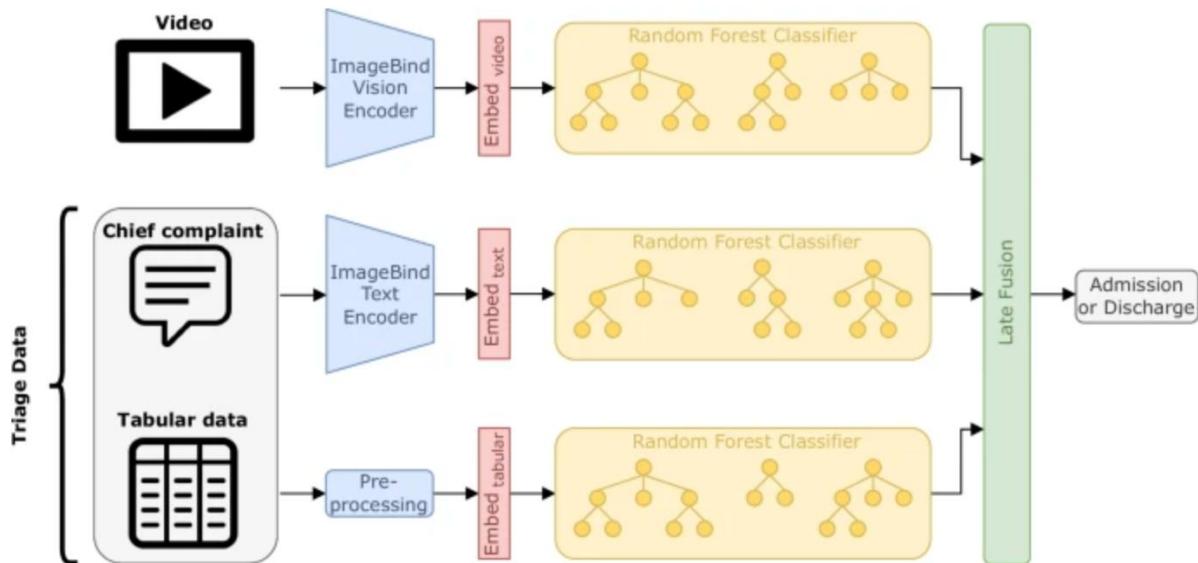
Fig. 1: Patient video recording and processing.



- This study investigates the use of computer vision AI models to predict hospital admissions from emergency department (ED) using short video clips of patients.
- The hypothesis of the paper is that computer vision AI can capture and extract meaningful clinical cues from patient appearances and movements similar to human clinicians' eye-balling assessment.

Hospitalization prediction from the emergency department using computer vision AI with short patient video clips

Fig. 2: Schematic representation of the predictive model.



- The results of the paper demonstrate that the AI model using video modality alone outperforms models using triage clinical data. The study further demonstrates that by combining video data and triage data, the multimodal AI model achieves the best performance.
- This work points out that video AI has the potential to enhance ED triage and contribute to more efficient healthcare resource allocation



Article | [Open access](#) | Published: 21 June 2024

Validation and application of computer vision algorithms for video-based tremor analysis

[Maximilian U. Friedrich](#) , [Anna-Julia Roenn](#), [Chiara Palmisano](#), [Jane Alty](#), [Steffen Paschen](#), [Guenther Deuschl](#), [Chi Wang Ip](#), [Jens Volkmann](#), [Muthuraman Muthuraman](#), [Robert Peach](#) & [Martin M. Reich](#) 

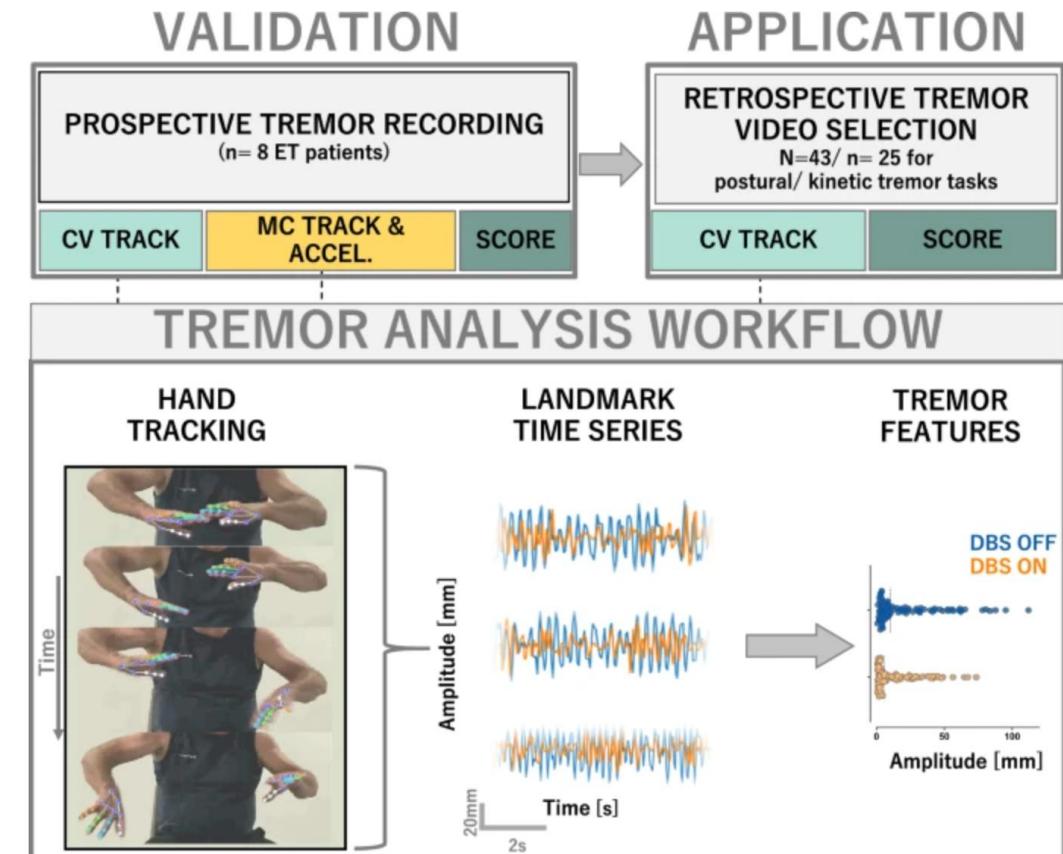
[npj Digital Medicine](#) 7, Article number: 165 (2024) | [Cite this article](#)

3492 Accesses | 23 Altmetric | [Metrics](#)

Validation and application of computer vision algorithms for video-based tremor analysis

- This study validates and applies computer-vision-based pose tracking for tremor analysis using clinical video recordings
- The paper demonstrates that open-source algorithms, such as Mediapipe, can effectively and accurately quantify tremor amplitude and frequency

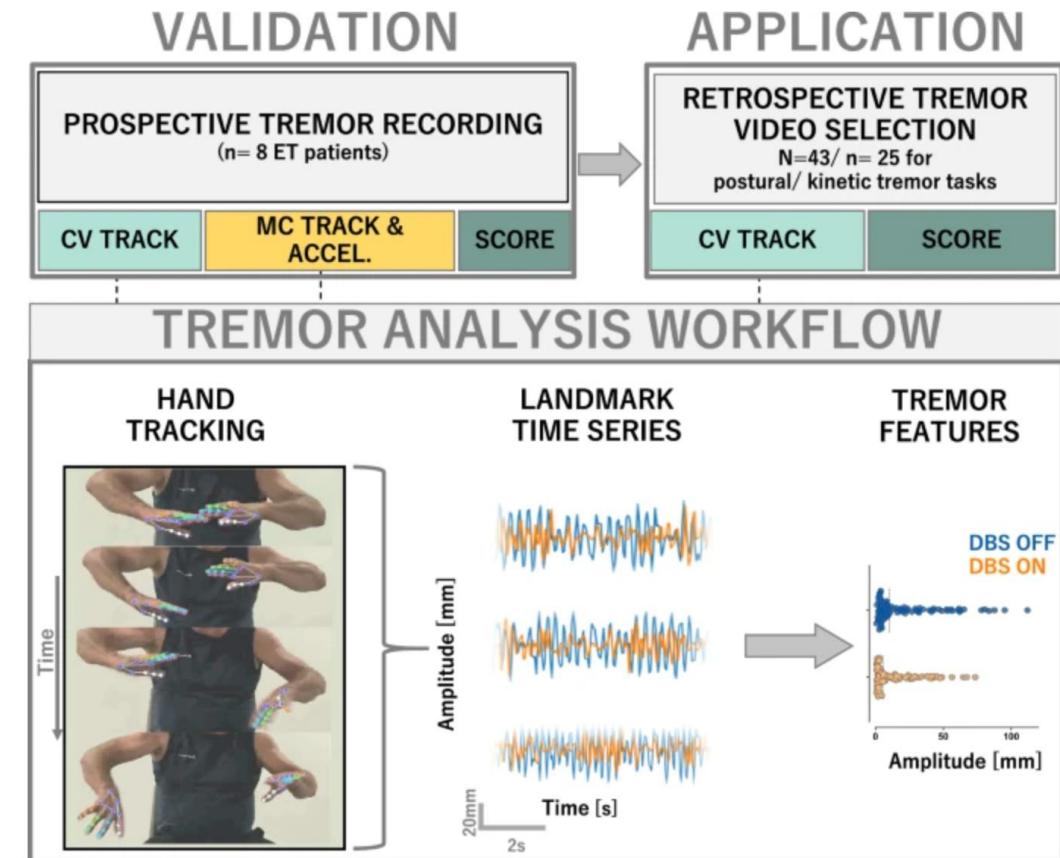
Fig. 7: Workflow.



Validation and application of computer vision algorithms for video-based tremor analysis

- Abilities match the gold-standard motion capture and tracking methods including Compared performance of Mediapipe to gold-standard methods including:
 - marker-based 3D motion capture
 - wrist-worn accelerometry
 - clinical rating scales
 - disease-specific pose-tracking algorithm called DLC-RCNN
- The results indicates that video-based AI tools can be used to achieve scalable and low-cost tremor assessment.

Fig. 7: Workflow.



eClinicalMedicine

Part of THE LANCET *Discovery Science*



This journal Journals Publish Clinical Global health Multimedia Events About

ARTICLES · Volume 78, 102940, December 2024 · Open Access

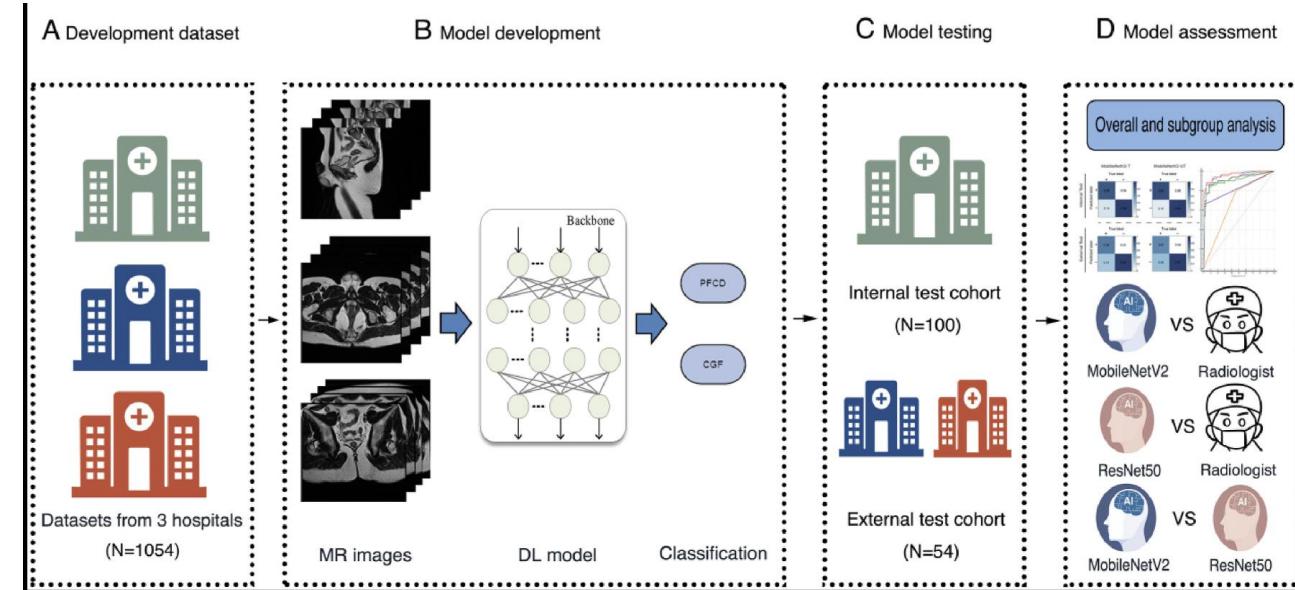
 Download Full Issue

Development and validation of the MRI-based deep learning classifier for distinguishing perianal fistulizing Crohn's disease from cryptoglandular fistula: a multicenter cohort study

Heng Zhang^{a,h,i,j} · Wenru Li^{c,h,i,j} · Tao Chen^{b,j} · Ke Deng^b · Bolin Yang^d · Jingen Luo^e · et al. [Show more](#)

Development and validation of the MRI-based deep learning classifier for distinguishing perianal fistulizing Crohn's disease from cryptoglandular fistula: a multicenter cohort study

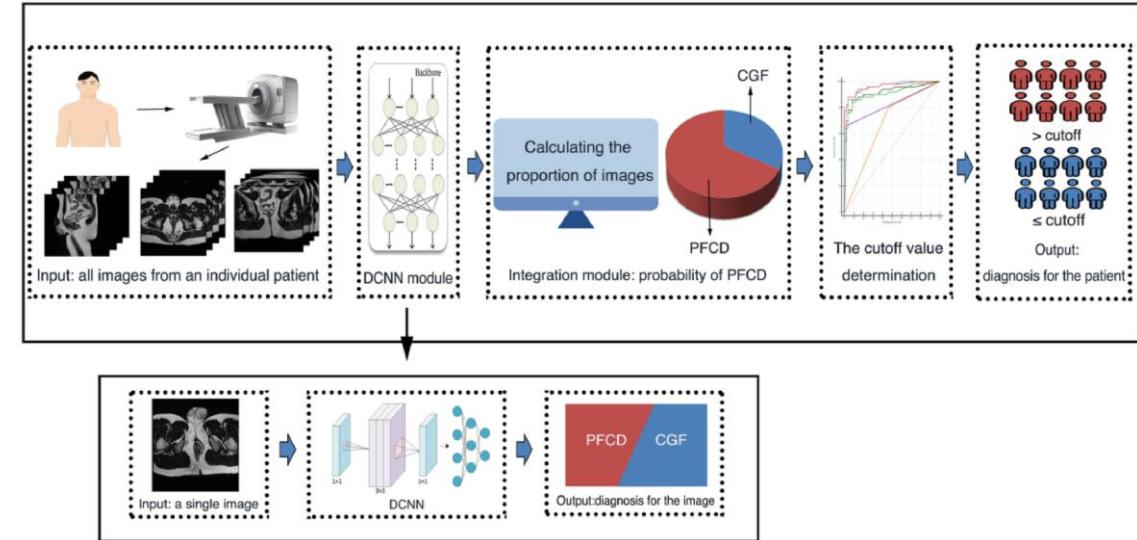
- This paper tackles a major challenge in diagnosing perianal fistulizing Crohn's disease (PFCD) from cryptoglandular fistula (CGF)
- Two conditions that look similar on imaging but requiring very different treatments
- Since misdiagnosis of PFCD and CGF can lead to delay of proper treatment, the study develops a novel deep learning classification framework based on MRI-based deep convolutional neural networks (DCNNs) to distinguish PFCD from CGF



Development and validation of the MRI-based deep learning classifier for distinguishing perianal fistulizing Crohn's disease from cryptoglandular fistula: a multicenter cohort study



- The authors of the paper trained two DCNNs:
 - MobileNetV2
 - ResNet50
- 44,871 MRI images from 1,054 patients across three hospitals to differentiate PFCD from CGF
- Compared performance of the DCNN framework to both senior and junior radiologists on the test datasets using accuracy, sensitivity, specificity, and AUC



- DCNN framework outperformed human radiologists and demonstrates the consistency and reliability of AI in diagnosis

eClinicalMedicine

Part of THE LANCET Discovery Science



This journal

Journals

Publish

Clinical

Global health

Multimedia

Events

About

ARTICLES · Volume 78, 102919, December 2024 · Open Access

 Download Full Issue

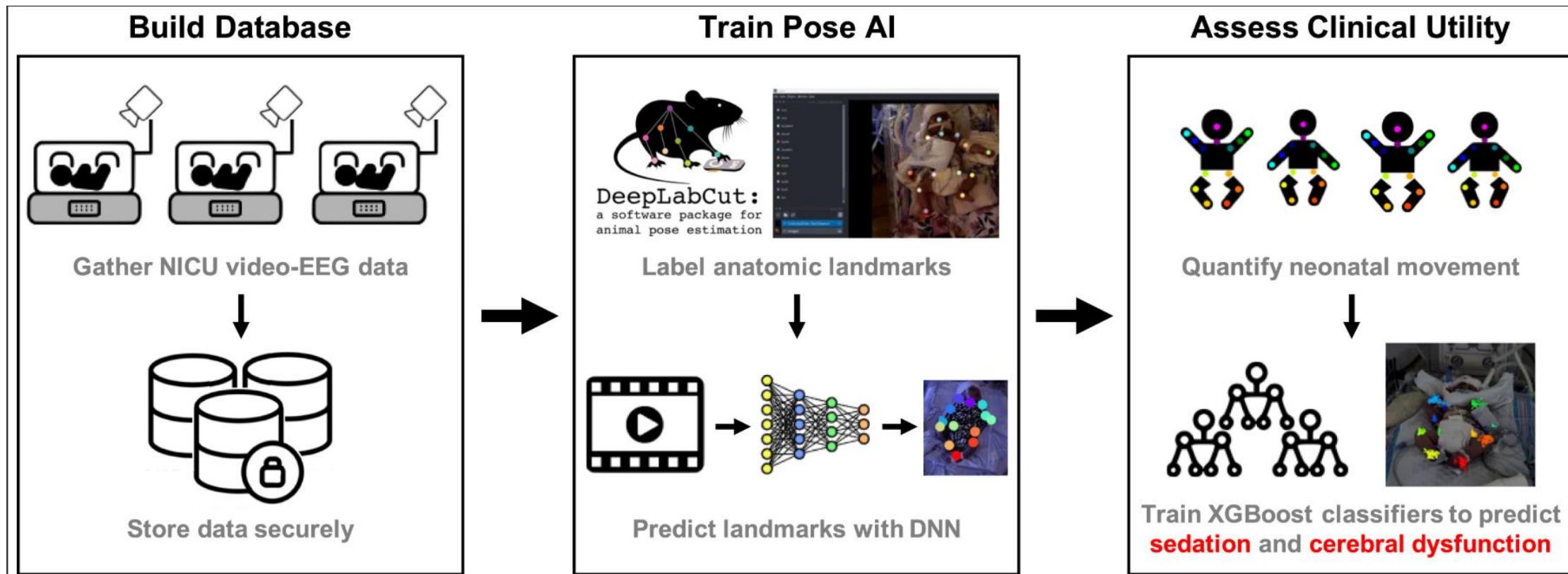
Detection of neurologic changes in critically ill infants using deep learning on video data: a retrospective single center cohort study

Alec Gleason^a · Florian Richter^b · Nathalia Beller^c · Naveen Arivazhagan^{d,e} · Rui Feng^f · Emma Holmes^{e,g} ·

Benjamin S. Glicksberg^c · Sarah U. Morton^{h,i} · Maite La Vega-Talbott^{j,k} · Madeline Fields^j · Katherine Guttmann^{g,l} ·

Girish N. Nadkarni^{d,e} · Felix Richter^{k,l}   Show less

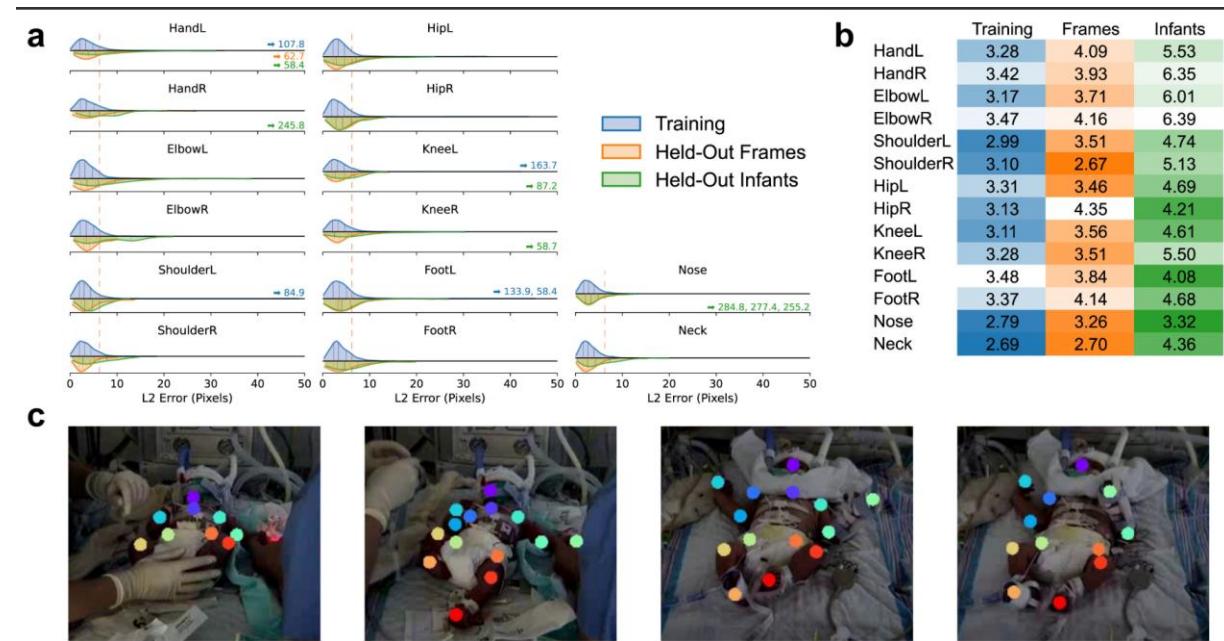
Detection of neurologic changes in critically ill infants using deep learning on video data: a retrospective single center cohort study



- Novel approach to detect neurologic changes in critical ill infants using deep learning approach; making the neonatal neuro-monitoring more accessible

Detection of neurologic changes in critically ill infants using deep learning on video data: a retrospective single center cohort study

- 282,000+ minutes of video from 115 infants
- Developed a deep-learning-based pose estimation system using DeepLabCut to recognize 14 anatomical landmarks in infants, which enabled automatic movement analysis from clinical video recordings
- Trained an XGBoost classifier to predict cerebral dysfunction
- Deep learning with pose AI may offer a scalable, minimally invasive method for neuro-telemetry in the NICU



[nature](#) > [nature biomedical engineering](#) > [articles](#) > [article](#)

Article | Published: 01 October 2024

Accurate prediction of disease-risk factors from volumetric medical scans by a deep vision model pre-trained with 2D scans

[Oren Avram](#) , [Berkin Durmus](#), [Nadav Rakocz](#), [Giulia Corradetti](#), [Ulzee An](#), [Muneeswar G. Nittala](#), [Prerit Terway](#), [Akos Rudas](#), [Zeyuan Johnson Chen](#), [Yu Wakatsuki](#), [Kazutaka Hirabayashi](#), [Swetha Velaga](#), [Liran Tiosano](#), [Federico Corvi](#), [Aditya Verma](#), [Ayesha Karamat](#), [Sophiana Lindenberg](#), [Deniz Oncel](#), [Louay Almidani](#), [Victoria Hull](#), [Sohaib Fasih-Ahmad](#), [Houri Esmaeilkhanian](#), [Maxime Cannesson](#), [Charles C. Wykoff](#), ... [Eran Halperin](#)  + Show authors

[Nature Biomedical Engineering](#) (2024) | [Cite this article](#)



Accurate prediction of disease-risk factors from volumetric medical scans by a deep vision model pre-trained with 2D scans

- Via transfer learning from 2D medical images, this model addresses a major bottleneck challenge in medical AI, which is the lack of large, annotated 3D datasets
- Model was pretrained on abundant 2D scans to improve predictions on 3D medical data
- Model has high generalizability over CT, MRI, ultrasound, and OCT modalities and shows superior performance over domain-specific models
- Model allowed reduction of annotation time by a factor of 5000 compared to manual labeling

Labeling is (cost- and) expert-time-prohibitive
<1 retina specialist per 100,000 individuals

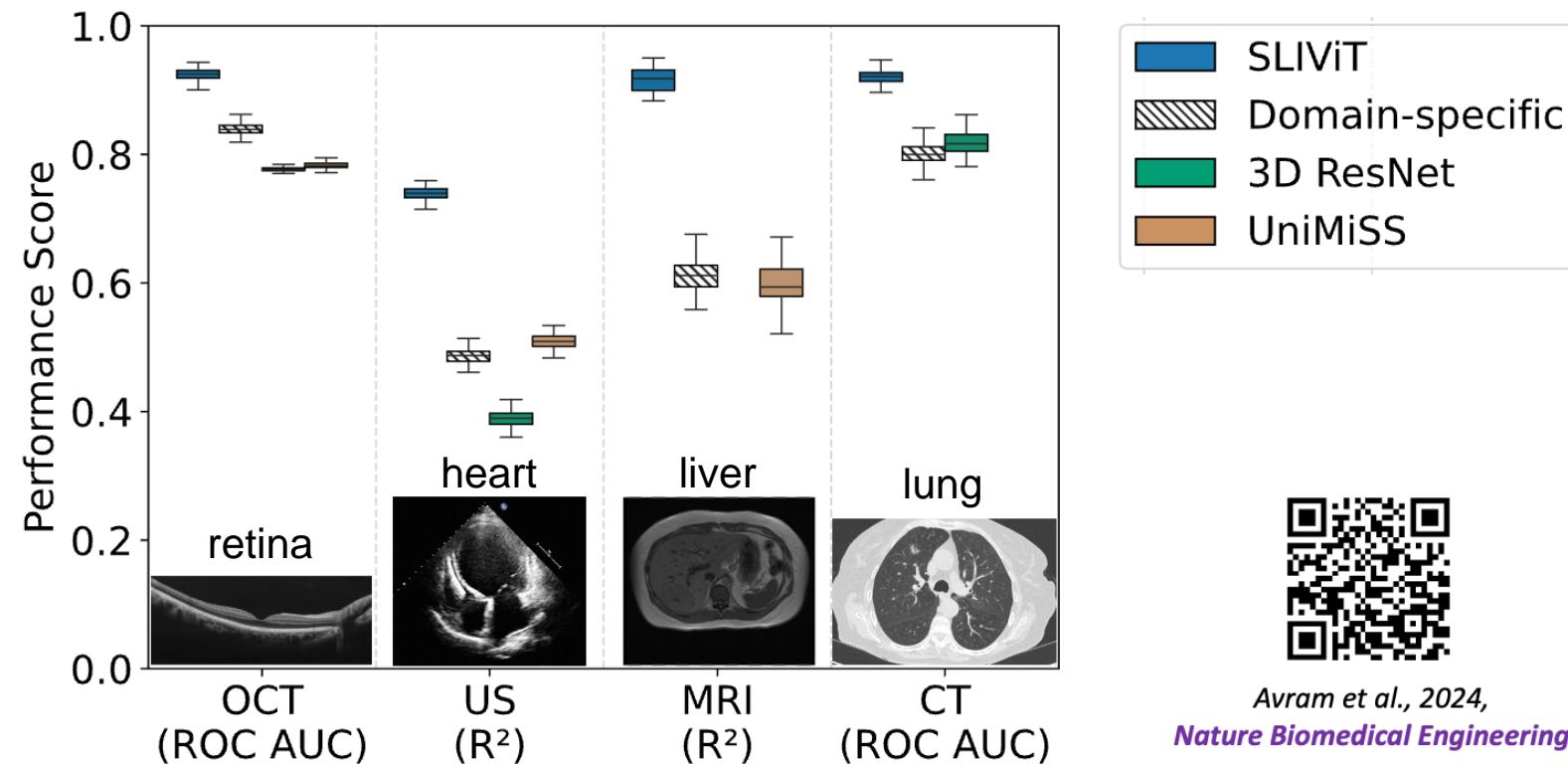


Accurate prediction of disease-risk factors from volumetric medical scans by a deep vision model pre-trained with 2D scans



SLIViT outperforms SOTA methods in different domains

SLIViT can handle multiple data modalities (OCT, US, MRI, CT, more?)
and significantly outperforms domain-specific SOTA methods



Project's GitHub



Avram et al., 2024,
Nature Biomedical Engineering

[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | Published: 15 August 2024

Brain aging patterns in a large and diverse cohort of 49,482 individuals

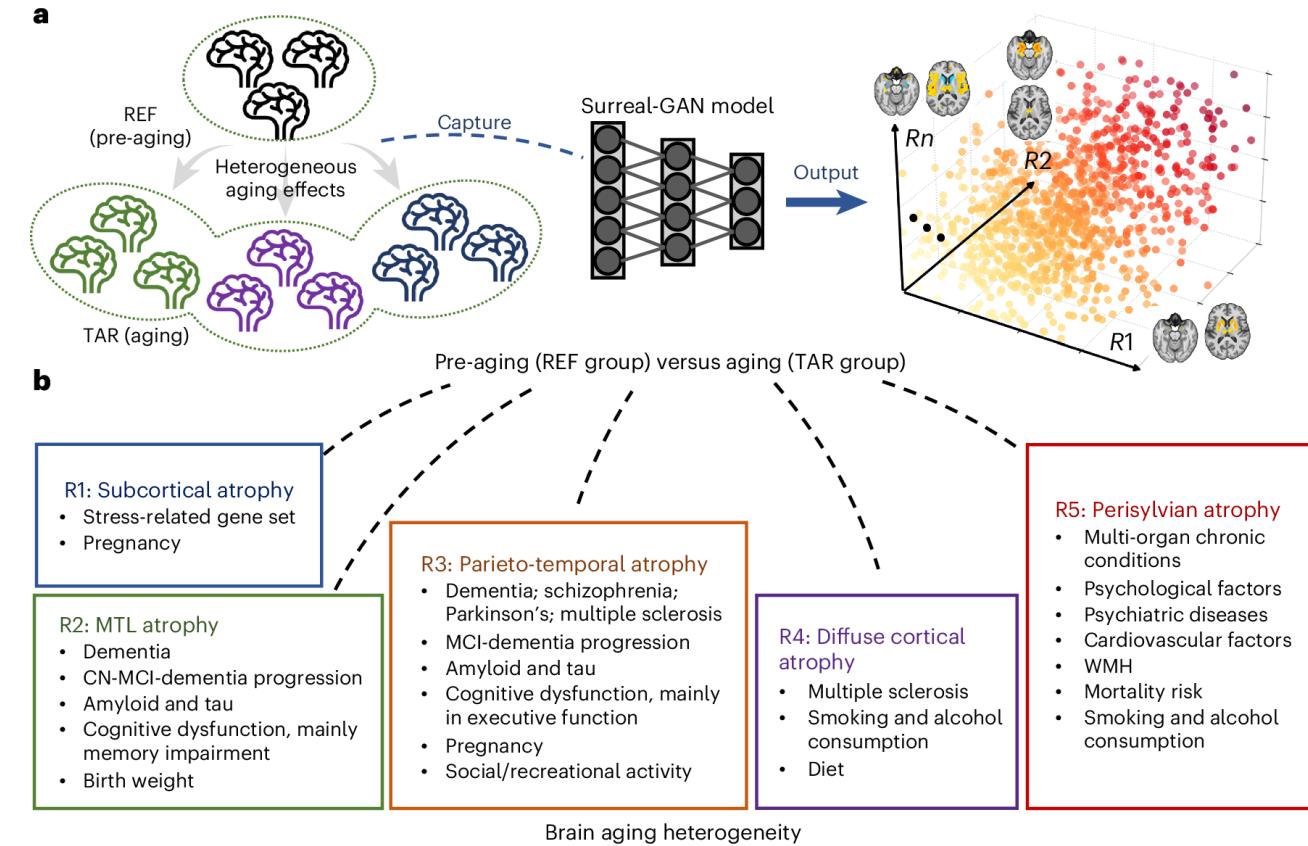


[Zhijian Yang](#), [Junhao Wen](#), [Guray Erus](#), [Sindhuja T. Govindarajan](#), [Randa Melhem](#), [Elizabeth Mamourian](#),
[Yuhan Cui](#), [Dhivya Srinivasan](#), [Ahmed Abdulkadir](#), [Paraskevi Parmpi](#), [Katharina Wittfeld](#), [Hans J. Grabe](#),
[Robin Bülow](#), [Stefan Frenzel](#), [Duygu Tosun](#), [Murat Bilgel](#), [Yang An](#), [Dahyun Yi](#), [Daniel S. Marcus](#), [Pamela](#)
[LaMontagne](#), [Tammie L. S. Benzinger](#), [Susan R. Heckbert](#), [Thomas R. Austin](#), [Shari R. Waldstein](#),
... [Christos Davatzikos](#)  + Show authors

[Nature Medicine](#) 30, 3015–3026 (2024) | [Cite this article](#)

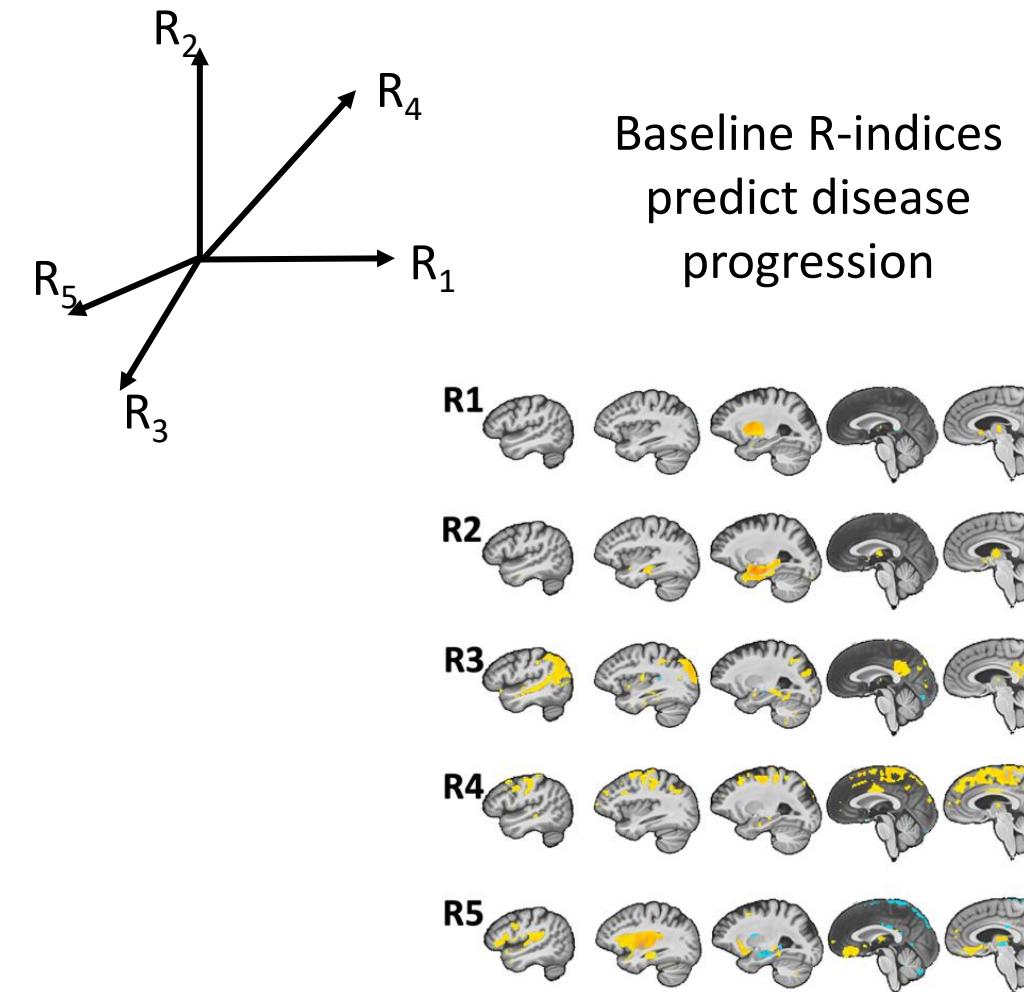
Brain aging patterns in a large and diverse cohort of 49,482 individuals

- Used state-of-the-art deep-representation learning method, Surreal-GAN
- Elucidating brain aging heterogeneity in a cohort of 49,482 individuals from 11 studies
- Five dominant patterns of brain atrophy were identified and quantified for each individual using respective measures, R-indices
- Their associations with biomedical, lifestyle and genetic factors provide insights into the etiology of observed variances, suggesting their potential as brain endophenotypes for genetic and lifestyle risks



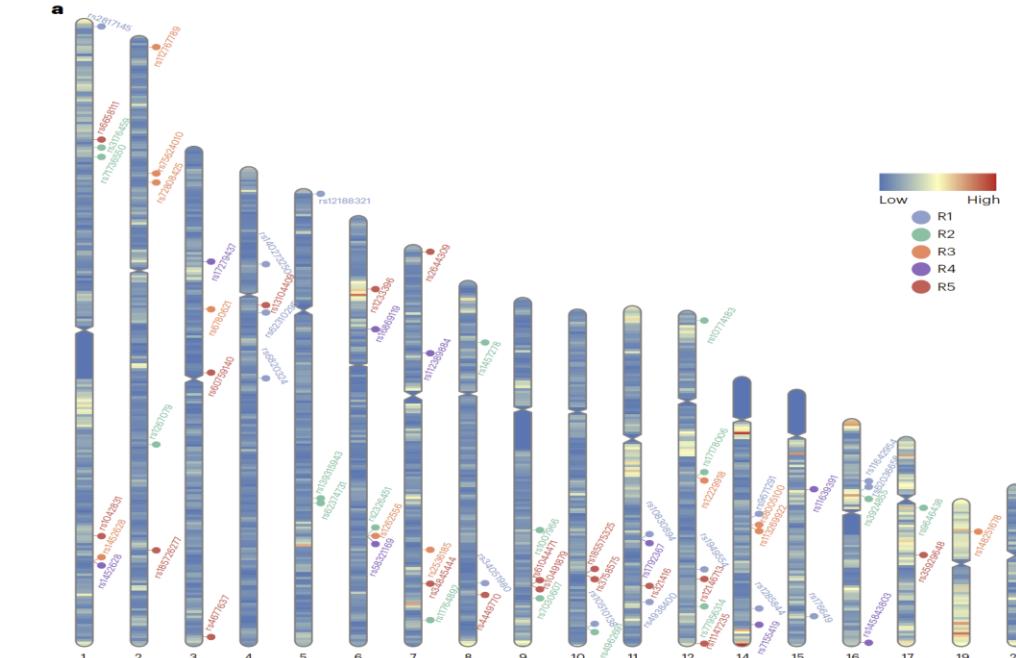
Brain aging patterns in a large and diverse cohort of 49,482 individuals

- Used state-of-the-art deep-representation learning method, Surreal-GAN
- Elucidating brain aging heterogeneity in a cohort of 49,482 individuals from 11 studies
- Five dominant patterns of brain atrophy were identified and quantified for each individual using respective measures, R-indices
- Their associations with biomedical, lifestyle and genetic factors provide insights into the etiology of observed variances, suggesting their potential as brain endophenotypes for genetic and lifestyle risks



Brain aging patterns in a large and diverse cohort of 49,482 individuals

- R-indices show distinct underlying genetic correlates – 73 genomic loci
- The dimensional system illuminates the intricate relationships between pathological factors and variations in brain aging
- R3 and R5 are associated with multiple sclerosis, schizophrenia and Parkinson's disease
- R5 exhibits broad associations with various systemic diseases, including neuropsychiatric disorders, cardiovascular diseases and immune health factors, and shows the strongest associations with WMH volumes
- R2, R3 and R5 are strongly correlated with MCI/dementia



eClinicalMedicine

Part of THE LANCET Discovery Science



This journal

Journals

Publish

Clinical

Global health

Multimedia

Events

About

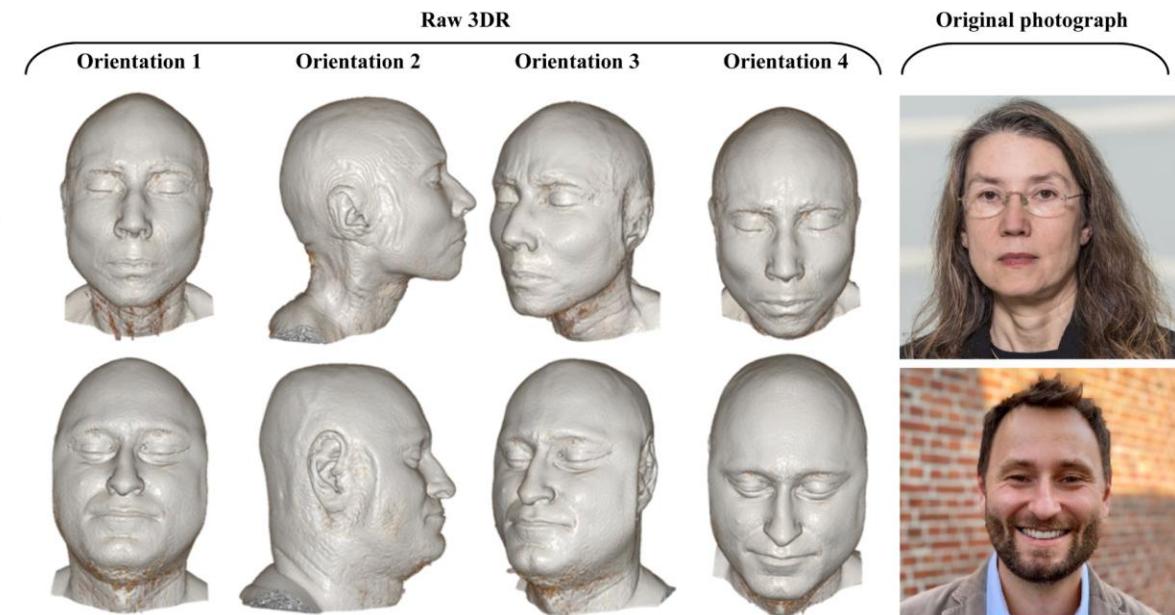
ARTICLES · Volume 78, 102930, December 2024 · Open Access

 Download Full Issue

Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy

Katharina Steeg  ^a  · Evelyn Bohrer ^a · Stefan Benjamin Schäfer ^a · Viet Duc Vu ^a · Jan Scherberich ^a · Anton George Windfelder ^{a,b,c} · et al. Show more

Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy

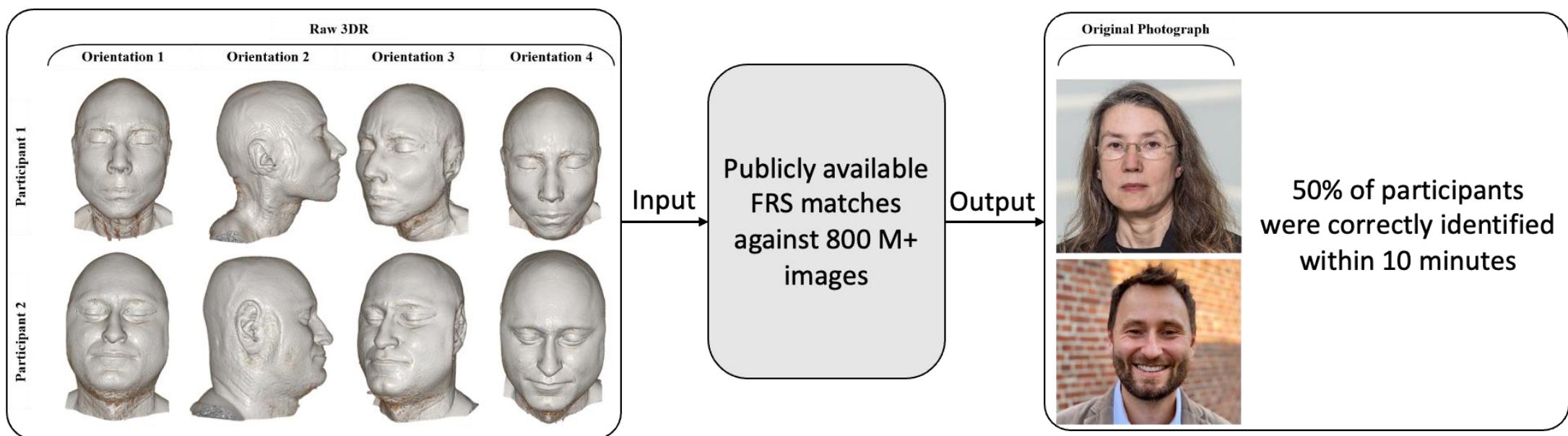


*Participants 1 and 2 provided explicit consent for the publication of their face images

- This paper points out a major privacy risk in medical imaging by demonstrating that publicly available facial recognition software (FRS) can re-identify individuals from MRI head scans
- Any person with minimum professional knowledge in computation can use free, easy-to-access software to match 3D facial reconstructions from MRI to online large public image databases

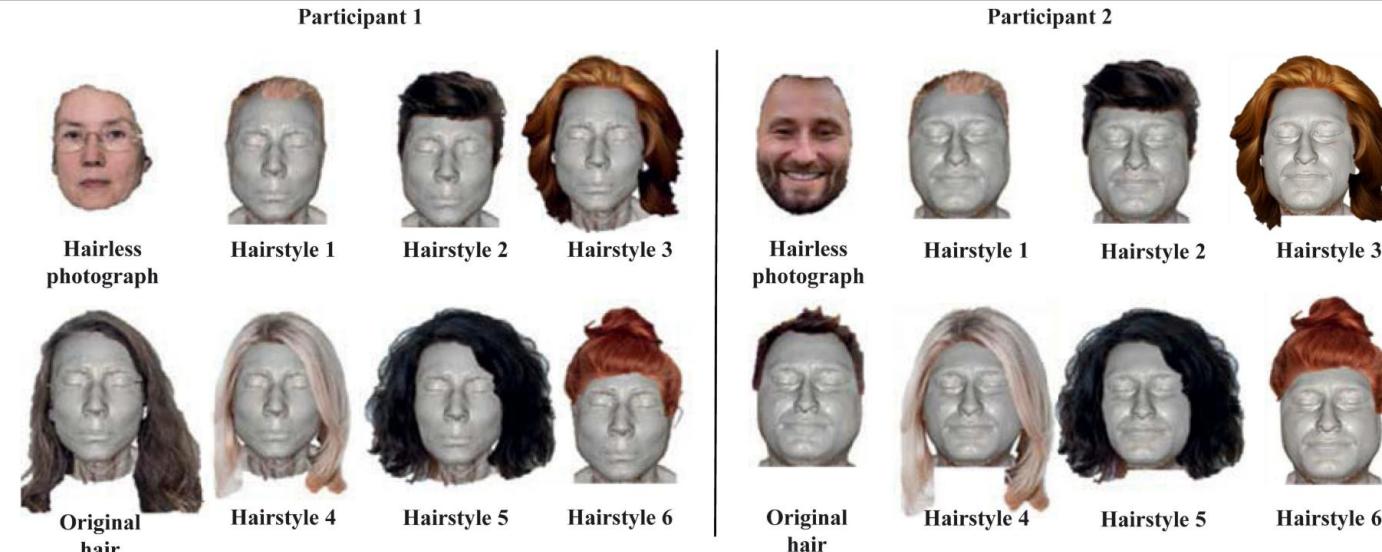
Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy

- Advances in AI and free 3D rendering software enable face reconstruction (3DR) of anonymised MRI head scans easily.
- Modern facial recognition software (FRS) is widely available and can match faces against massive online databases.
- Re-identification of participants from anonymised MRI head images is possible within minutes.



Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy. Steeg, Katharina et al. eClinicalMedicine, Volume 78, 102930

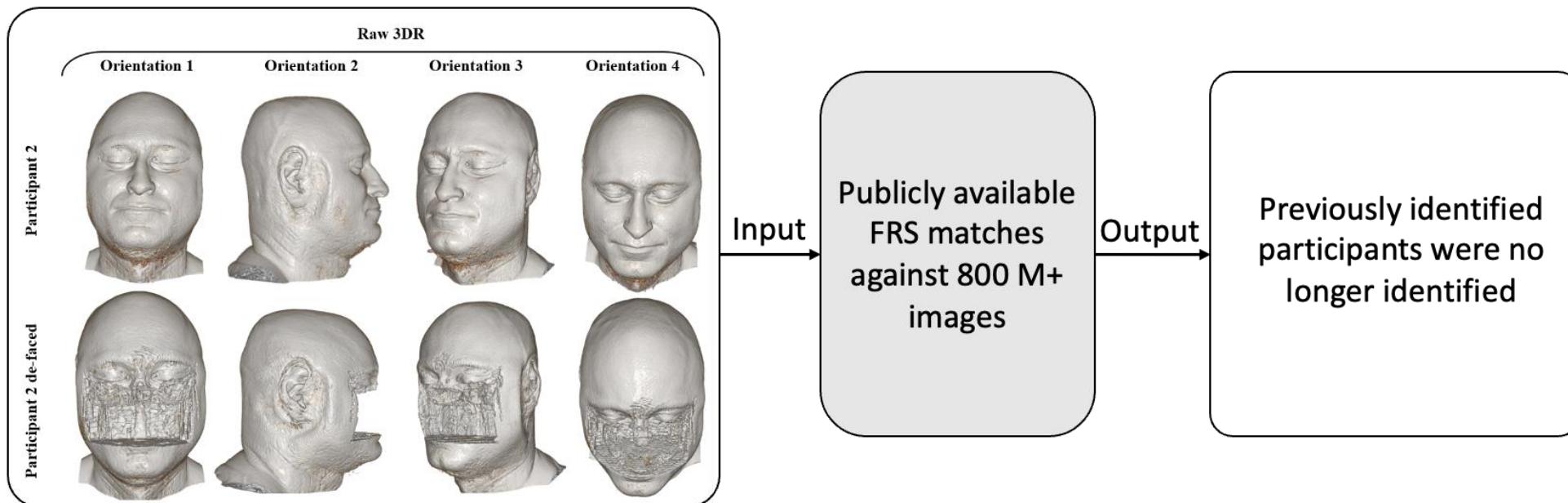
Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy



- Slightly modifying the 3D reconstructions, such as adjusting head orientation or adding hair, can *improve* the accuracy of identifying participants
- To tackle the re-identification issue, the study also tested an anonymization tool called `mri_deface` which obscures facial details in MRI images
- The study successfully prevented re-identification and showed that by de-facing the MRI head scans, patient privacy can be protected

Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy

- Re-identification could expose personal and medical information, increasing privacy risks.
- ***De-facing*** can prevent re-identification, but lacks global standard regulations.
- Standardised regulations for de-facing are needed to mitigate medical data misuse.



Re-identification of anonymised MRI head images with publicly available software: investigation of the current risk to patient privacy. Steeg, Katharina et al. eClinicalMedicine, Volume 78, 102930

Clinical Decision Support, Evaluation and Implementation



Original Investigation | Health Informatics

October 28, 2024

Large Language Model Influence on Diagnostic Reasoning

A Randomized Clinical Trial



Ethan Goh, MBBS, MS^{1,2}; Robert Gallo, MD³; Jason Hom, MD⁴; Eric Strong, MD⁴; Yingjie Weng, MHS⁵; Hannah Kerman, MD^{6,7}; Joséphine A. Cool, MD^{6,7}; Zahir Kanjee, MD, MPH^{6,7}; Andrew S. Parsons, MD, MPH⁸; Neera Ahuja, MD⁴; Eric Horvitz, MD, PhD^{9,10}; Daniel Yang, MD¹¹; Arnold Milstein, MD²; Andrew P. J. Olson, MD¹²; Adam Rodman, MD, MPH^{6,7}; Jonathan H. Chen, MD, PhD^{1,2,13}

[» Author Affiliations](#) | [Article Information](#)

JAMA Netw Open. 2024;7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969

Augmented Diagnostic Reasoning?



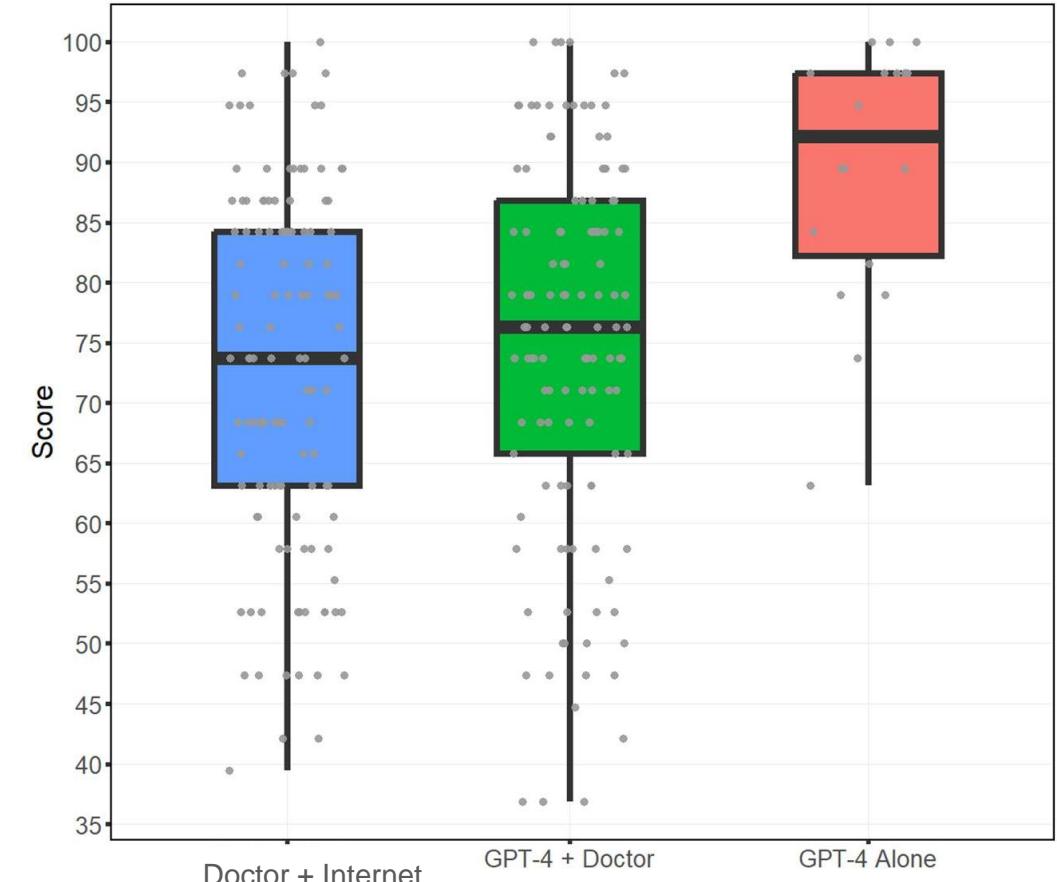
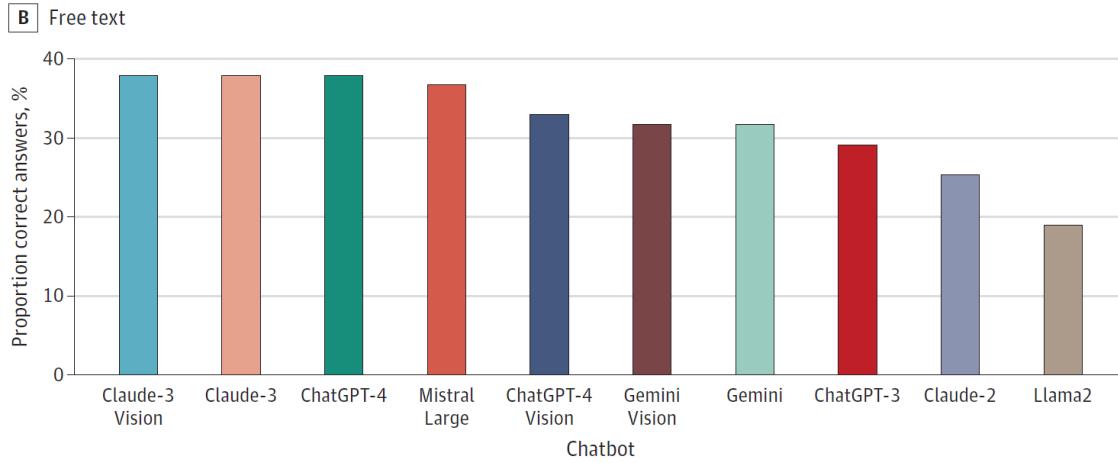
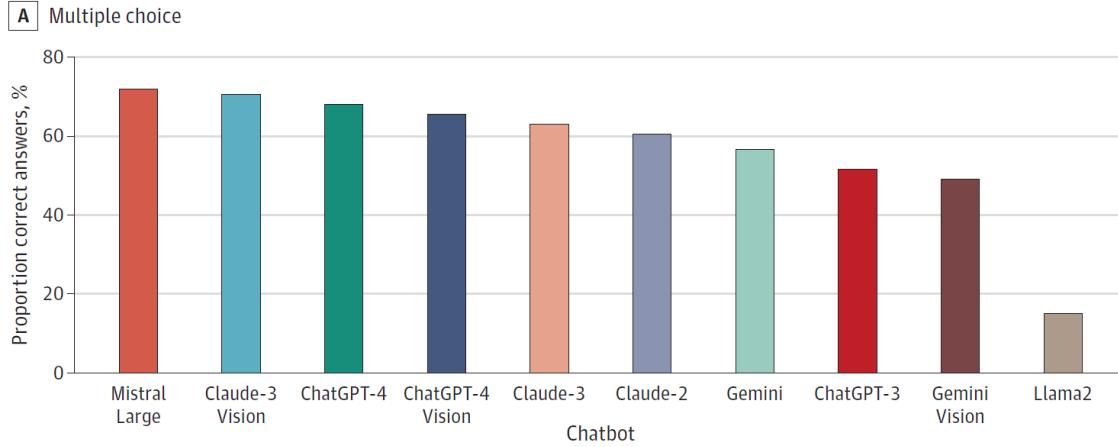
- Clinical Vignettes (6 Cases), reviewed >250 times
- 50 licensed physicians
- Consensus Grade on Reasoning

Table. Medical Accuracy of Chatbots Evaluated on Multiple-Choice and Free-Text Examinations of Clinical Oncology Cases, Stratified by Management and Diagnosis-Related Questions

Chatbot	Multiple choice				P value ^a	Free text			
	No./No. (%)	Management	Diagnosis	Total		No./No. (%)	Management	Diagnosis	Total
Claude-3 Vision	31/42 (73.81)	25/37 (67.57)	56/79 (70.89)	.15	.15	15/42 (35.71)	15/37 (40.54)	30/79 (37.97)	>.99
ChatGPT-4 Vision	23/42 (54.76)	29/37 (78.38)	52/79 (65.82)	.16	.16	14/42 (33.33)	12/37 (32.43)	26/79 (32.91)	.53
Gemini Vision	22/42 (52.38)	17/37 (45.95)	39/79 (49.37)	.17	.17	15/42 (35.71)	10/37 (27.03)	25/79 (31.65)	.09
Mistral Large	32/42 (76.19)	25/37 (67.57)	57/79 (72.15)	.09	.09	14/42 (33.33)	15/37 (40.54)	29/79 (36.71)	.77
ChatGPT-4	24/42 (57.14)	30/37 (81.08)	54/79 (68.35)	.16	.16	17/42 (40.48)	13/37 (35.14)	30/79 (37.97)	.22
ChatGPT-3	24/42 (57.14)	17/37 (45.95)	41/79 (51.9)	.06	.06	13/42 (30.95)	10/37 (27.03)	23/79 (29.11)	.31
Claude-3	29/42 (69.05)	21/37 (56.76)	50/79 (63.29)	.04	.04	18/42 (42.86)	12/37 (32.43)	30/79 (37.97)	.06
Claude-2	24/42 (57.14)	24/37 (64.86)	48/79 (60.76)	>.99	>.99	12/42 (28.57)	8/37 (21.62)	20/79 (25.32)	.14
Gemini	26/42 (61.9)	19/37 (51.35)	45/79 (56.96)	.07	.07	14/42 (33.33)	11/37 (29.73)	25/79 (31.65)	.33
Llama2	6/42 (14.29)	6/37 (16.22)	12/79 (15.19)	>.99	>.99	9/42 (21.43)	6/37 (16.22)	15/79 (18.99)	.20

^a P values calculated using χ^2 test.

Diagnostic Reasoning Case Example



Original Investigation

March 11, 2024



Clinical Decision Support for Hypertension Management in Chronic Kidney Disease

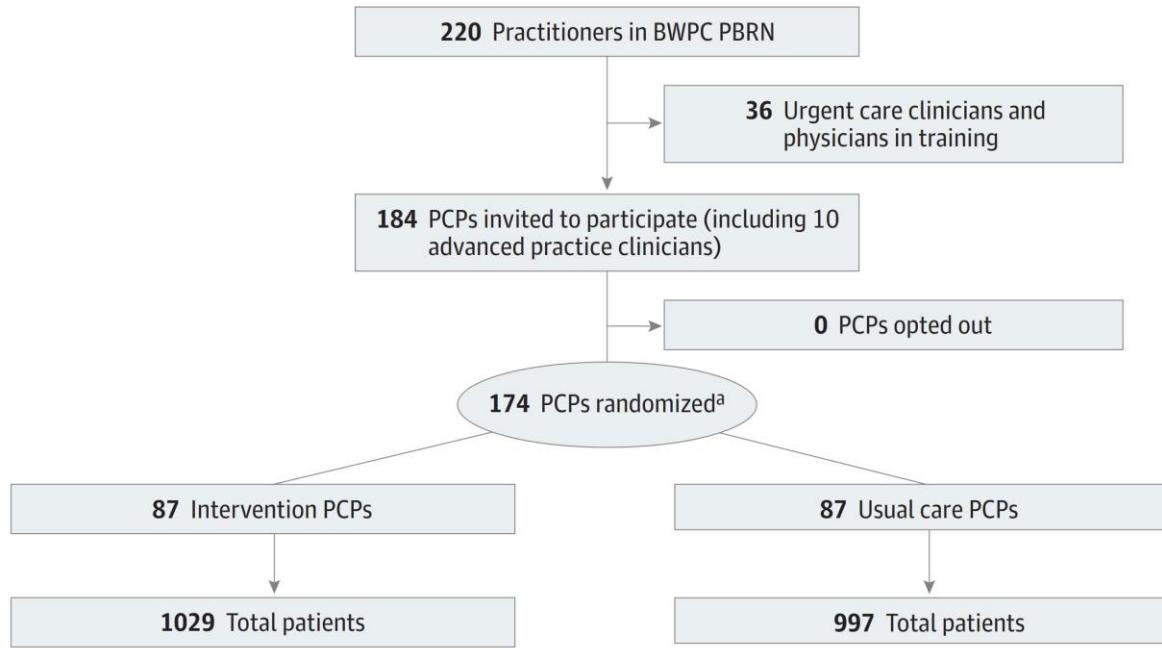
A Randomized Clinical Trial

Lipika Samal, MD, MPH^{1,3}; John L. Kilgallon, BA^{1,4}; Stuart Lipsitz, ScD^{1,3,5}; Heather J. Baer, ScD^{1,3,5}; Allison McCoy, PhD⁶; Michael Gannon, ScM^{1,7}; Sarah Noonan, BS^{1,8}; Ryan Dunk, BA¹; Sarah W. Chen, BS¹; Weng Ian Chay, BS¹; Richard Fay, MEd¹; Pamela M. Garabedian, MS⁹; Edward Wu, BS^{1,10}; Matthew Wien, BS¹; Saul Blecker, MD, MHS¹¹; Hojjat Salmasian, MD, MPH, PhD¹²; Joseph V. Bonventre, MD, PhD^{2,3}; Gearoid M. McMahon, MBBCH^{2,3}; David W. Bates, MD^{1,3}; Sushrut S. Waikar, MD, MPH¹³; Jeffrey A. Linder, MD, MPH¹⁴; Adam Wright, PhD⁶; Patricia Dykes, RN, PhD, MA^{1,3}

[» Author Affiliations](#) | [Article Information](#)

JAMA Intern Med. 2024;184(5):484-492. doi:10.1001/jamainternmed.2023.8315

Study design and example of Best Practice Advisory



Important (1)

⚠ Patient has CKD, 2 SBP's ≥ 140 mmHg, and is on an ACE, which could be increased

provide feedback: 😊😊😊

Why did this alert fire?

- Pt has CKD: 2 eGFR < 60 within the past 2 years, at least 90 days apart - Most recent eGFR: **40** (7/28/2020)
- Pt has had 2 elevated SBP's ≥ 140 mmHg - Today's SBP: **160 mmHg**
- Pt is on Lisinopril - Current Dose: **5 mg**
- Most recent K: 3.5 (7/28/2020)

Consider increasing Lisinopril dose. Consider ordering a BMP to monitor creatinine.

Order	Do Not Order	Lisinopril 10 MG Tablet
Order	Do Not Order	Basic Metabolic Panel in 1 week
Order	Do Not Order	Ambulatory BWH Renal E-Consult

Acknowledge Reason

Will repeat BP Will order different dose Will discuss with pt Will review chart Remind me next visit Other

Results

Table 2. SBP at Baseline and 180 Days, Change in Mean SBP from Baseline, and BP Control

Measurement variable	Intervention	Usual care	P value
Baseline SBP, mean (SD), mm Hg ^a	154.3 (14.2)	153.7 (14.4)	.54
SBP at 180 d, mean (SD), mm Hg ^a	139.5 (19.7)	142.1 (19.9)	.009
Change in SBP, % (95% CI), mm Hg ^b	-14.6 (-13.1 to -16.0)	-11.7 (-10.2 to -13.1)	.005
BP control, % (95% CI) ^c	50.4 (46.5 to 54.3)	47.1 (43.3 to 51.0)	.23

Abbreviations: BP, blood pressure; SBP, systolic blood pressure.

^a SBP at baseline is the observed mean SBP presented in Table 1. SBP at 180 days is a product of the model.

^b Mean change in SBP represents an adjusted estimate after controlling for baseline difference in sex, using a repeated measures model incorporating SBP measurements at baseline and 90 days, and after accounting for clustering by primary care practitioner matched pair, primary care practitioner, and patient.

^c By design, there were no patients with controlled BP at baseline. Percentage represents adjusted estimates after controlling for baseline difference in sex, using a repeated measures model incorporating BP measurements at 90 days, and after accounting for clustering by primary care practitioner matched pair, primary care practitioner, and patient. BP control was defined as BP less than 140/90 mm Hg.



Original Investigation

September 18, 2024

Tailoring Risk Prediction Models to Local Populations

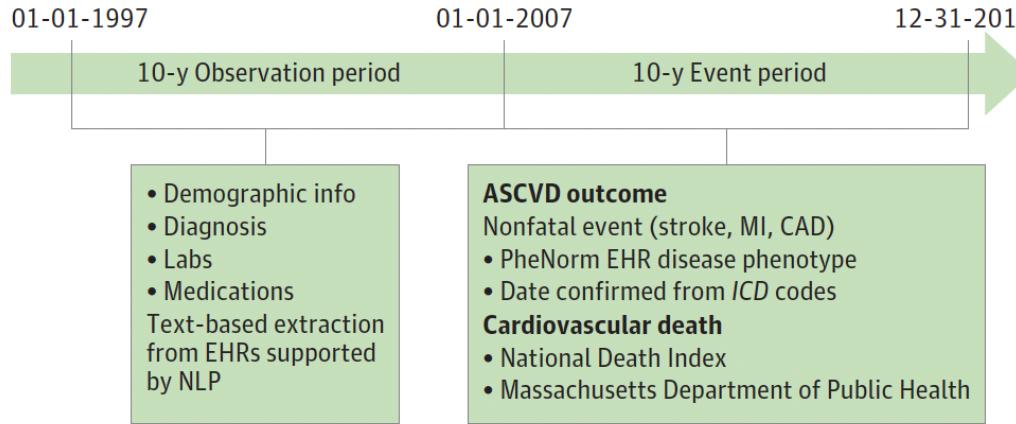
Aniket N. Zinzuwadia, MD¹; Olga Mineeva, MS²; Chunying Li, PhD¹; et al

[» Author Affiliations](#) | [Article Information](#)

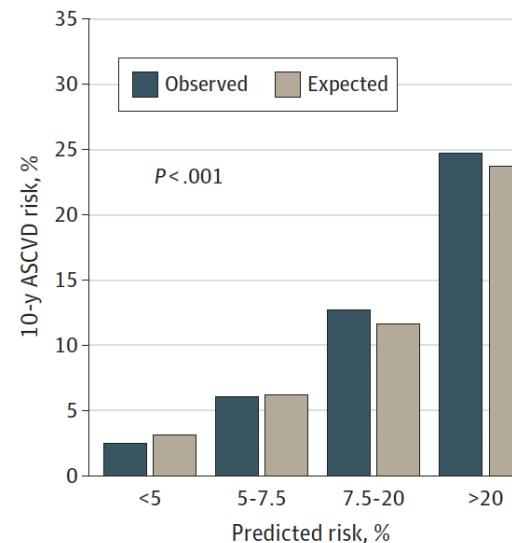
JAMA Cardiol. 2024;9(11):1018-1028. doi:10.1001/jamacardio.2024.2912

Validation of PREVENT model in EHR cohort from large Boston-area hospitals: room for improvement

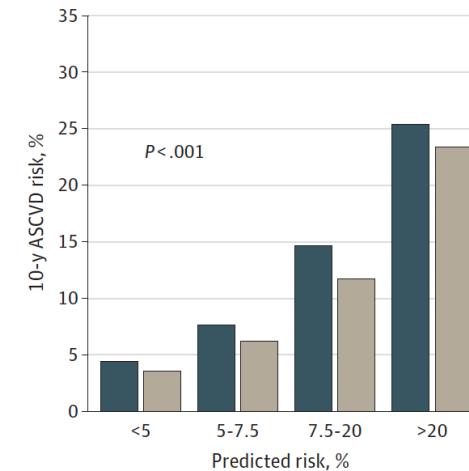
A Data collection



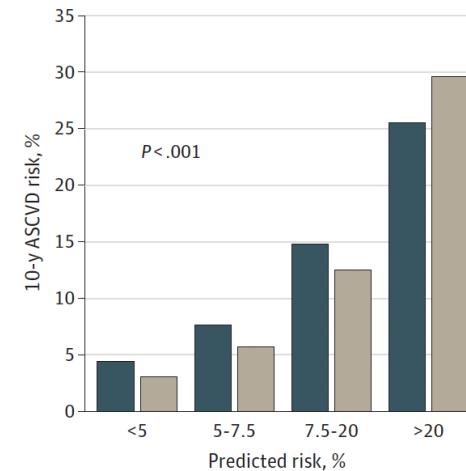
N=95K



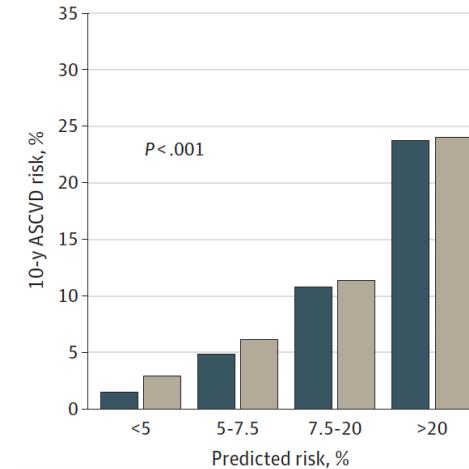
D Male sex: PREVENT



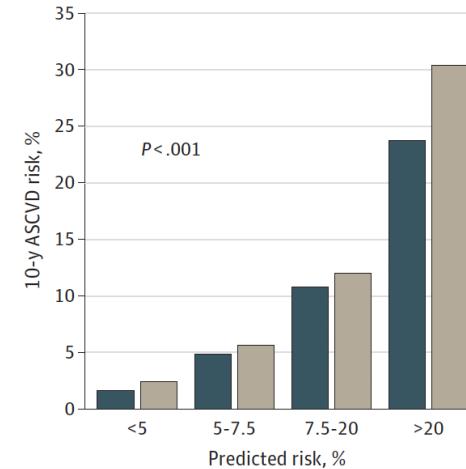
E Male sex: PREVENT-CS



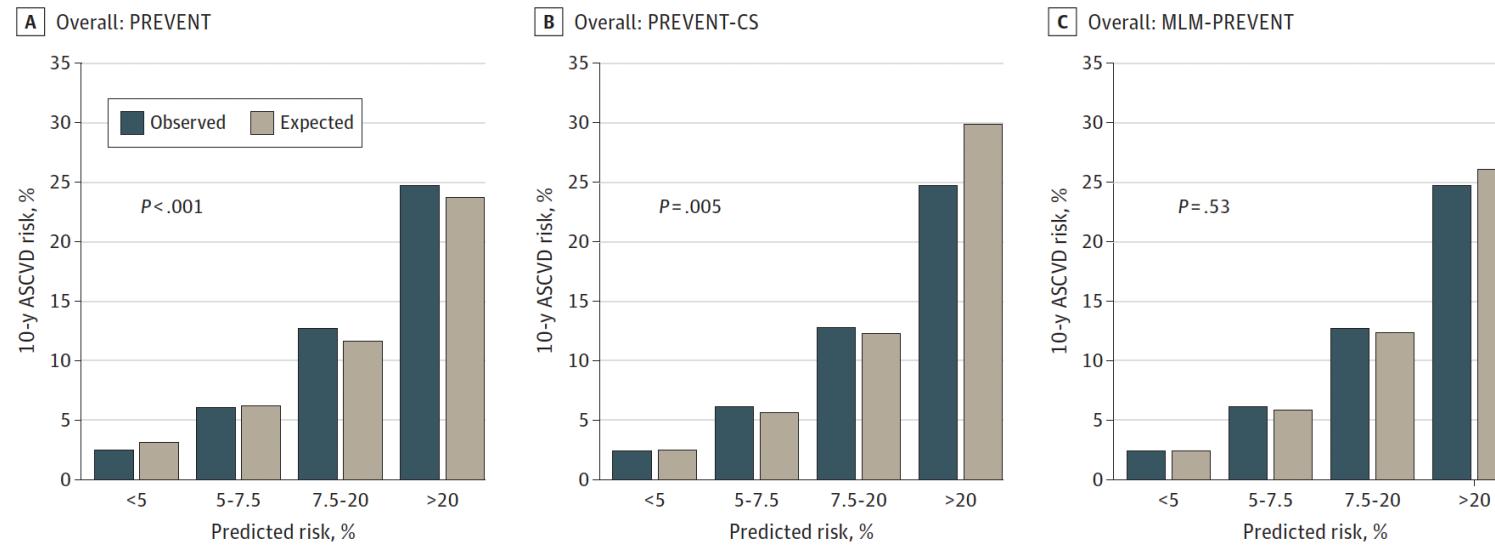
G Female sex: PREVENT



H Female sex: PREVENT-CS



Adapting PREVENT to local population with XGBoost



47 665 PREVENT recalibration
in training set

MLM-PREVENT
PREVENT, age, and sex were used
as MLM input variables

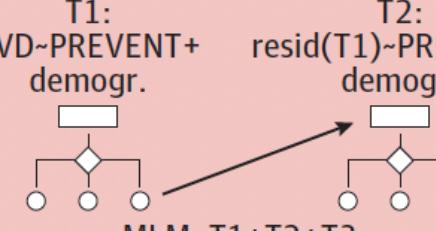
Monotonicity constraints
PREVENT and age

A validation set of 9524 patients was
used to terminate model training after a drop in incremental
improvement of additional boosting iterations

T1:
 $\text{CVD} \sim \text{PREVENT} + \text{demogr.}$

T2:
 $\text{resid}(\text{T1}) \sim \text{PREVENT} + \text{demogr.}$

MLM: $\text{T1} + \text{T2} + \text{T3} \dots$



Article | [Open access](#) | Published: 18 March 2024

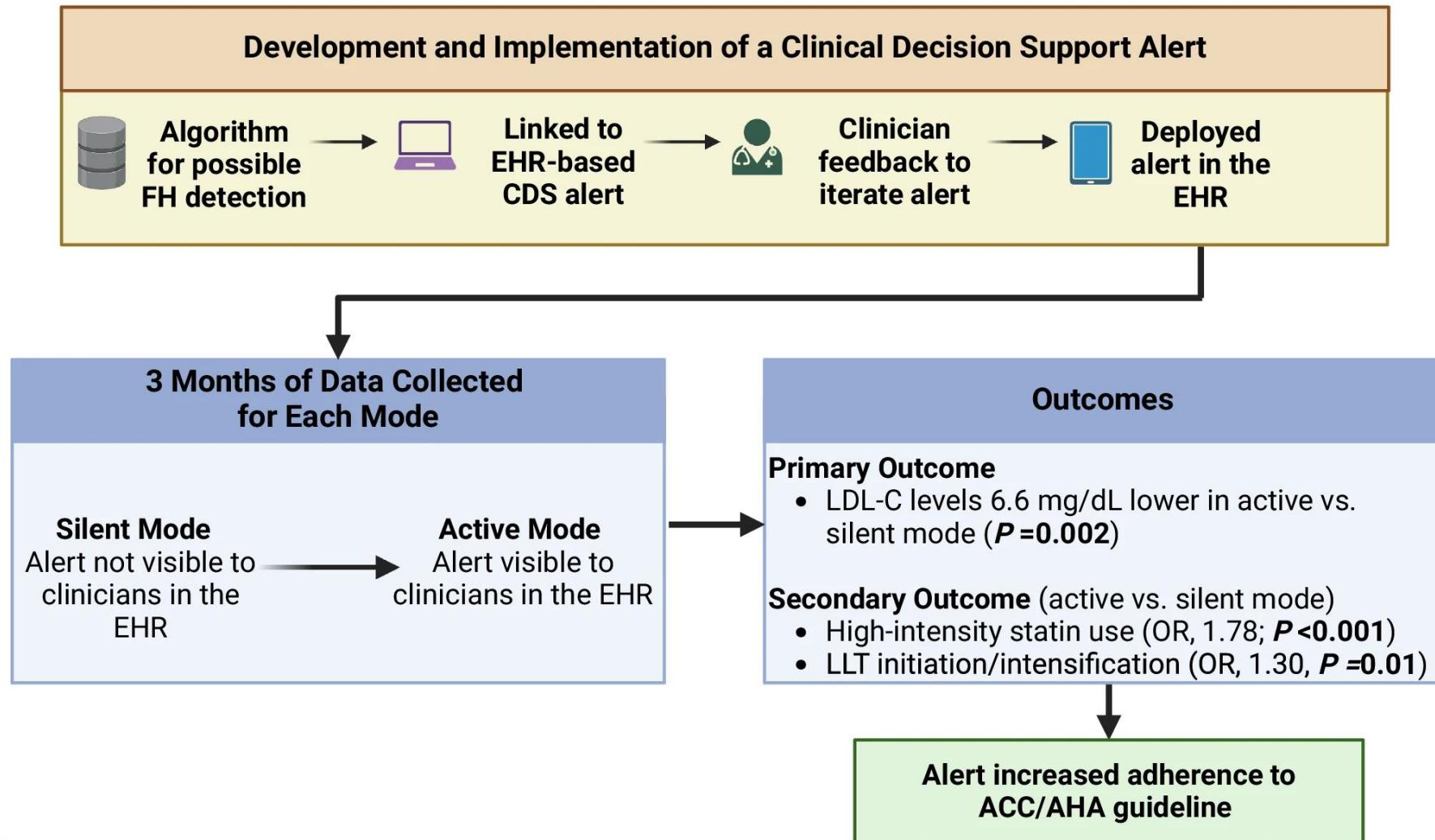


Effect of clinical decision support for severe hypercholesterolemia on low-density lipoprotein cholesterol levels

[Hana Bangash](#), [Seyedmohammad Saadatagah](#), [Mohammadreza Naderian](#), [Marwan E. Hamed](#), [Lubna Alhalabi](#), [Alborz Sherafati](#), [Joseph Sutton](#), [Omar Elsekaily](#), [Ali Mir](#), [Justin H. Gundelach](#), [Daniel Gibbons](#), [Paul Johnsen](#), [Christina M. Wood-Wentz](#), [Carin Y. Smith](#), [Pedro J. Caraballo](#), [Kent R. Bailey](#) & [Iftikhar J. Kullo](#) 

npj Digital Medicine 7, Article number: 73 (2024) | [Cite this article](#)

Study Design



Results



LDL-C levels were **6.6 mg/dL lower** (95% CI, −10.7 to −2.5; $p=0.002$) when the CDS was in active vs. silent mode.



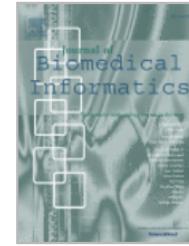
The odds of high-intensity statin use (OR, 1.78; 95% CI, 1.41–2.23; $p< 0.001$) and lipid lowering therapy initiation/intensification (OR, 1.30, 95% CI, 1.06–1.58, $p=0.01$) were higher in active vs. silent mode.

Implementation of CDS was associated with lowering of LDL-C levels in patients with severe hypercholesterolemia/possible familial hypercholesterolemia, likely due to higher rates of clinician led lipid lowering therapy initiation/intensification.



Journal of Biomedical Informatics

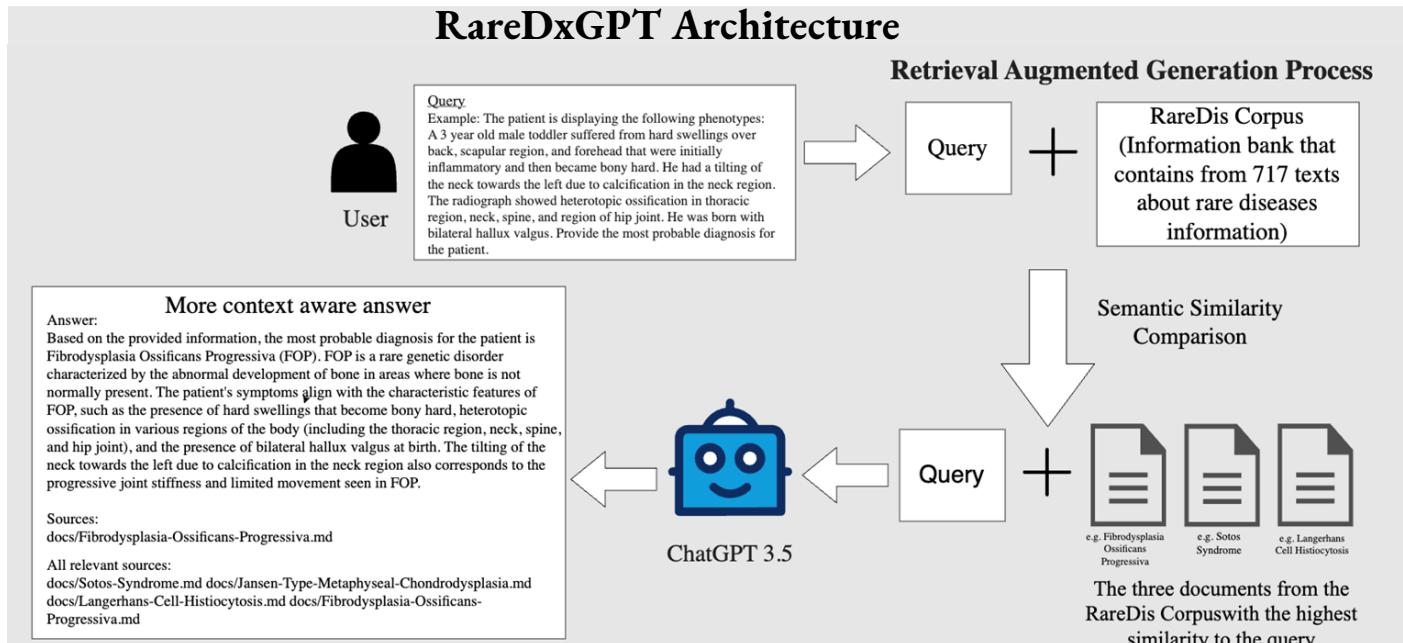
Volume 157, September 2024, 104702



Rare disease diagnosis using knowledge guided retrieval augmentation for ChatGPT

Charlotte Zelin ^a, Wendy K. Chung ^{b c}, Mederic Jeanne ^{b c}, Gongbo Zhang ^d,
Chunhua Weng ^d  

Study Design



Case Reports: PubMed Free-Text

Extracted case information of thirty different rare diseases

Responses evaluated with the aid of rare disease specialists on:

1. Accuracy
2. Global Quality Scale

Three Prompt Types Tested

Prompt: “The patient is displaying the following phenotypes: [case information]. Provide the most probable diagnosis for the patient.”

Prompt + Explanation: [Prompt] + “Explain what made you say this.”

Prompt + Roleplay: “You are a rare disease diagnostician. Your goal is to provide your patient with the most accurate diagnosis from the phenotypes described.” + [Prompt]

Results

- RareDx had a higher accuracy than ChatGPT 3.5 for each prompt
- Prompt + Explanation with RareDxGPT yielded the best results (43% accuracy)
- RareDxGPT was able to diagnosis disease with both short and long differential diagnosis (DDx)
 - Short DDx: Hutchinson's Gilford Progeria
 - Long DDx: Sotos Syndrome → case study could have lead to other ID/ND genes
- RareDxGPT was able to diagnosis less common manifestations: Pentalogy of Cantrell

Significance

- Potential to reduce rare disease diagnosis time
 - More specific rare disease knowledge than the average clinician
- Recommends Large Language Model testing procedures
 - Leveraging in-context learning and choosing the correct prompt style

Future Studies

- Explore ways to reduce bot hallucinations
 - Persisted even RareDxGPT
- Database utilizing a photo bank for rare skin disease diagnosis
- Test for top three possible diagnoses
- Compare results with RareDis Corpus to results with other information corpuses

JOURNAL ARTICLE

TrajVis: a visual clinical decision support system to translate artificial intelligence trajectory models in the precision management of chronic kidney disease

✉

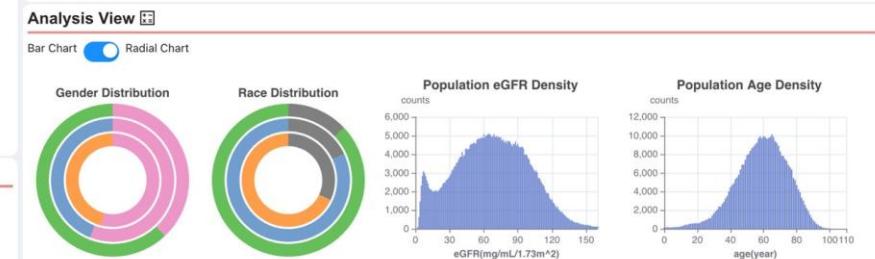
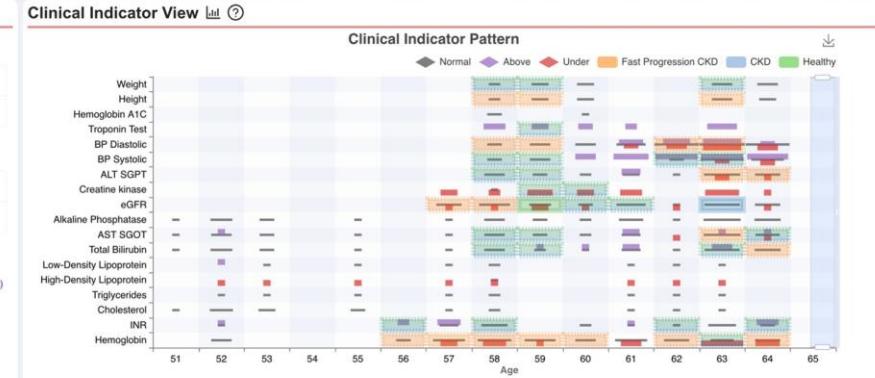
Zuotian Li, MS, Xiang Liu, PhD, Ziyang Tang, PhD, Nanxin Jin, MS, Pengyue Zhang, PhD,
Michael T Eadon, MD, Qianqian Song, PhD ✉, Yingjie V Chen, PhD ✉, Jing Su, PhD ✉

Journal of the American Medical Informatics Association, Volume 31, Issue 11, November
2024, Pages 2474–2485, <https://doi.org/10.1093/jamia/ocae158>

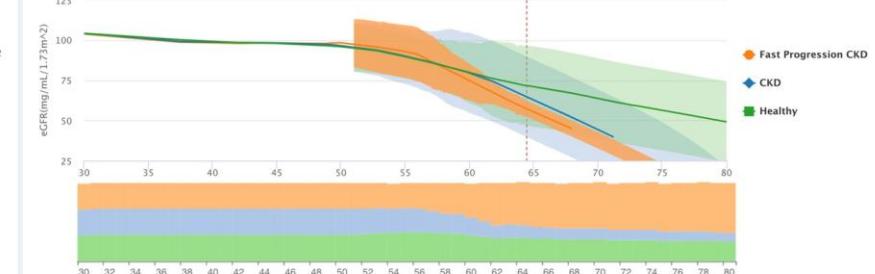
Published: 25 June 2024 **Article history ▾**



System interface

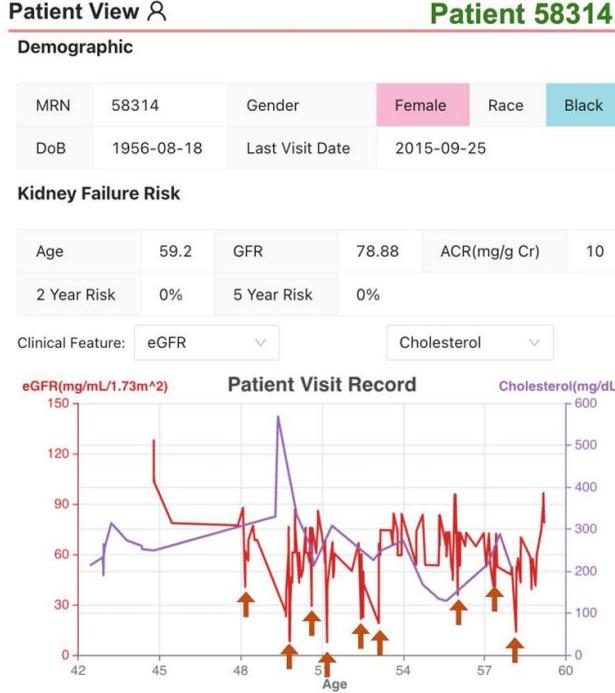
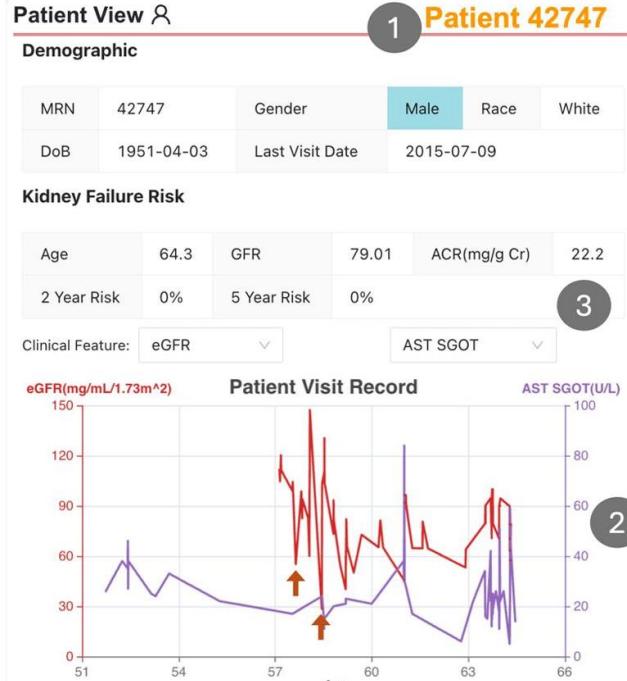


Trajectory Uncertainty Evolution

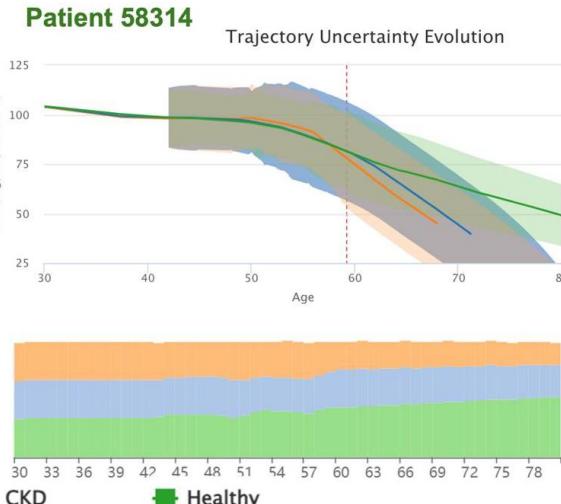
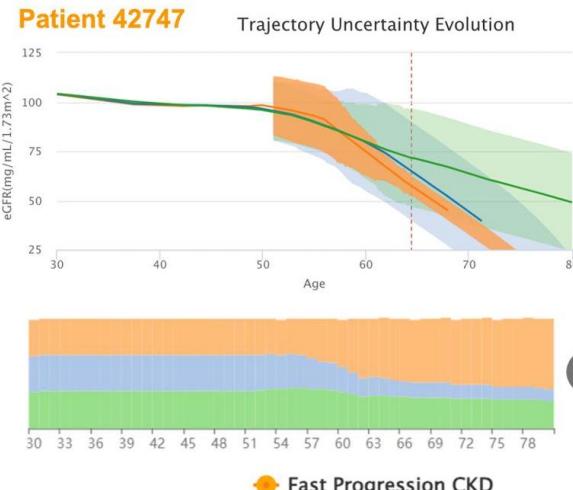


Case study

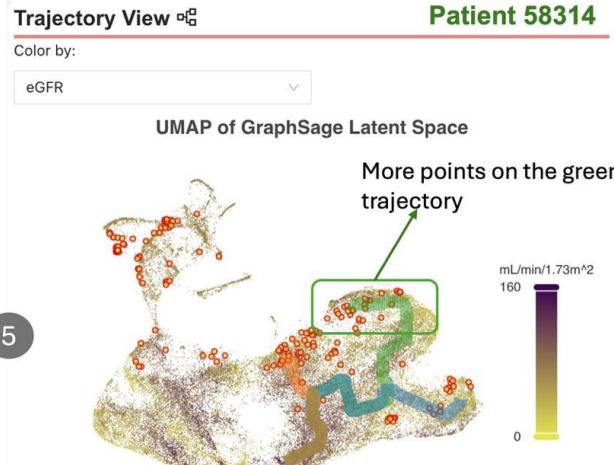
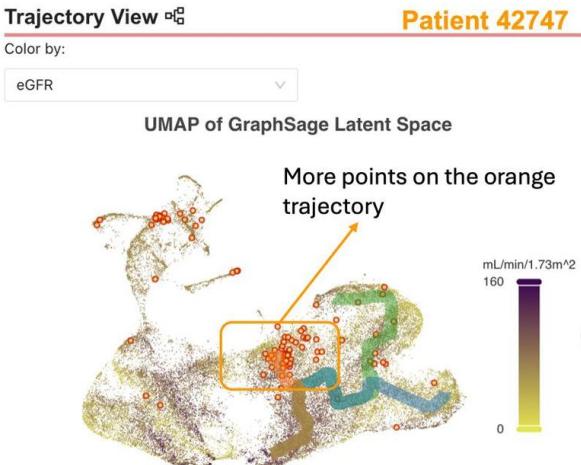
A Longitudinal series of clinical measurements



B Predicted CKD progression trajectories



C Explainable visualization: latent representations by graph AI





Article | [Open access](#) | Published: 04 July 2024

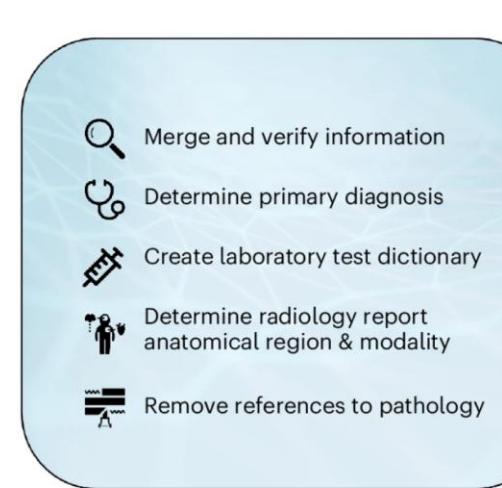
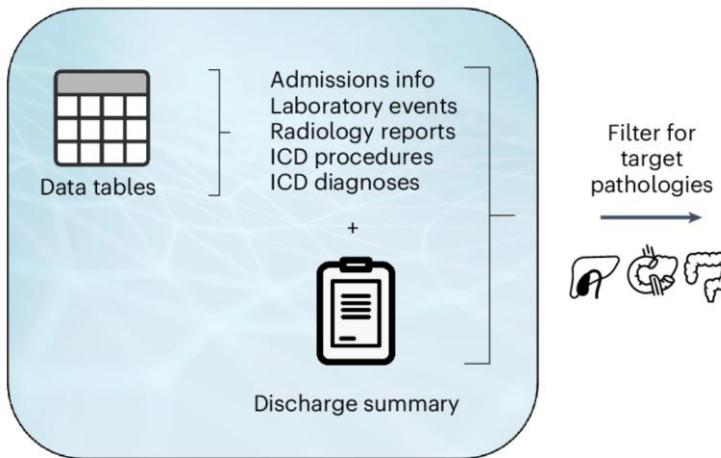
Evaluation and mitigation of the limitations of large language models in clinical decision-making

[Paul Hager](#)✉, [Friederike Jungmann](#), [Robbie Holland](#), [Kunal Bhagat](#), [Inga Hubrecht](#), [Manuel Knauer](#), [Jakob Vielhauer](#), [Marcus Makowski](#), [Rickmer Braren](#), [Georgios Kaassis](#) & [Daniel Rueckert](#)

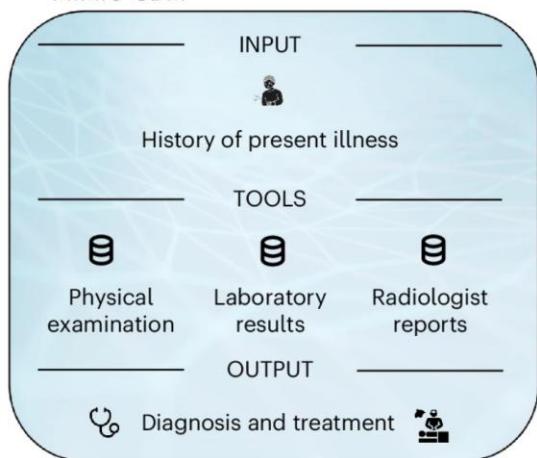
[Nature Medicine](#) **30**, 2613–2622 (2024) | [Cite this article](#)

Dataset creation and evaluation framework

a MIMIC-IV



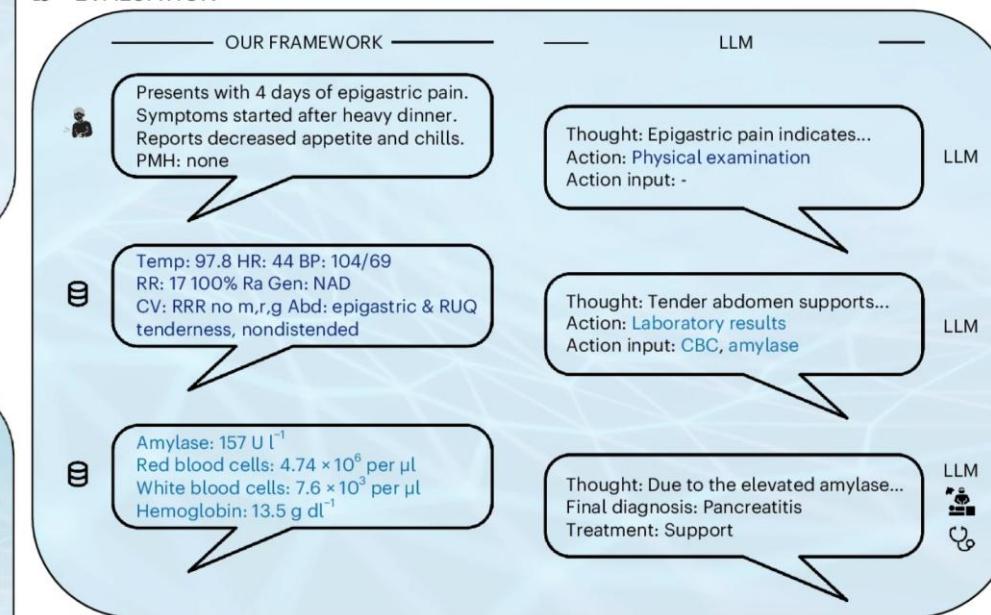
MIMIC-CDM



MIMIC-CDM statistics

2,400 cases			
Appendicitis	957	CT abdomen	1,836
Cholecystitis	648	Radiograph chest	1,728
Diverticulitis	257	US abdomen	1,325
Pancreatitis	538	Radiograph abdomen	342
		MRCP abdomen	227
		Other	501
<hr/>			
2,400 physical examinations	143,191 laboratory tests	5,959 radiologist reports	

b EVALUATION



Diagnostic accuracy



Clinical guidelines



Instruction following

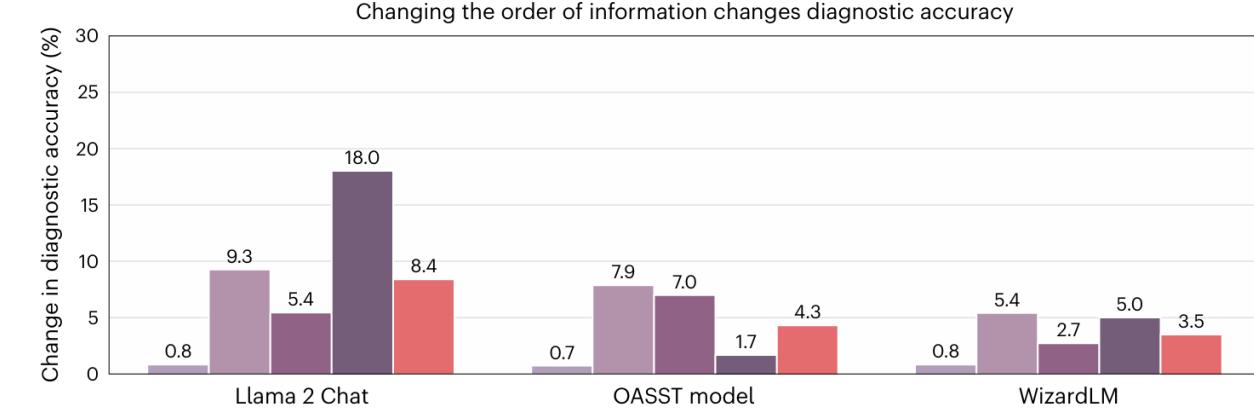
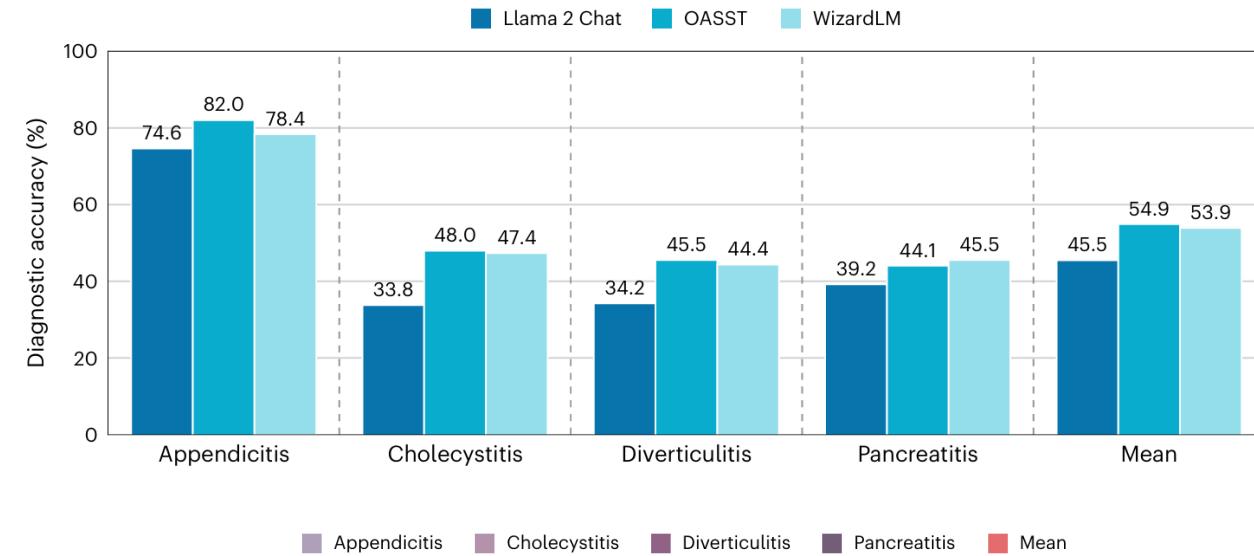
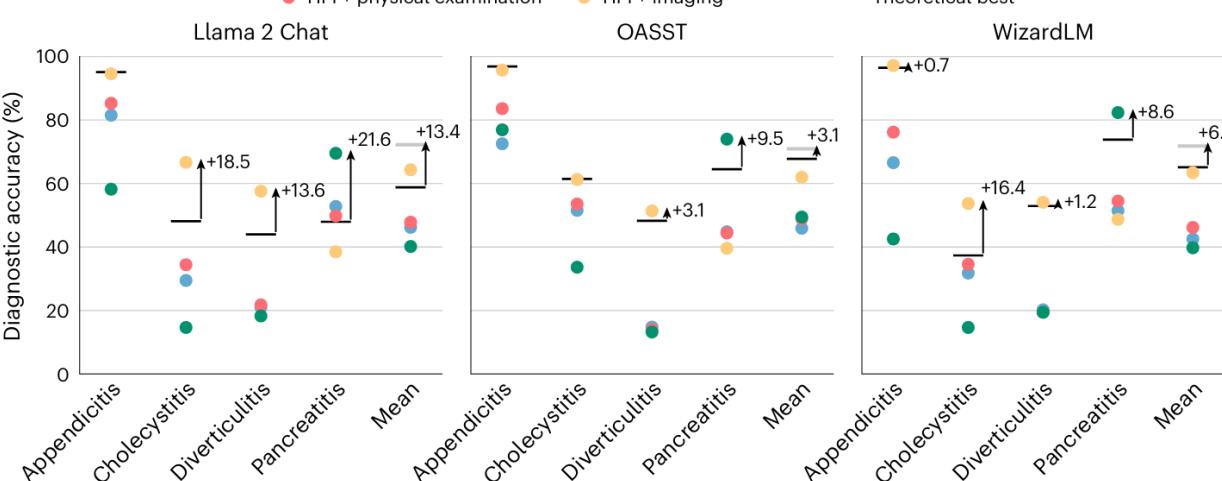
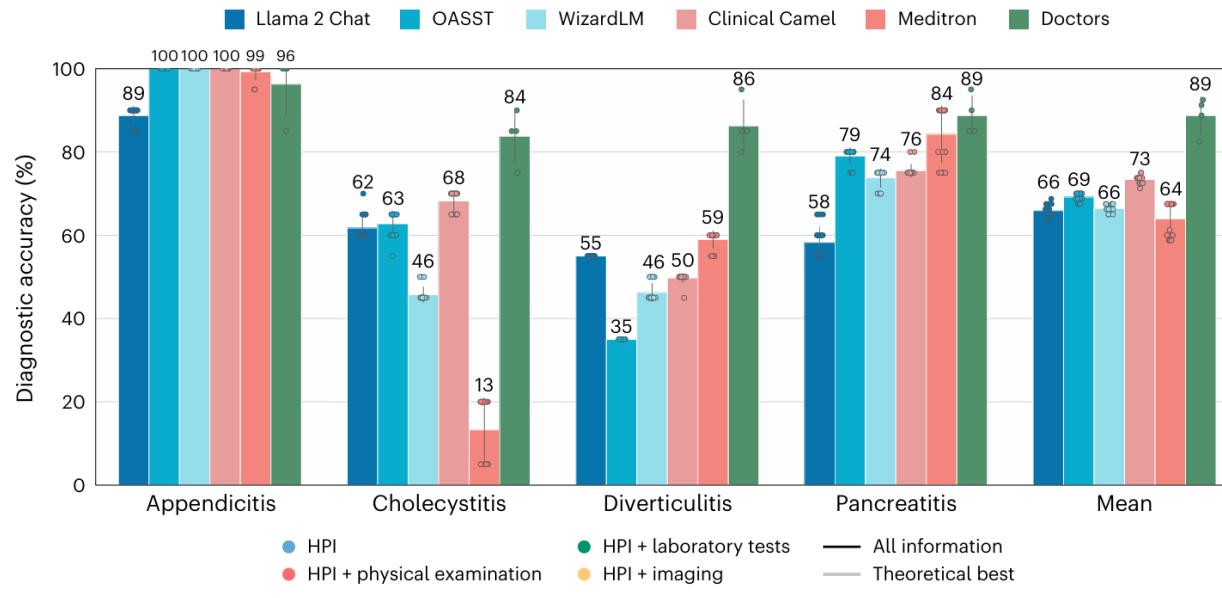


Lab test interpretation



Robustness

Challenges of LLMs in CDS



Article | [Open access](#) | Published: 08 October 2024



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

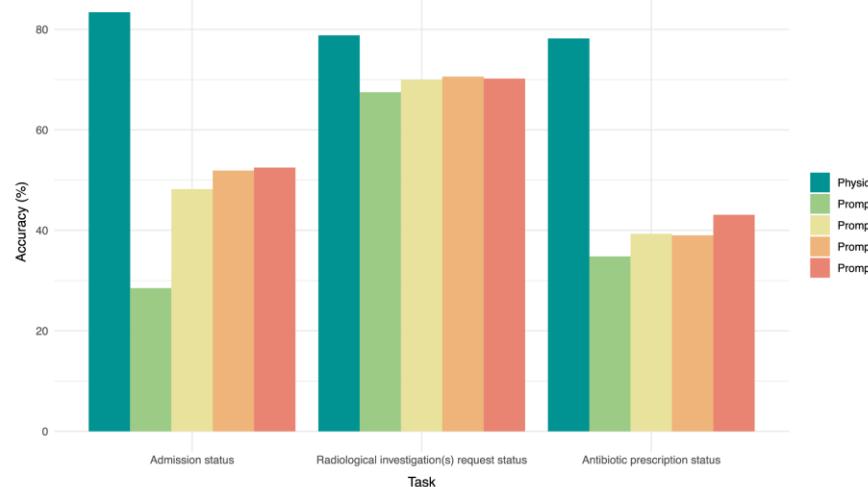
[Christopher Y. K. Williams](#)  , [Brenda Y. Miao](#), [Aaron E. Kornblith](#) & [Atul J. Butte](#)

[Nature Communications](#) **15**, Article number: 8236 (2024) | [Cite this article](#)

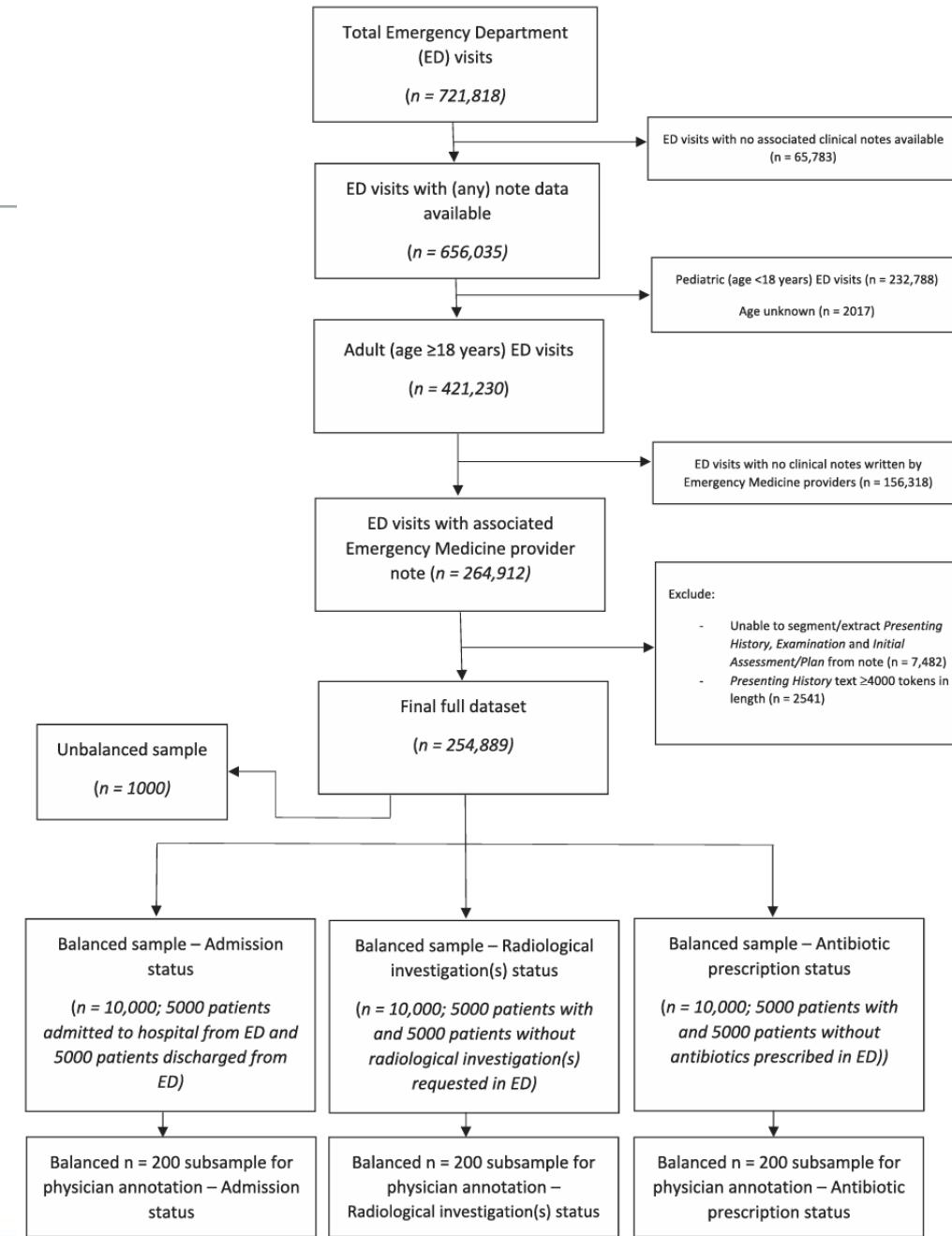
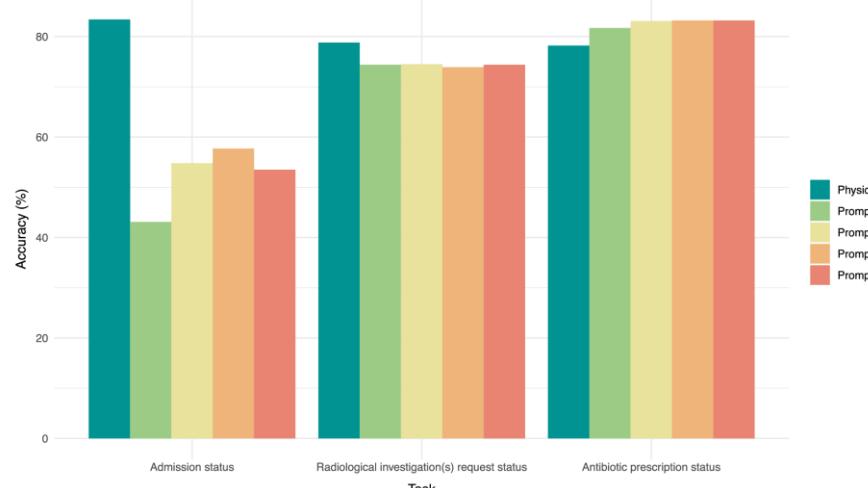
11k Accesses | **430** Altmetric | [Metrics](#)

Patient flowchart

A)



B)



Comparison of physician performance with GPT



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	Physician	73 (36.5)	26 (13)	74 (37)	27 (13.5)	0.73 (0.64–0.81) 0.74 (0.66–0.82)
		Prompt A	100 (50)	93 (46.5)	7 (3.5)	0 (0)	1 (1–1) 0.07 (0.03–0.13)
		Prompt B	98 (49)	67 (33.5)	33 (16.5)	2 (1)	0.98 (0.95–1.00) 0.33 (0.24–0.43)
		Prompt C	95 (47.5)	61 (30.5)	39 (19.5)	5 (2.5)	0.95 (0.90–0.99) 0.39 (0.30–0.49)
		Prompt D	93 (46.5)	60 (30)	40 (20)	7 (3.5)	0.93 (0.87–0.97) 0.40 (0.31–0.50)
		Physician	76 (38)	21 (10.5)	79 (39.5)	24 (12)	0.76 (0.67–0.84) 0.79 (0.70–0.87)
	2) Radiological investigation(s) request status	Prompt A	96 (48)	91 (45.5)	9 (4.5)	4 (2)	0.96 (0.92–0.99) 0.09 (0.04–0.15)
		Prompt B	93 (46.5)	83 (41.5)	17 (8.5)	7 (3.5)	0.93 (0.87–0.98) 0.17 (0.09–0.24)
		Prompt C	95 (47.5)	83 (41.5)	17 (8.5)	5 (2.5)	0.95 (0.90–0.99) 0.17 (0.10–0.24)
		Prompt D	95 (47.5)	84 (42)	16 (8)	5 (2.5)	0.95 (0.90–0.99) 0.16 (0.09–0.24)
		Physician	64 (32)	22 (11)	78 (39)	36 (18)	0.64 (0.55–0.73) 0.78 (0.70–0.86)
		Prompt A	93 (46.5)	74 (37)	26 (13)	7 (3.5)	0.93 (0.88–0.97) 0.26 (0.18–0.35)
b) GPT-4-turbo	1) Admission status	Prompt B	91 (45.5)	71 (35.5)	29 (14.5)	9 (4.5)	0.91 (0.85–0.96) 0.29 (0.20–0.39)
		Prompt C	92 (46)	68 (34)	32 (16)	8 (4)	0.92 (0.87–0.97) 0.32 (0.23–0.42)
		Prompt D	89 (44.5)	63 (31.5)	37 (18.5)	11 (5.5)	0.89 (0.83–0.95) 0.37 (0.27–0.47)
		Physician	73 (36.5)	26 (13)	74 (37)	27 (13.5)	0.73 (0.64–0.81) 0.74 (0.66–0.82)
		Prompt A	100 (50)	78 (39)	22 (11)	0 (0)	1 (1–1) 0.22 (0.14–0.31)
		Prompt B	96 (48)	59 (29.5)	41 (20.5)	4 (2)	0.96 (0.92–0.99) 0.41 (0.32–0.51)
	2) Radiological investigation(s) request status	Prompt C	94 (47)	56 (28)	44 (22)	6 (3)	0.94 (0.89–0.99) 0.44 (0.34–0.54)
		Prompt D	99 (49.5)	67 (33.5)	33 (16.5)	1 (0.5)	0.99 (0.97–1) 0.33 (0.24–0.42)
		Physician	76 (38)	21 (10.5)	79 (39.5)	24 (12)	0.76 (0.67–0.84) 0.79 (0.70–0.87)
		Prompt A	88 (44)	61 (30.5)	39 (19.5)	12 (6)	0.88 (0.82–0.94) 0.39 (0.29–0.48)
		Prompt B	79 (39.5)	37 (18.5)	63 (31.5)	21 (10.5)	0.79 (0.71–0.86) 0.63 (0.53–0.72)
		Prompt C	76 (38)	35 (17.5)	65 (32.5)	24 (12)	0.76 (0.67–0.84) 0.65 (0.56–0.75)
3) Antibiotic prescription status	1) Admission status	Prompt D	76 (38)	43 (21.5)	57 (28.5)	24 (12)	0.76 (0.67–0.84) 0.57 (0.47–0.67)
		Physician	64 (32)	22 (11)	78 (39)	36 (18)	0.64 (0.55–0.73) 0.78 (0.70–0.86)
		Prompt A	51 (25.5)	7 (3.5)	93 (46.5)	49 (24.5)	0.51 (0.41–0.6) 0.93 (0.88–0.98)
		Prompt B	44 (22)	5 (2.5)	95 (47.5)	56 (28)	0.44 (0.34–0.54) 0.95 (0.90–0.99)
	2) Radiological investigation(s) request status	Prompt C	39 (19.5)	5 (2.5)	95 (47.5)	61 (30.5)	0.39 (0.30–0.49) 0.95 (0.90–0.99)
		Prompt D	41 (20.5)	5 (2.5)	95 (47.5)	59 (29.5)	0.41 (0.32–0.51) 0.95 (0.90–0.99)
		Physician	64 (32)	22 (11)	78 (39)	36 (18)	0.64 (0.55–0.73) 0.78 (0.70–0.86)
		Prompt A	51 (25.5)	7 (3.5)	93 (46.5)	49 (24.5)	0.51 (0.41–0.6) 0.93 (0.88–0.98)

Public/global Health and Precision Medicine Informatics



Original Investigation | AI in Cardiology

May 1, 2024

Deep Learning-Based Assessment of Built Environment From Satellite Images and Cardiometabolic Disease Prevalence

Zhuo Chen, PhD^{1,2}; Jean-Eudes Dazard, PhD²; Yassin Khalifa, PhD²; Issam Motairek, MD¹; Catherine Kreatsoulas, PhD³; Sanjay Rajagopalan, MD^{1,2}; Sadeer Al-Kindi, MD^{1,2,4}

[» Author Affiliations](#) | [Article Information](#)

JAMA Cardiol. 2024;9(6):556-564. doi:10.1001/jamacardio.2024.0749



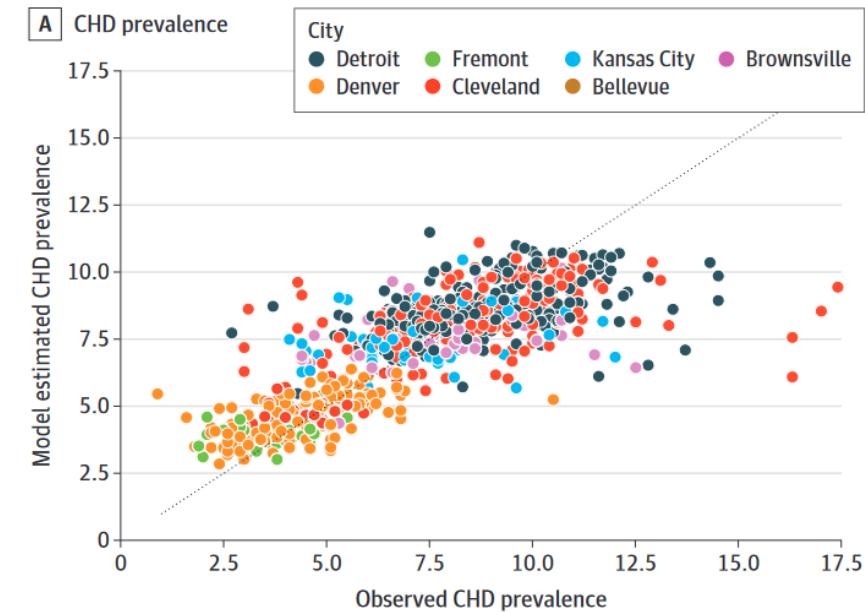
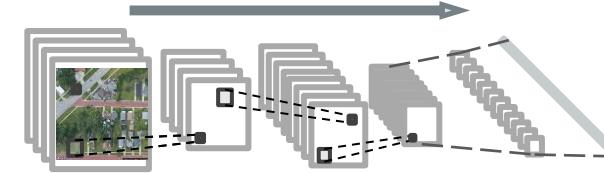
Deep Learning-Based Assessment of Built Environment From Satellite Images and Cardiometabolic Disease Prevalence



Background & Objective:

- Built environment influences cardiovascular and metabolic health, but large-scale assessment has been limited.
- This study used **deep learning on Google Satellite Images (GSI)** to extract built environment features and **analyzed their association** with cardiometabolic diseases (CHD, stroke, CKD) across **7 U.S. cities**.

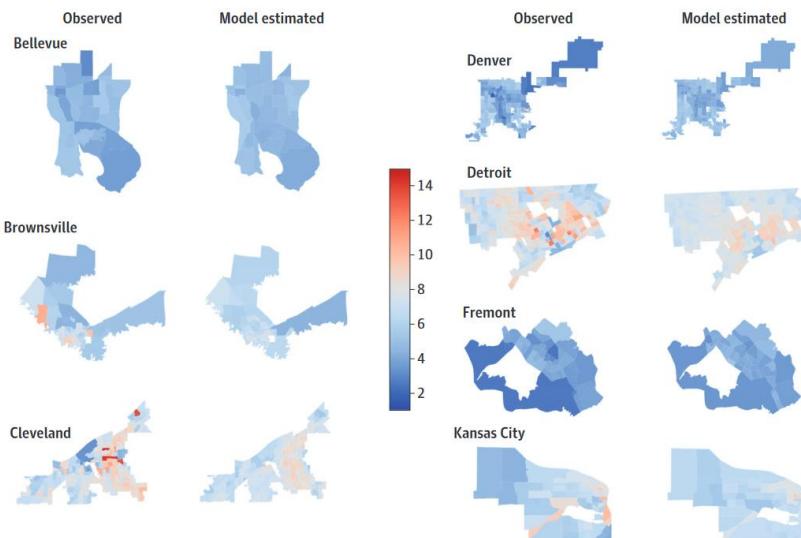
Feature Extraction Using Deep ConvNet



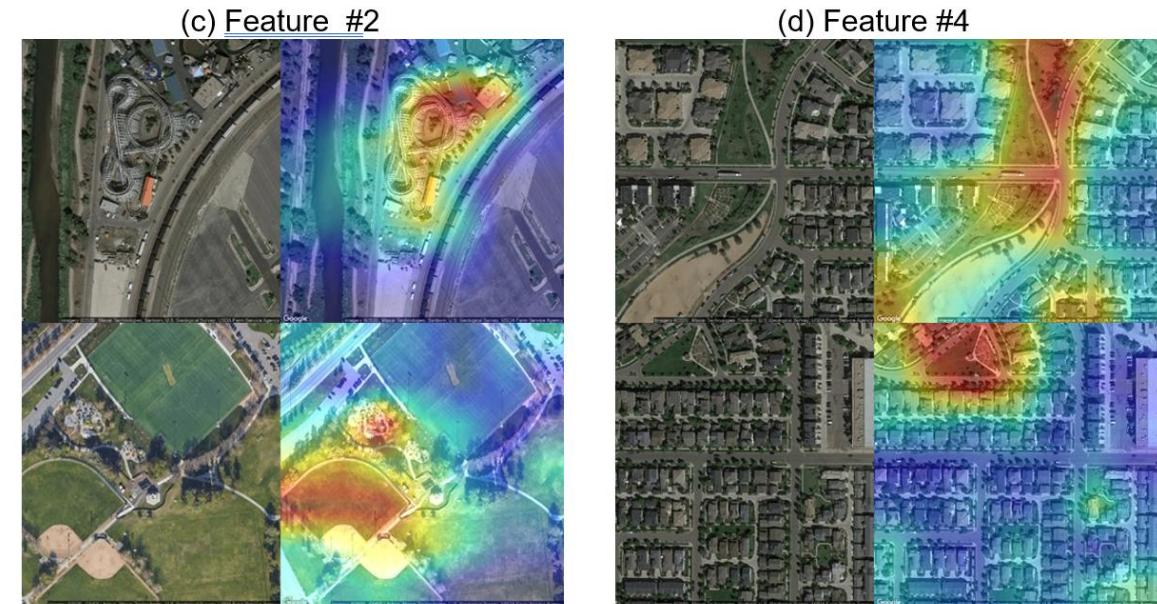
$R^2 = 0.60$ (CHD), 0.65 (stroke), 0.64 (CKD) at the census tract level.

Deep Learning-Based Assessment of Built Environment From Satellite Images and Cardiometabolic Disease Prevalence

- **Key Findings:**
- Deep learning-extracted features from **satellite imagery improved disease prevalence estimation** compared to demographic & social determinants alone.



Feature Interpretations Using Gradient-Weighted Class Activation Mapping (**Grad-CAMs**)



THE LANCET Digital Health



This journal Journals Publish Clinical Global health Multimedia Events About

ARTICLES · Volume 6, Issue 2, E93-E104, February 2024 · Open Access

 Download Full Issue

A scalable federated learning solution for secondary care using low-cost microcomputing: privacy-preserving development and evaluation of a COVID-19 screening test in UK hospitals

Andrew A S Soltan, MRCP  ^{a,b,c,d,e}  · Anshul Thakur, PhD ^c · Jenny Yang, MSc ^c · Prof Anoop Chauhan, FRCP ^f ·

Leon G D'Cruz, PhD ^f · Phillip Dickson, BSc ^g · Marina A Soltan, MRCP ^{h,i} · Prof David R Thickett, FRCP ^{h,i} ·

Prof David W Eyre, DPhil ^{a,d,j,k} · Prof Tingting Zhu, DPhil ^c · Prof David A Clifton, DPhil ^{c,k,l} Show less

A scalable federated learning solution for secondary care using low-cost microcomputing: privacy-preserving development and evaluation of a COVID-19 screening test in UK hospitals



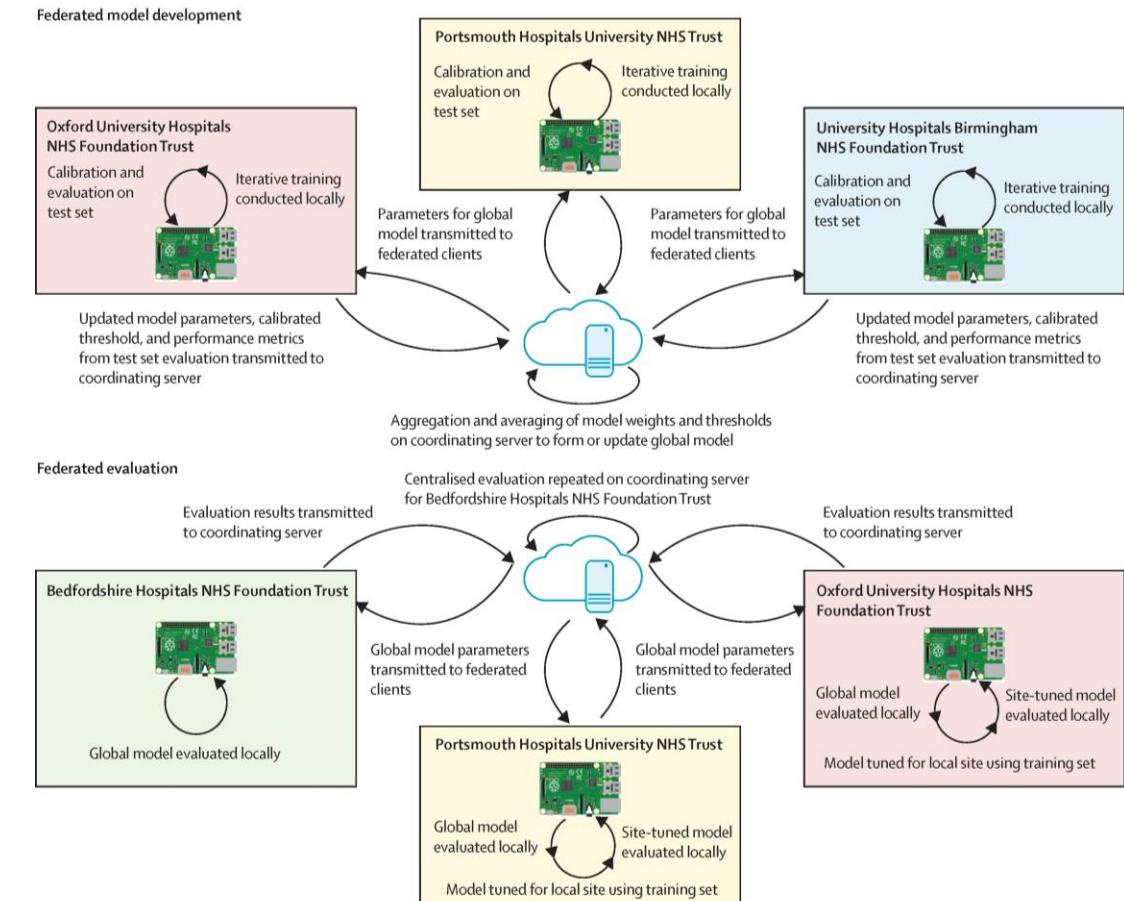
- The study developed full-stack federated learning using a low-cost embedded system using Raspberry Pi 4 to enable federated training of a COVID-19 screening AI model across four UK hospital groups without centralizing patient data
- 150 rounds of federated training using the FedAvg algorithm across three hospital groups (OUH, PUH, UHB). Two models were trained: logistic regression with L2 regularization and a deep neural network with dropout regularization

Feature set	Constituents
Vital signs	Heart rate, respiratory rate, oxygen saturations, blood pressure, temperature, oxygen delivery device level
Full blood count	Haemoglobin, haematocrit, mean cell volume, white cell count, neutrophil count, lymphocyte count, monocyte count, eosinophil count, basophil count, platelets
Urea and electrolytes	Sodium, potassium, creatinine, urea, estimated glomerular filtration rate
Liver function tests and C-reactive protein	Albumin, alkaline phosphatase, alanine aminotransferase, bilirubin, C-reactive protein

A scalable federated learning solution for secondary care using low-cost microcomputing: privacy-preserving development and evaluation of a COVID-19 screening test in UK hospitals



- Authors present an inexpensive and easy-to-use embedded system for federated learning and assess deployment of this system in the real-world secondary care setting
- Study is among the first to implement full-stack federated learning using an embedded microcomputing system, eliminating the need for specialist software or expertise at each site
- Established that federated learning can enable hospitals to contribute to AI model development while maintaining data privacy, reducing ethical and technical barriers to AI adoption in healthcare.





[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 05 February 2024

Deep learning-aided decision support for diagnosis of skin disease across skin tones

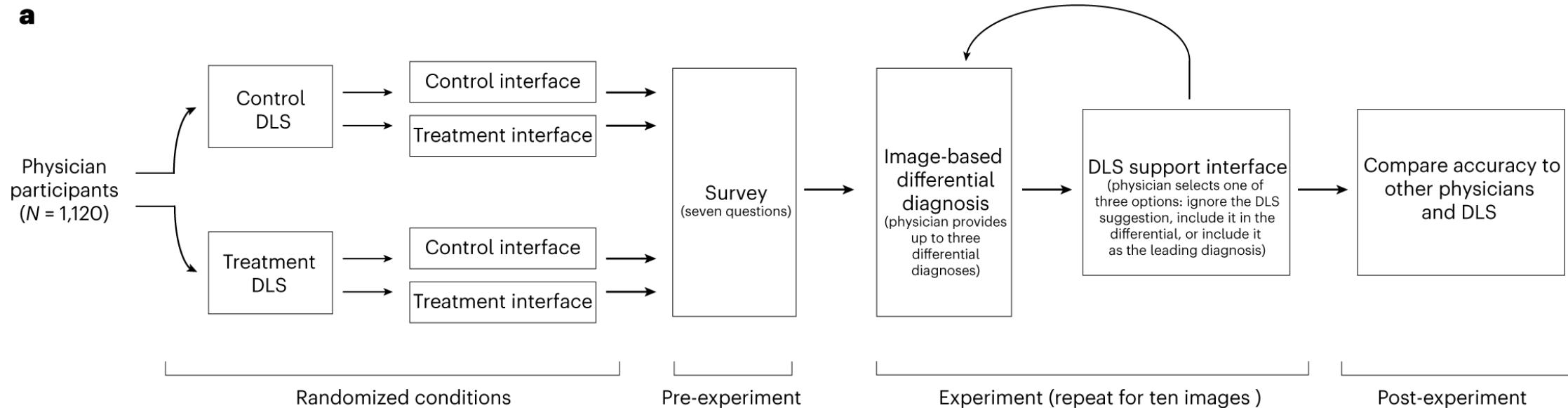
[Matthew Groh](#)✉, [Omar Badri](#), [Roxana Daneshjou](#), [Arash Koochek](#), [Caleb Harris](#), [Luis R. Soenksen](#), [P. Murali Doraiswamy](#) & [Rosalind Picard](#)

Nature Medicine **30**, 573–583 (2024) | [Cite this article](#)

Deep learning-aided decision support for diagnosis of skin disease across skin tones



a

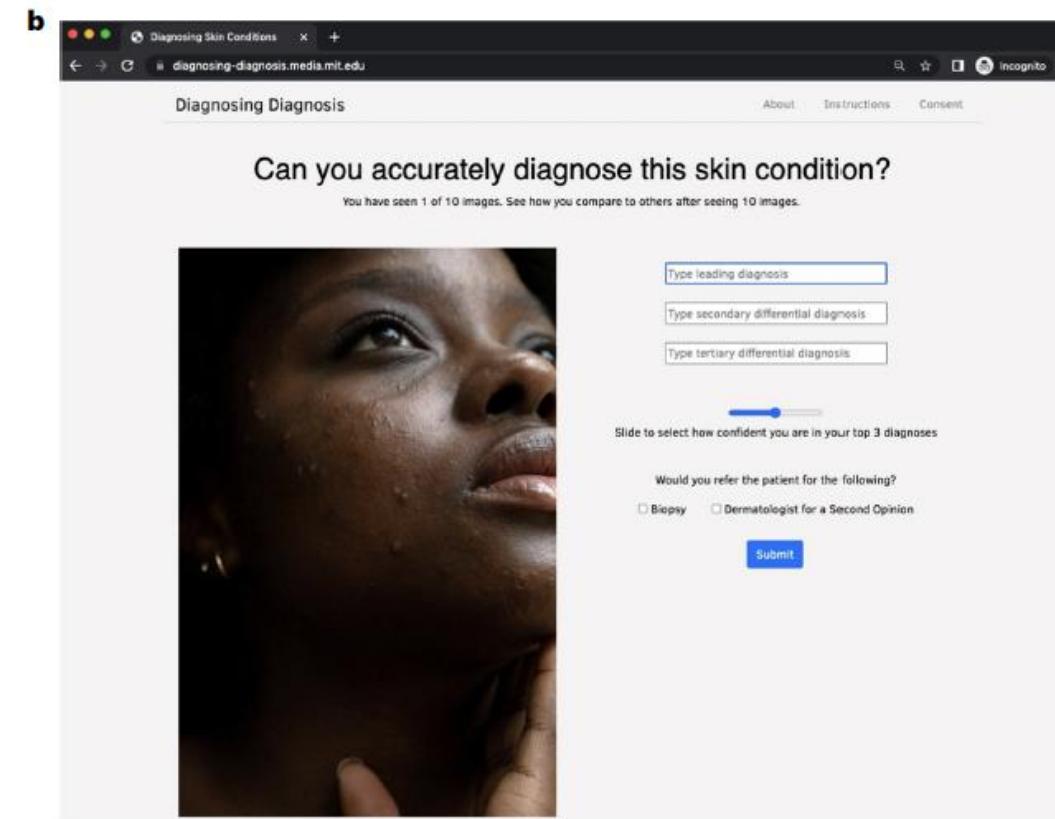


- This study examined the effectiveness of AI-assisted diagnosis in dermatology using a large-scale digital experiment involving 389 board-certified dermatologists (BCDs) and 459 primary-care physicians (PCPs) from 39 countries to evaluate the accuracy of diagnoses in a store-and-forward teledermatology simulation

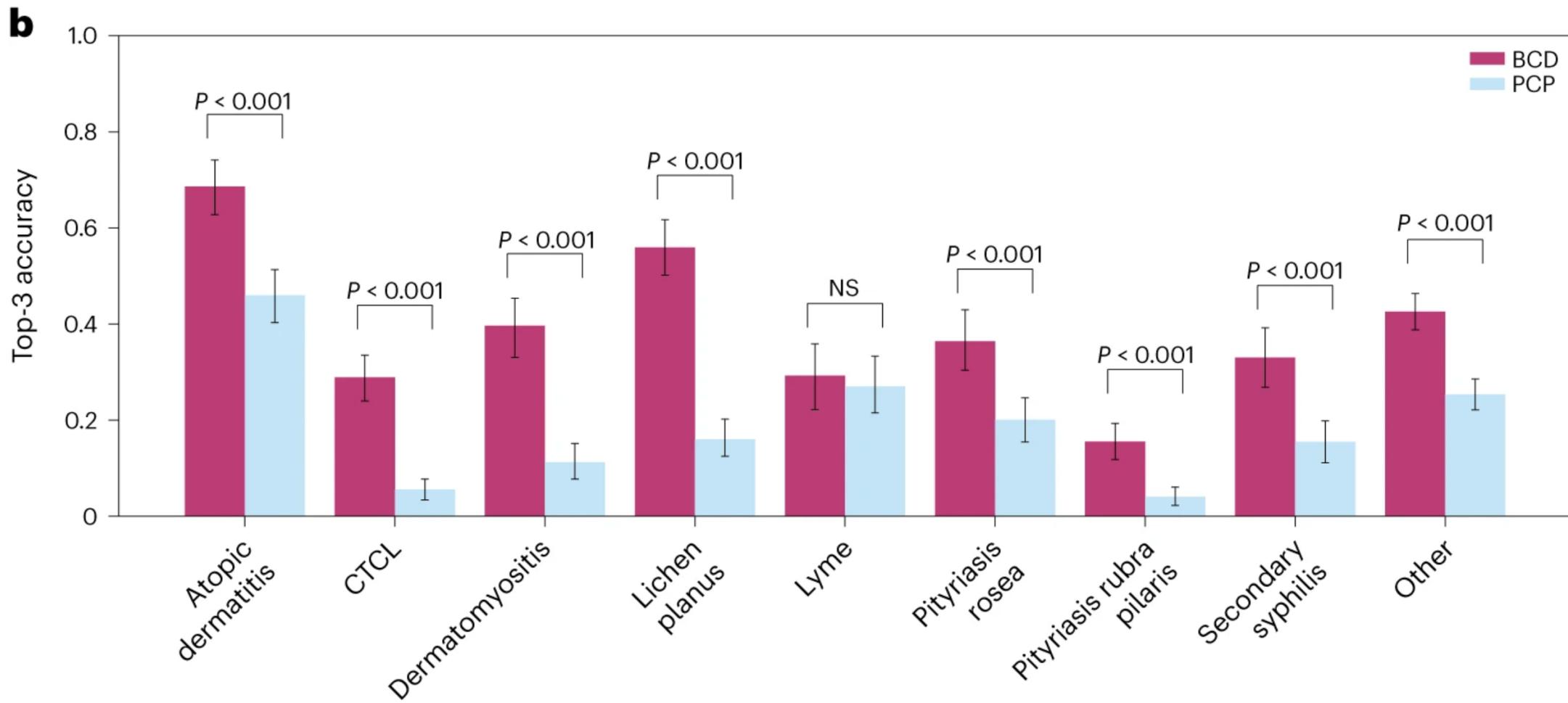
Deep learning-aided decision support for diagnosis of skin disease across skin tones



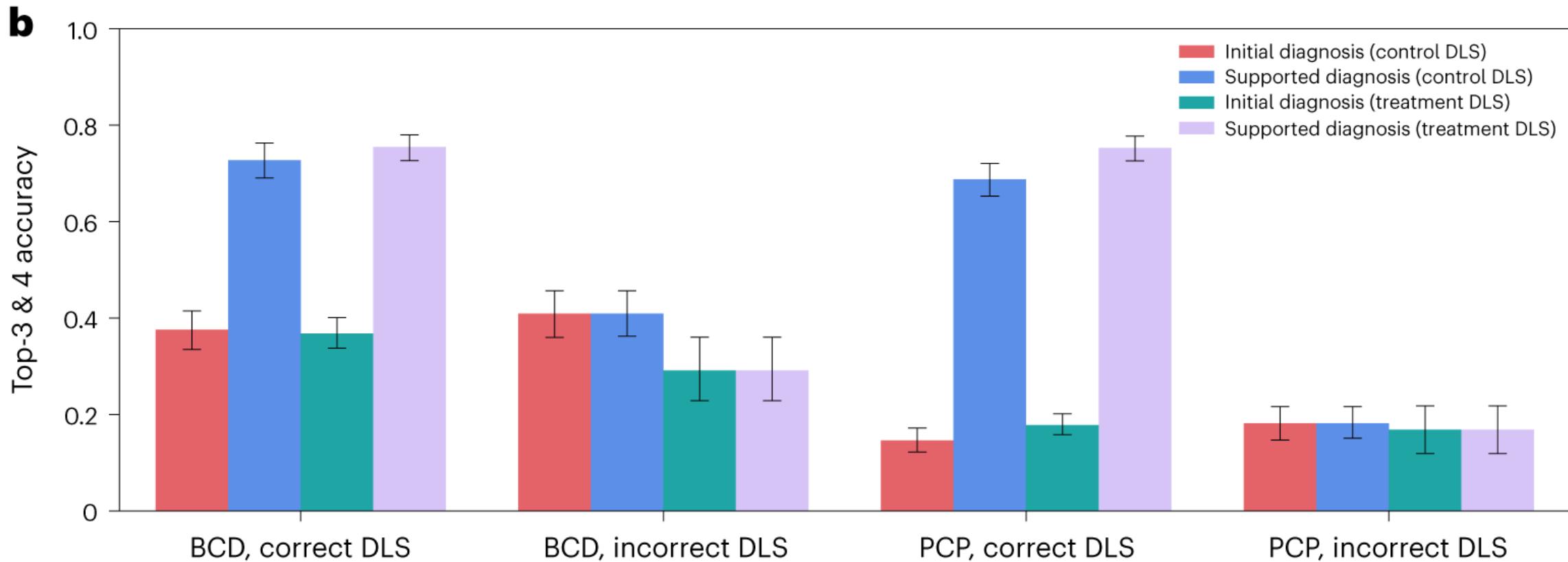
- DLS-based decision support increases top-1 diagnostic accuracy by 33% for BCDs and 69% for PCPs - improved sensitivity in diagnosing specific skin diseases
- Large increases in overall accuracy by physician–machine partnerships can inform diagnostic procedures to improve triage and reduce delayed and missed diagnoses
- Found differences in how often BCDs and PCPs refer patients with light and dark skin for biopsy



Deep learning-aided decision support for diagnosis of skin disease across skin tones



Deep learning-aided decision support for diagnosis of skin disease across skin tones



[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 22 July 2024

A foundation model for clinical-grade computational pathology and rare cancers detection

[Eugene Vorontsov](#), [Alican Bozkurt](#), [Adam Casson](#), [George Shaikovski](#), [Michal Zelechowski](#), [Kristen Severson](#), [Eric Zimmermann](#), [James Hall](#), [Neil Tenenholtz](#), [Nicolo Fusi](#), [Ellen Yang](#), [Philippe Mathieu](#), [Alexander van Eck](#), [Donghun Lee](#), [Julian Viret](#), [Eric Robert](#), [Yi Kan Wang](#), [Jeremy D. Kunz](#), [Matthew C. H. Lee](#), [Jan H. Bernhard](#), [Ran A. Godrich](#), [Gerard Oakley](#), [Ewan Millar](#), [Matthew Hanna](#), [Hannah Wen](#), [Juan A. Retamero](#), [William A. Moye](#), [Razik Yousfi](#), [Christopher Kanan](#), [David S. Klimstra](#), [Brandon Rothrock](#), [Siqi Liu](#)✉ & [Thomas J. Fuchs](#) — Show fewer authors

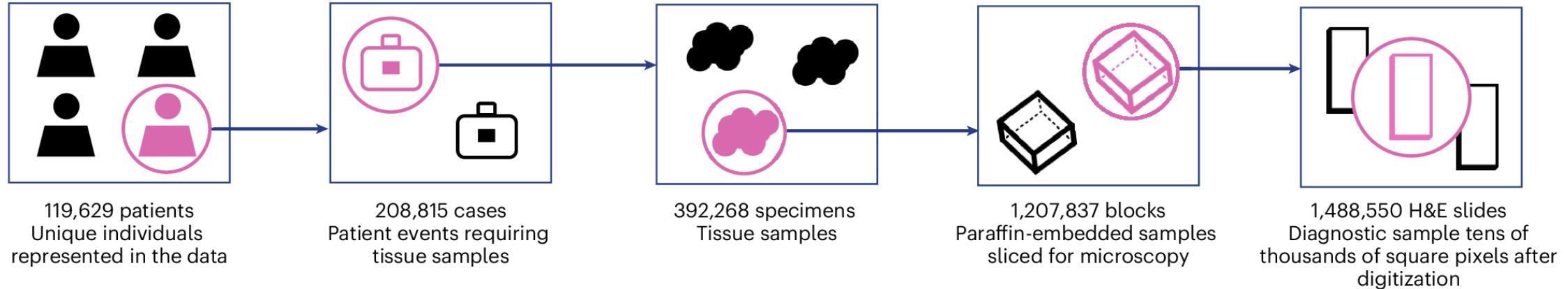
[Nature Medicine](#) **30**, 2924–2935 (2024) | [Cite this article](#)



A foundation model for clinical-grade computational pathology and rare cancers detection



a



- This study introduces Virchow, the largest foundation model for computational pathology, designed to model the diverse patterns observed in pathology images for clinical decision support
- The model enables pan-cancer detection with a 0.95 AUROC across 9 common and 7 rare cancers, demonstrating strong biomarker prediction and cell identification capabilities

A foundation model for clinical-grade computational pathology and rare cancers detection



- Virchow trained on data from ~100,000 patients; 1.5 million H&E stained WSIs acquired from Memorial Sloan Kettering Cancer Center (MSKCC). Training data composed of cancerous and benign tissues, collected via biopsy (63%) and resection (37%), from 17 high-level tissues
- Study assesses the performance of a pan-cancer model trained using the Virchow embeddings to predict specimen-level cancer across different tissues
- Virchow achieves comparable or superior performance to clinical-grade tissue-specific models (UNI, Phikon, and CTransPath), even with less training data, particularly for rare cancer variants
- These findings highlight the potential of foundation models to advance pathology AI applications, especially with limited labeled data

[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 26 August 2024

Brain clocks capture diversity and disparities in aging and dementia across geographically diverse populations

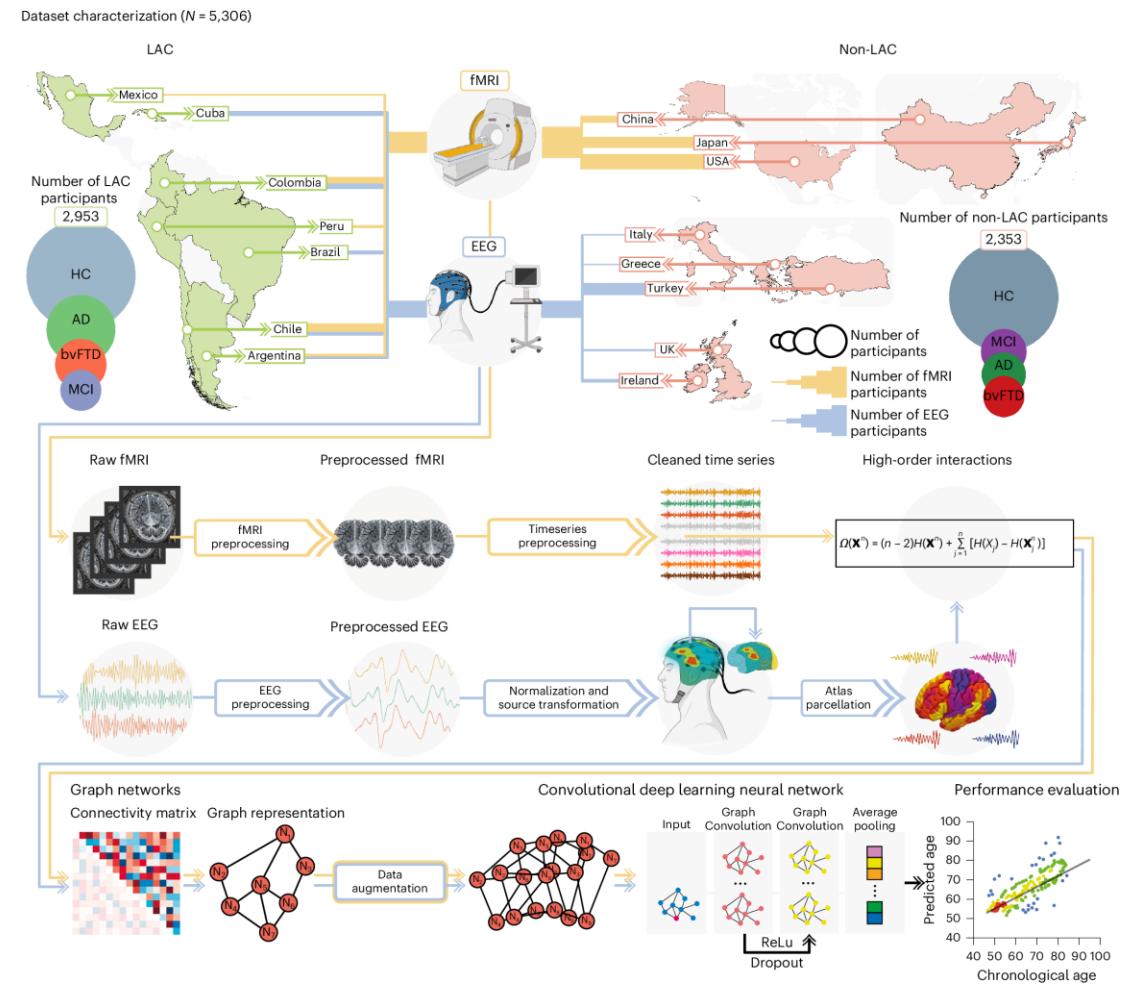


[Sebastian Moguilner](#), [Sandra Baez](#), [Hernan Hernandez](#), [Joaquín Migeot](#), [Agustina Legaz](#), [Raul Gonzalez-Gomez](#), [Francesca R. Farina](#), [Pavel Prado](#), [Jhosmary Cuadros](#), [Enzo Tagliazucchi](#), [Florencia Altschuler](#), [Marcelo Adrián Maito](#), [María E. Godoy](#), [Josephine Cruzat](#), [Pedro A. Valdes-Sosa](#), [Francisco Lopera](#), [John Fredy Ochoa-Gómez](#), [Alfredis Gonzalez Hernandez](#), [Jasmin Bonilla-Santos](#), [Rodrigo A. Gonzalez-Montealegre](#), [Renato Anghinah](#), [Luís E. d'Almeida Manfrinati](#), [Sol Fittipaldi](#), [Vicente Medel](#), ... [Agustin Ibanez](#)  + Show authors

[Nature Medicine](#) **30**, 3646–3657 (2024) | [Cite this article](#)

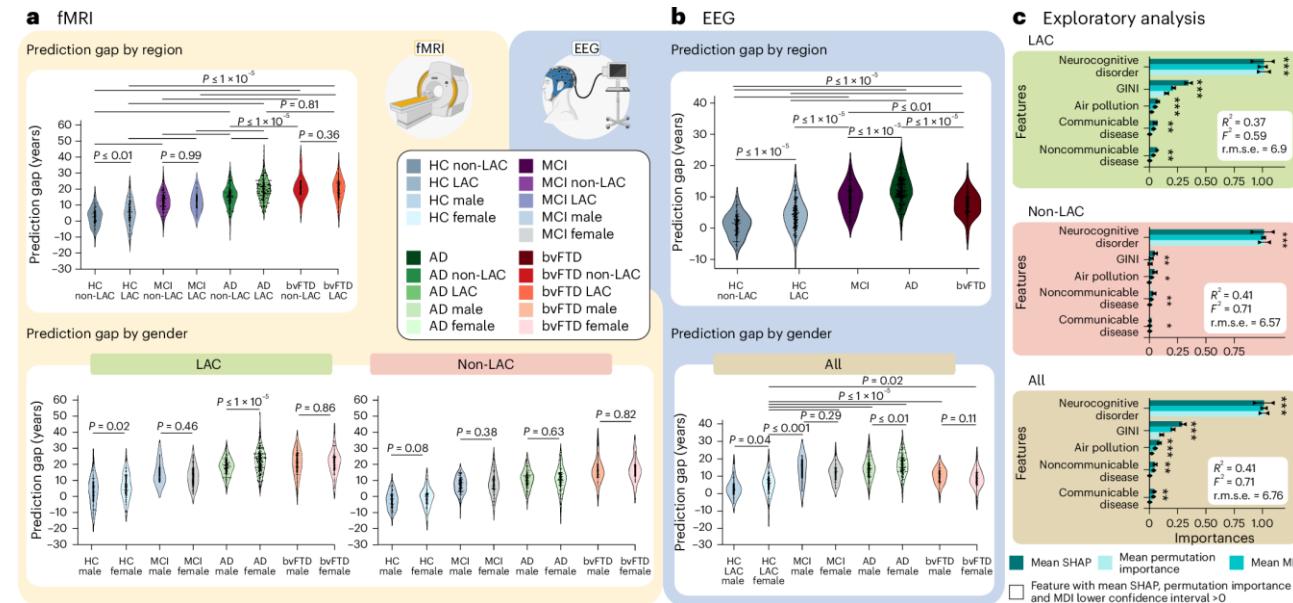
Brain clocks capture diversity and disparities in aging and dementia across geographically diverse populations

- This study develops brain-age gap models using deep learning architectures and fMRI and ECG datasets from 5,306 participants across 15 countries
- By incorporating diverse factors such as geography, socioeconomic status, and neurodegeneration, the models reveal significant discrepancies in brain-age gaps, particularly in Latin American and Caribbean (LAC) populations

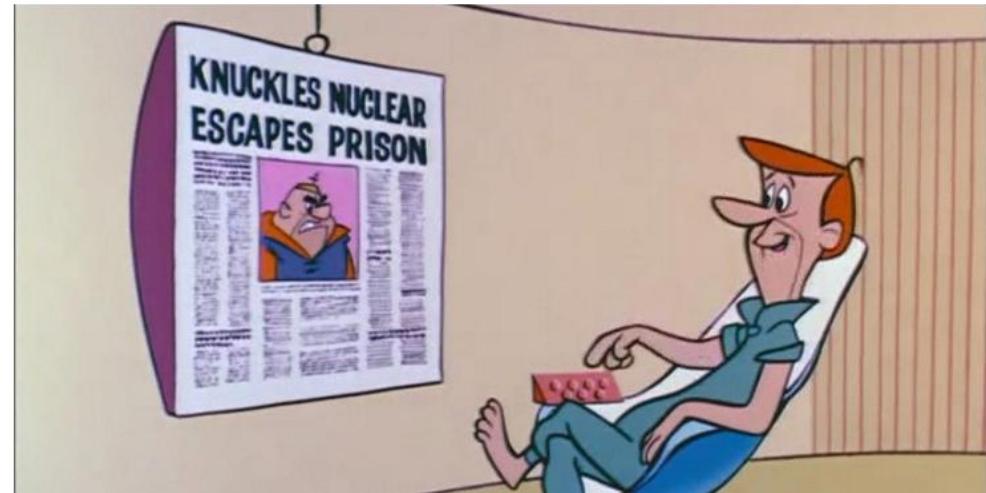
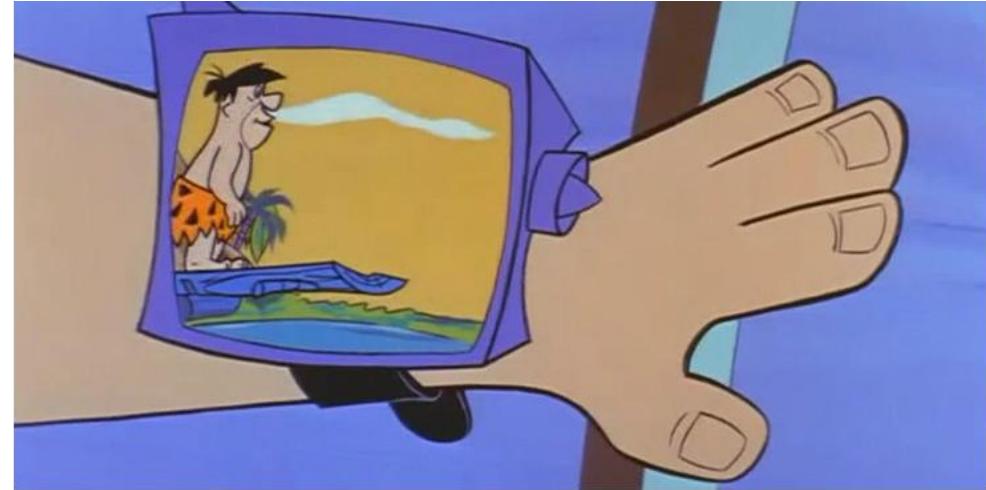
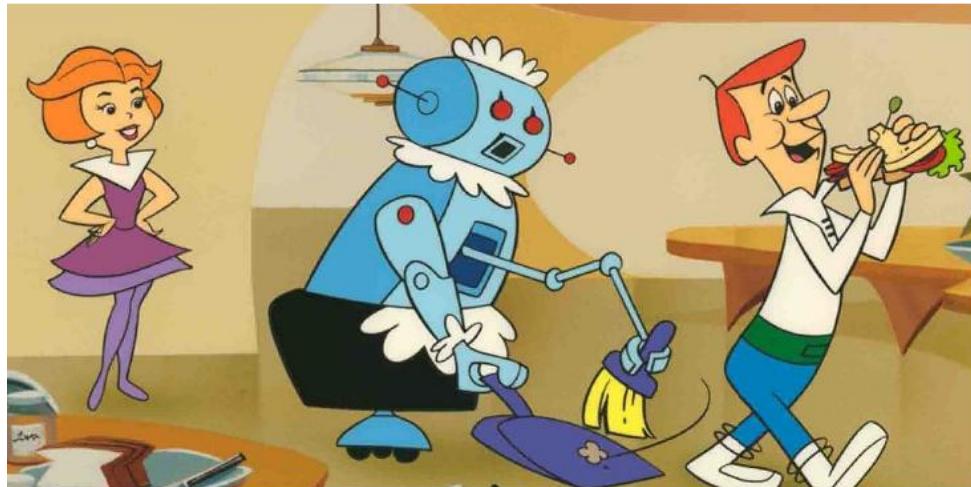


Brain clocks capture diversity and disparities in aging and dementia across geographically diverse populations

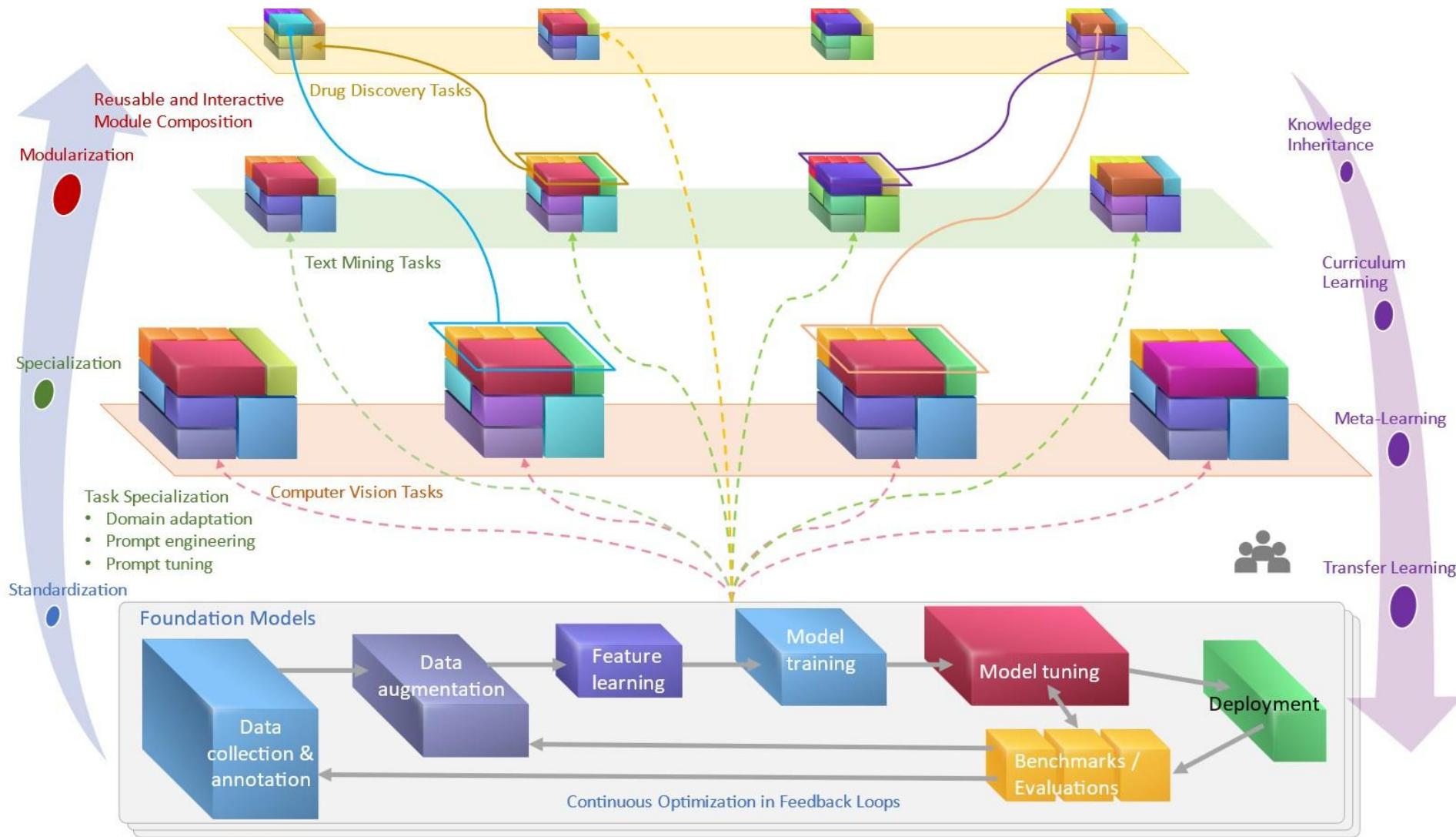
- Graph convolutional networks (GCN) used to integrate and cross-validate fMRI and EEG data to predict brain-age gaps
- Total dataset consisted of 5,306 participants, with 2,953 undergoing fMRI and 2,353 EEG acquisitions. 3,509 controls, 517 had MCI, 828 had AD and 463 had behavioral variant frontotemporal dementia (bvFTD).
- The findings highlight the potential of deep learning techniques to track longitudinal structural changes to quantify accelerated brain aging and its associations with structural inequality, health disparities, and gender differences.



Science Fiction or Reality?



Industrial Revolution for Health Care AI



<https://arxiv.org/abs/2311.02278>