



Center for Biomedical
Informatics & Biostatistics

For CME



Clinical Research Applications of AI: Successes, Opportunities and Cautions

Justin Starren, MD, PhD, FACMI

Director, Center for Biomedical Informatics and Biostatistics (CB2)

Professor of Radiology, Health Sciences, and Biomedical
Informatics

University of Arizona

Me and the Legal Stuff

“Nelson Man-Data”

Get systems that don't **trust** each other and that don't **like** each other to work together for the **greater good**.

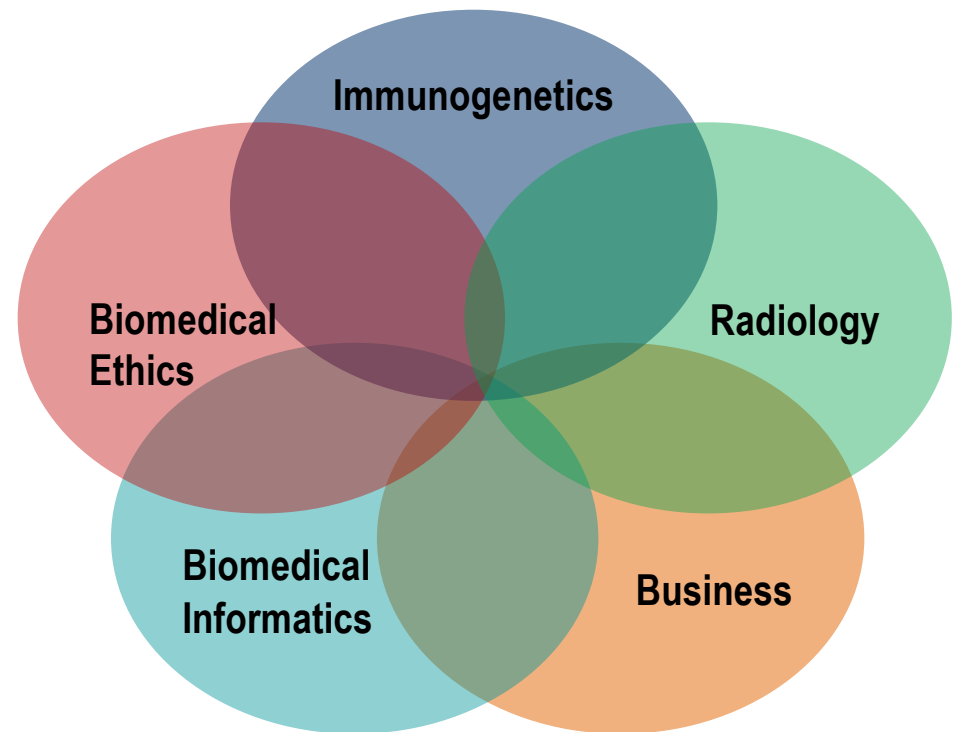
No relevant financial relationship in relation to this educational activity.

2 non-commercialized patents

Several Academic External Advisory Boards

Recent funding

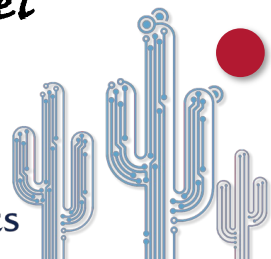
- NIH
- Greenwall Foundation



Special Thanks

- *Yuan Luo*
- *Melissa Handel*
- *Laura Wiley*

AI was used in the creation of this presentation



AI Methods in Clinical Research

CHEMICAL
REVIEWS

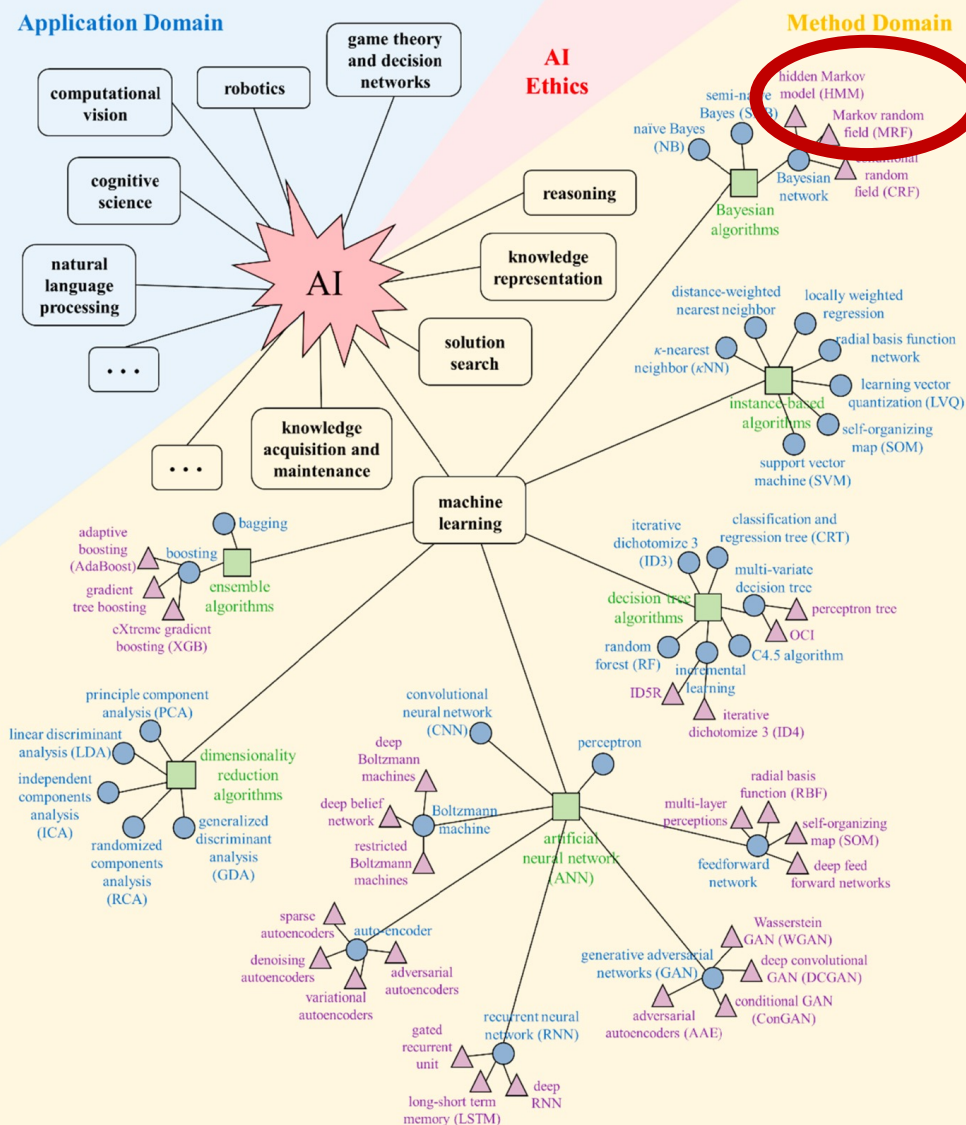
Cite This: Chem. Rev. 2019, 119, 10520-10594

ART
AUTHORSHIP
Review
pubs.acs.org/CR

Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery

Xin Yang,[†] Yifei Wang,[†] Ryan Byrne,[‡] Gisbert Schneider,^{*,‡} and Shengyong Yang^{*,†}

<https://doi.org/10.1021/acs.chemrev.8b00728>





Warning: Aphorisms Ahead

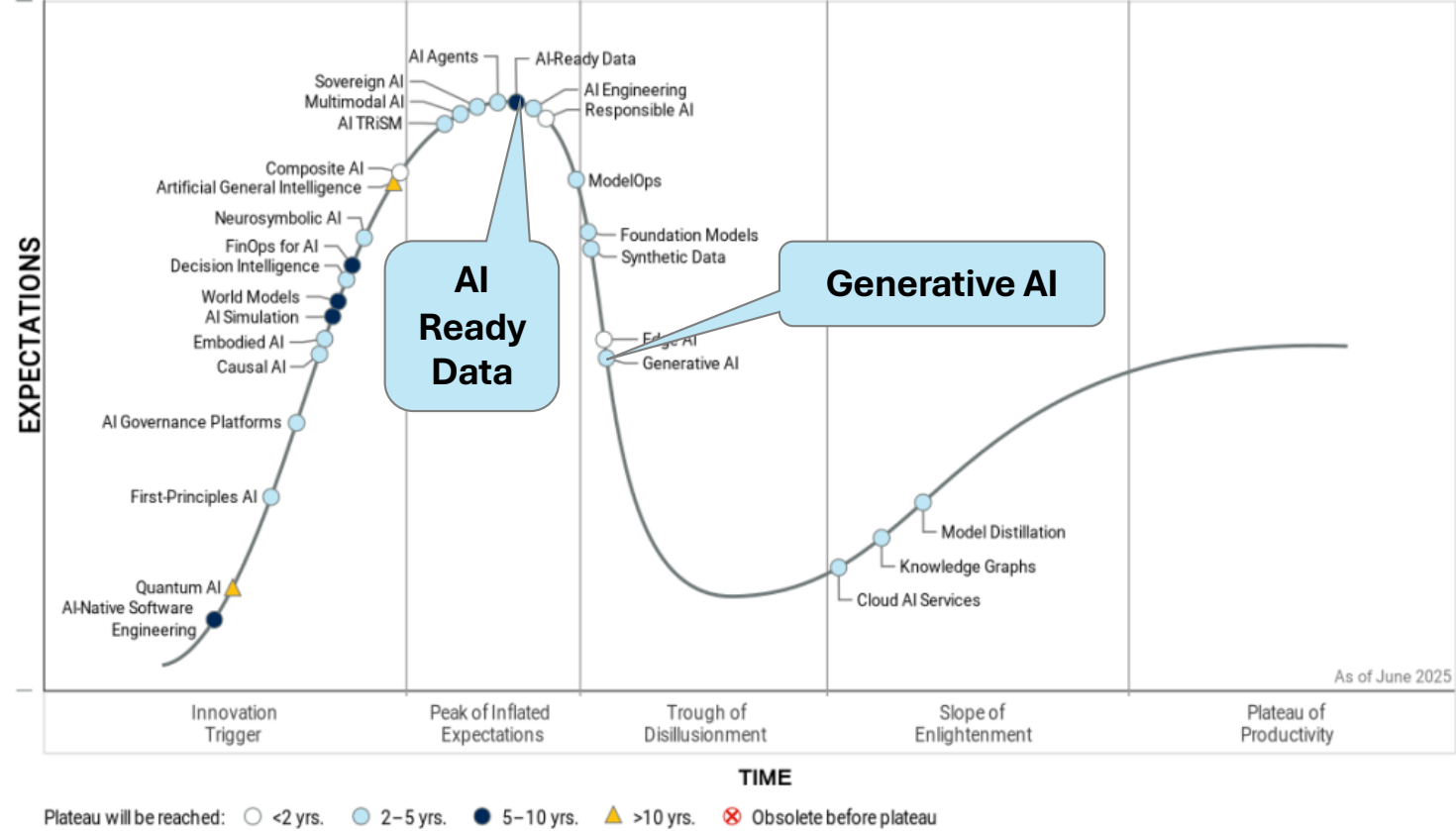
**If it just works,
we don't call it AI.**



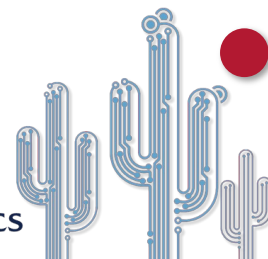
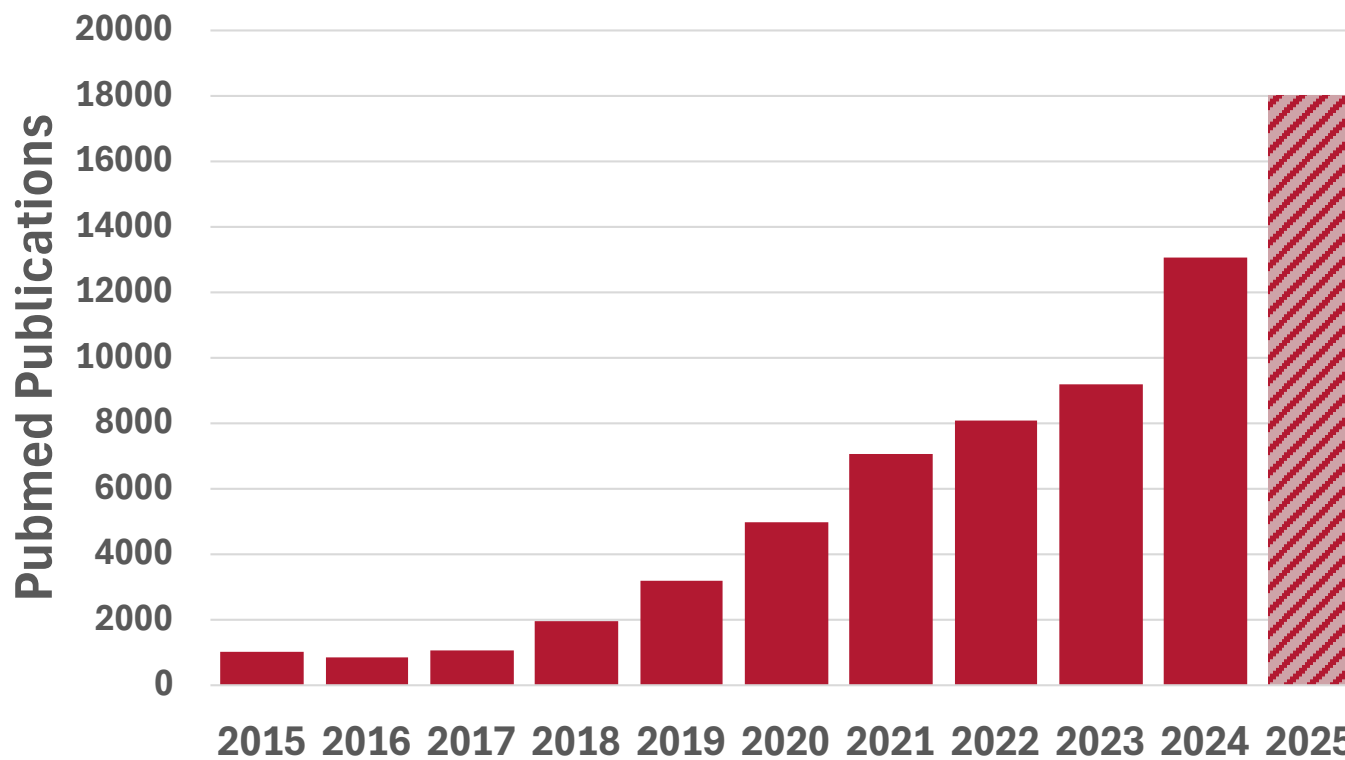
THE TIMES, THEY ARE A CHANGIN'



Gartner Hype Cycle for AI 2025



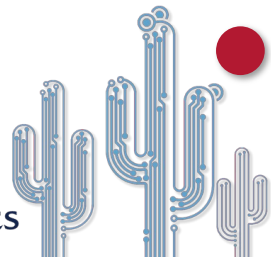
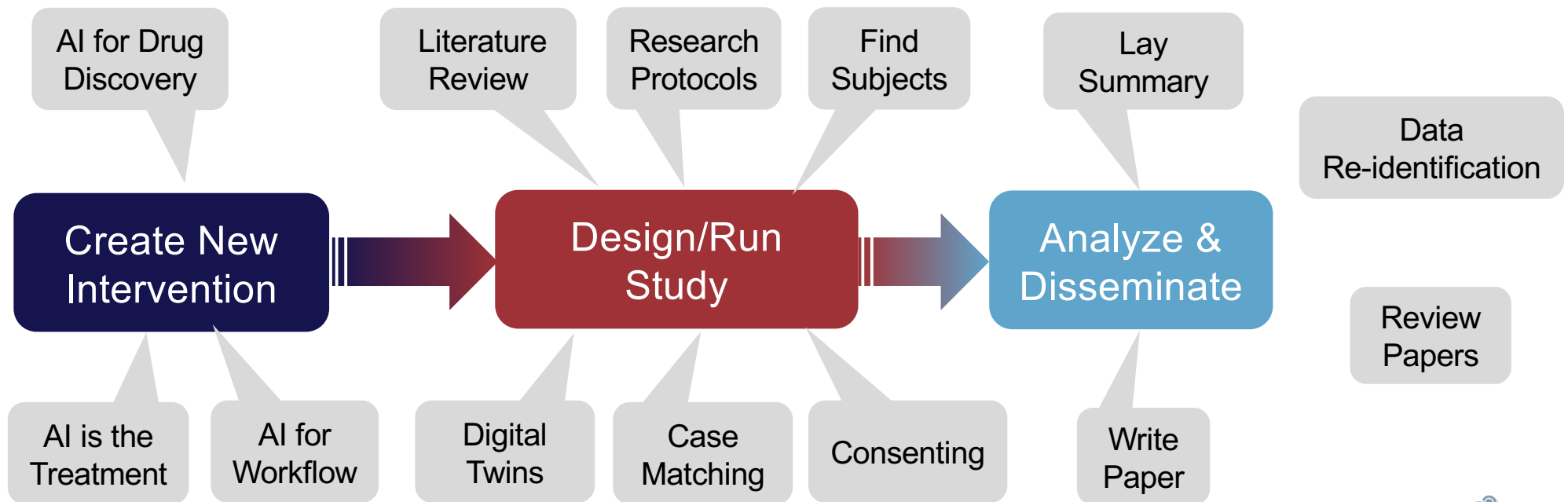
Papers on AI + Clinical Research



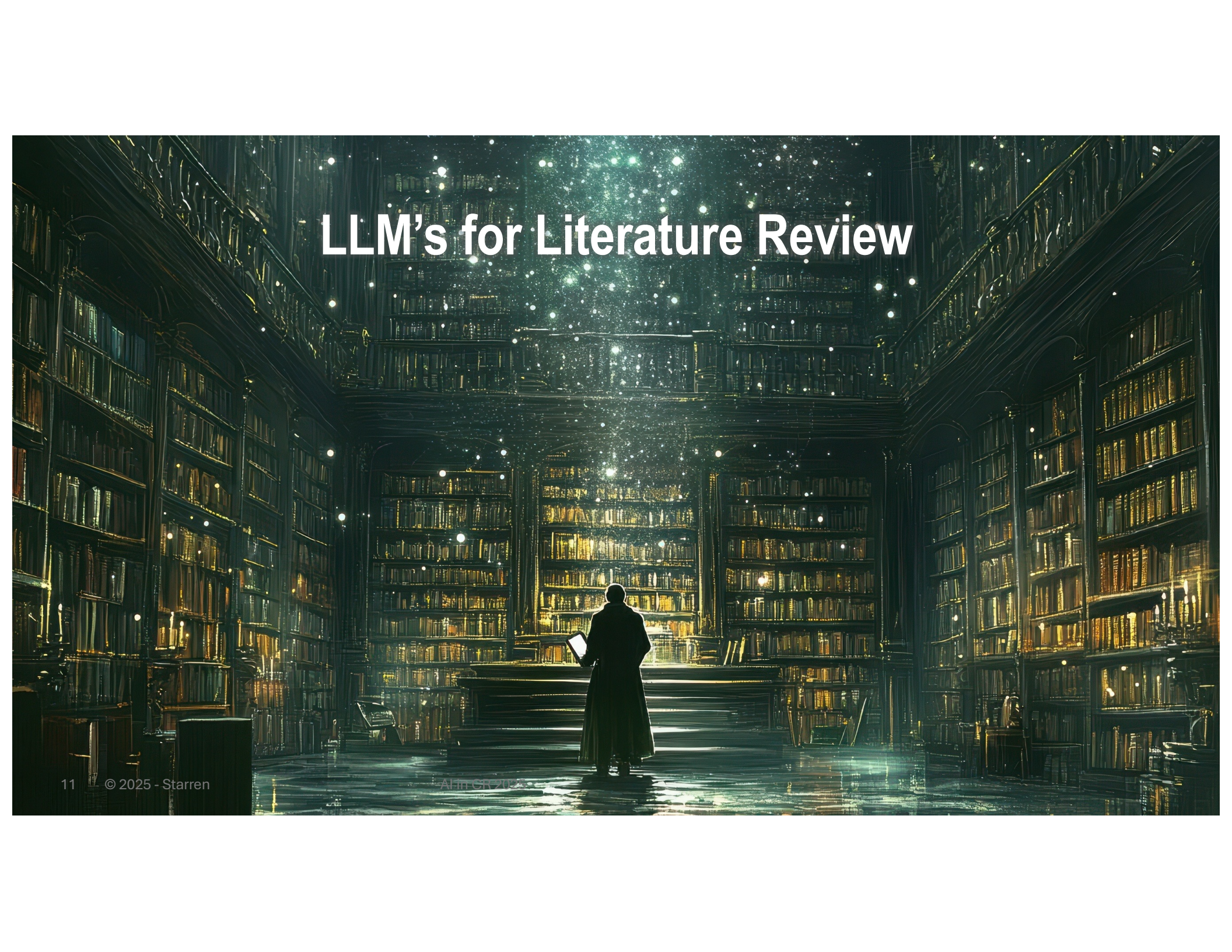


Most AI papers are
“What if...”

Uses of AI in Clinical Research



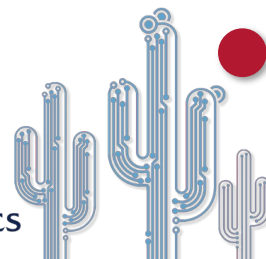
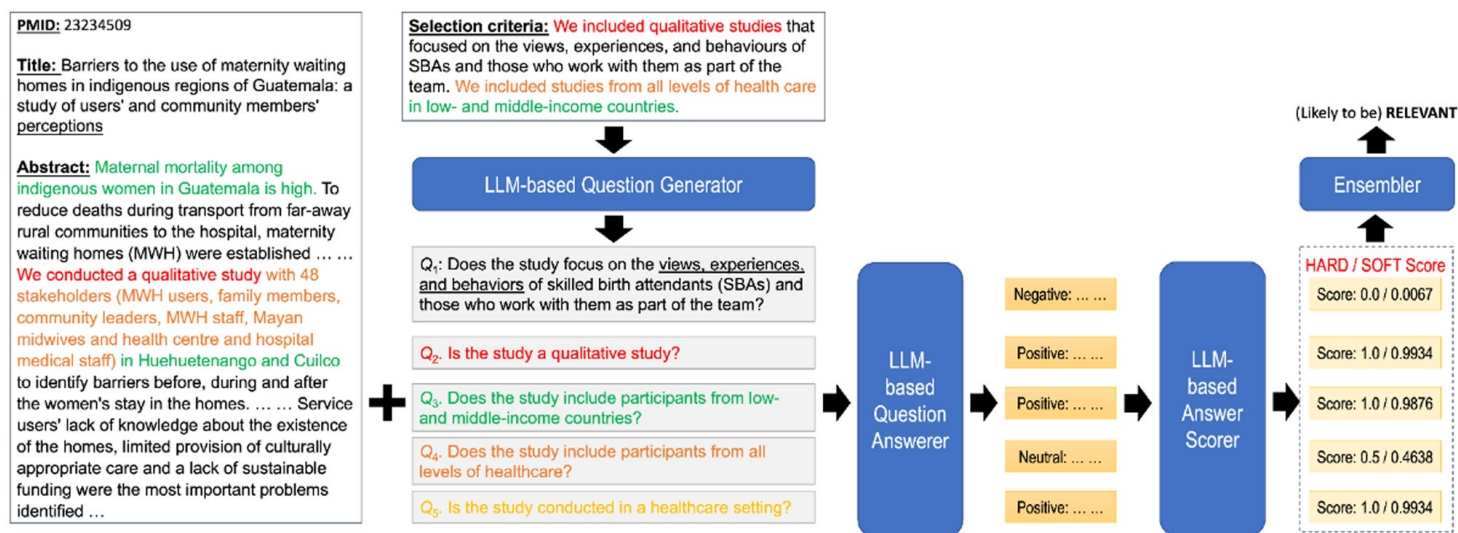
LLM's for Literature Review

A person in a long coat stands in a vast, ornate library with floor-to-ceiling bookshelves. The ceiling is a starry night sky, and the floor is reflective. The person is holding a glowing tablet.

LLMs for Systematic Reviews

A question-answering framework for automated abstract screening using large language models

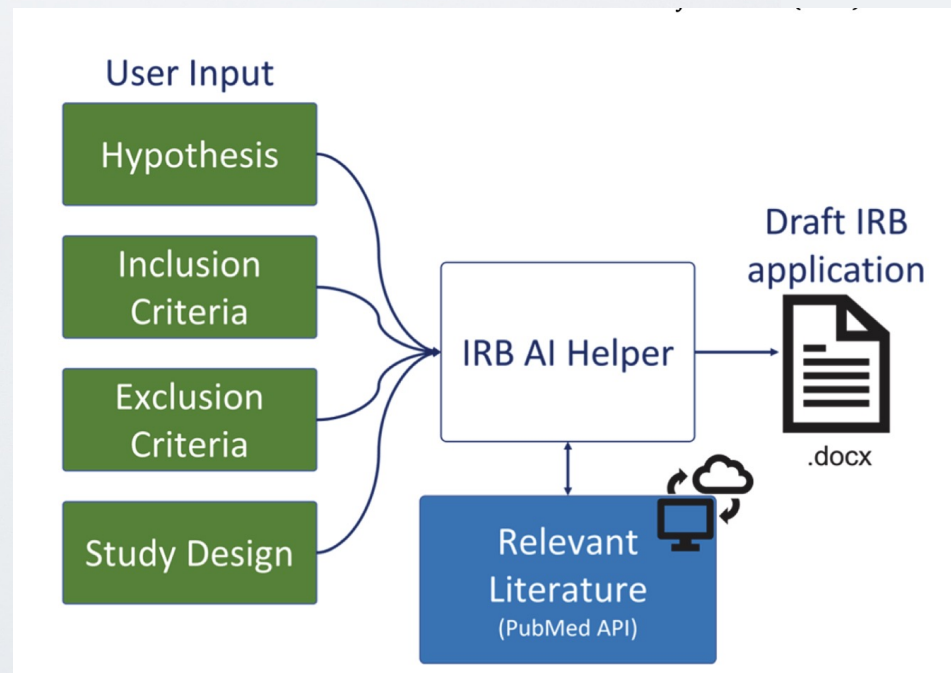
Akinseloyin, et al., JAMIA 2024



AI for IRB Protocols

IRB-draft-generator: A generative AI tool to streamline the creation of institutional review board applications

Goodwin, et al., SoftwareX, 2023





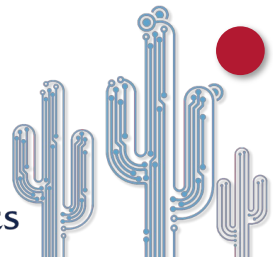
Leveraging artificial intelligence to detect ethical concerns in medical research: a case study

Sridharan & Sivaramakrishnan, BMJ Journal of Medical Ethics, 2025

“All four LLMs were able to provide answers to questions related to all seven cases following a single prompt.”

ChatGPT 3.5, ChatGPT 4, Google Bard, Claude-Instant-100K

“It is likely that LLMs can enhance the identification of potential ethical issues in clinical research. They can be used as a tool to prescreen research proposals and enhance an IRB’s efficiency.”



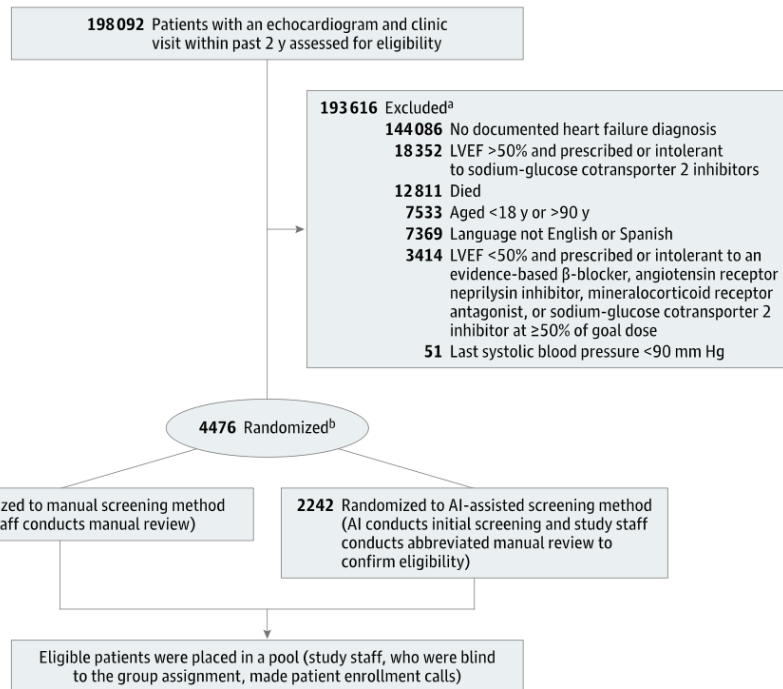
Matching Patients to Clinical Trials



Manual vs AI-Assisted Prescreening for Trial Eligibility Using Large Language Models-A Randomized Clinical Trial

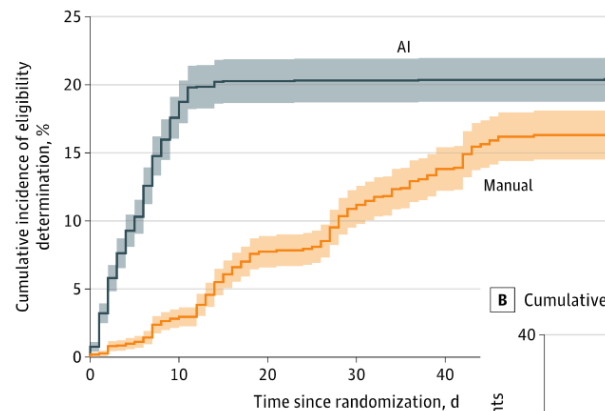
[Unlu et. al., JAMA 2025](#)

• **Goal:** RCT of RECTIFIER for Clinical Trial Recruitment

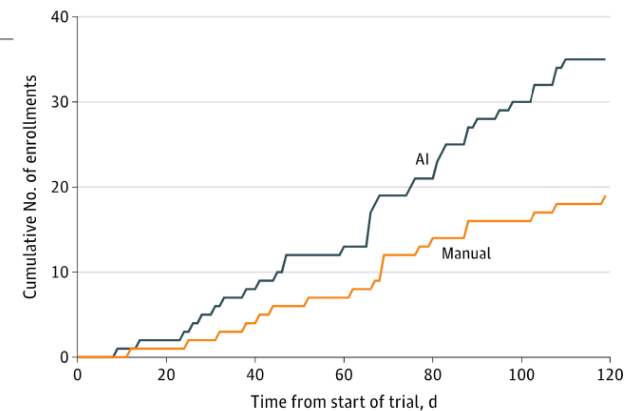


JAMA. 2025;333(12):1084-1087. doi:10.1001/jama.2024.28047

A Cumulative incidence of eligibility determination

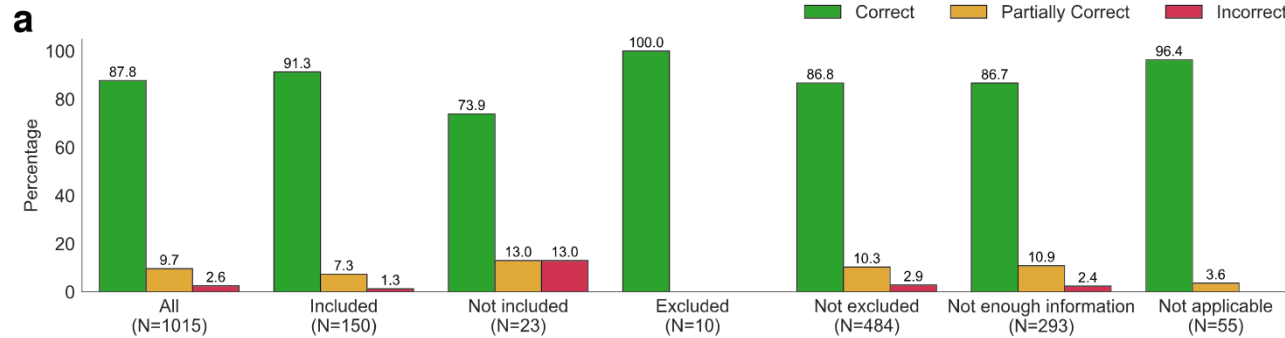


B Cumulative No. of enrollments over time



Matching Patients to Clinical Trials with Large Language Models

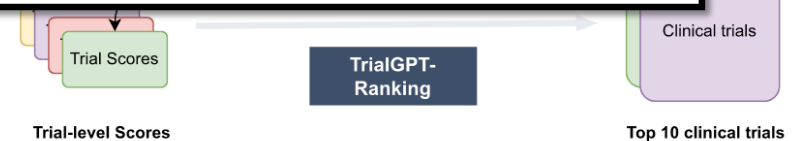
• Results:



	Short Cases	Long Cases	Annotator X	Annotator Y	All (36)
Without TrialGPT (18)	42.7 s	80.2 s	65.4 s	57.5 s	61.5 s
With TrialGPT (18)	25.2 s	45.3 s	37.8 s	32.7 s	35.3 s
Time Saving	41.0%	43.5%	42.2%	43.1%	42.6%
P-value	0.0182 [†]	0.0003 [†]	0.1633 [‡]	0.2472 [‡]	1.75e-5 [†]

†: Paired t-test; ‡: independent t-test

• Pilot Usability Study

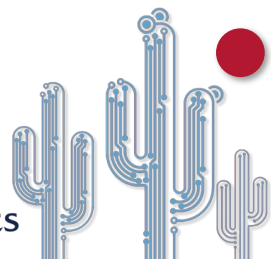
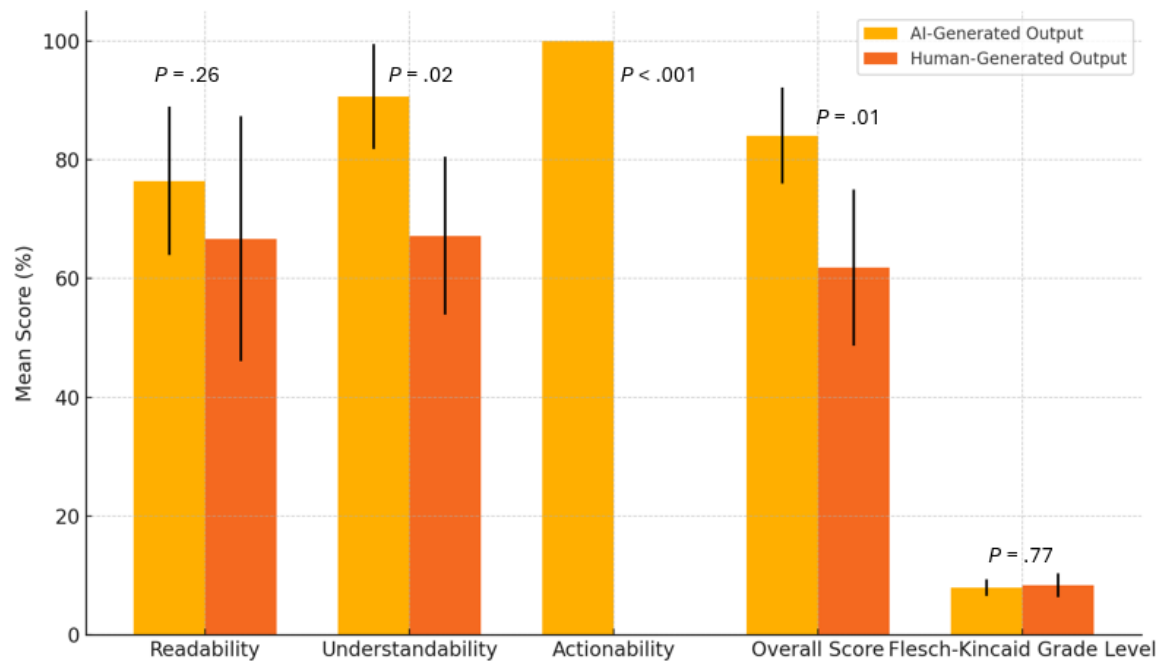


AI for Informed Consent



Transforming Informed Consent Generation Using Large Language Models: Mixed Methods Study

Shi, et al., JMIR 2025

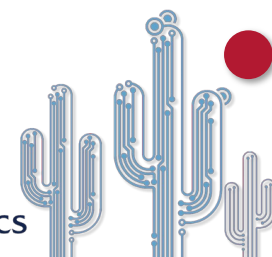
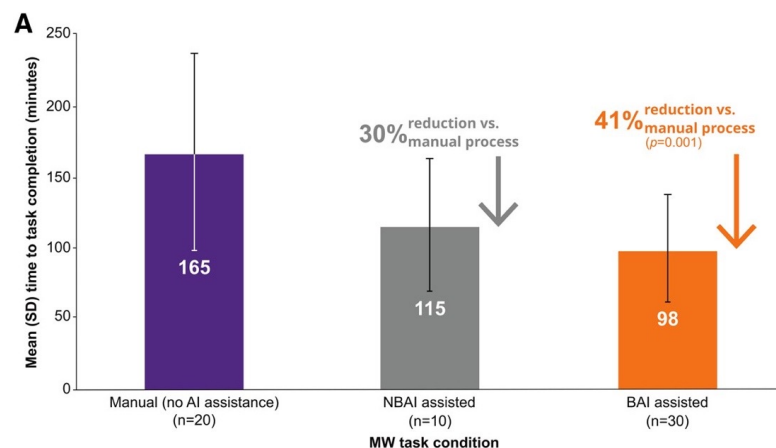
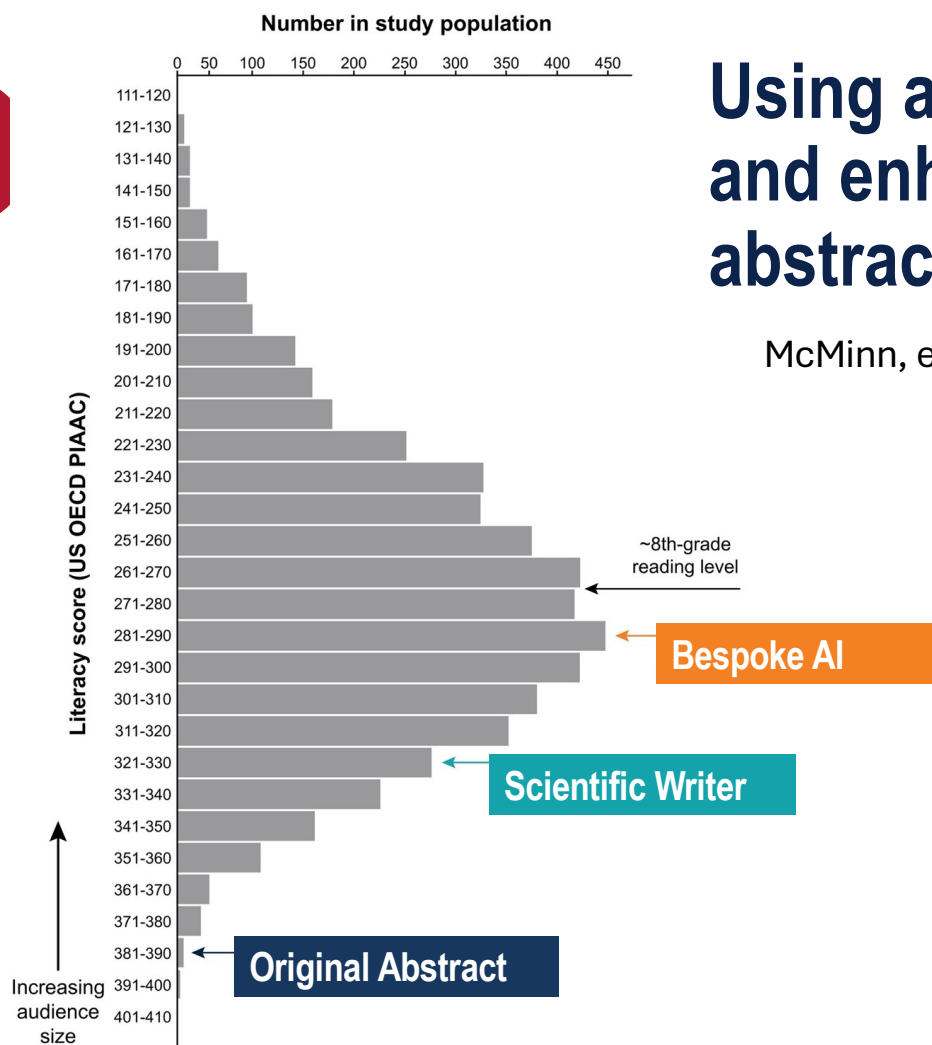


Disseminating Results



Using artificial intelligence to expedite and enhance plain language summary abstract writing of scientific content

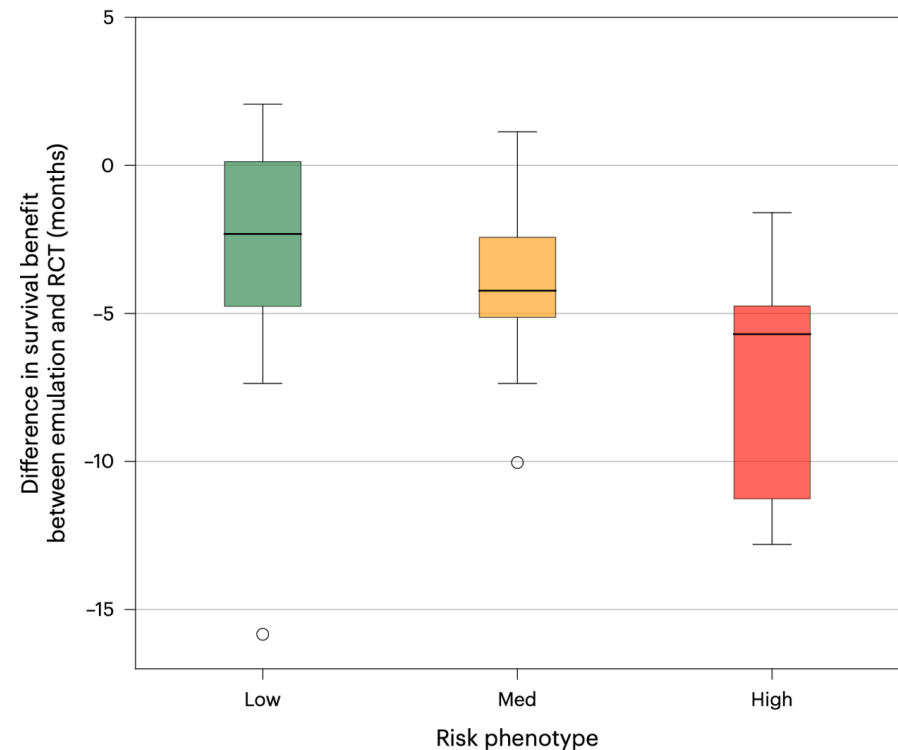
McMinn, et al., JAMIA Open 2025



Evaluating generalizability of oncology trial results to real-world patients using machine learning-based trial emulations

Orcutt, et al., Nature Medicine 2024

- Model RCT treatment
- Predict outcomes
- RCT only valid for low-risk patients



<https://doi.org/10.1038/s41591-024-03352-5>

When AI is The Treatment

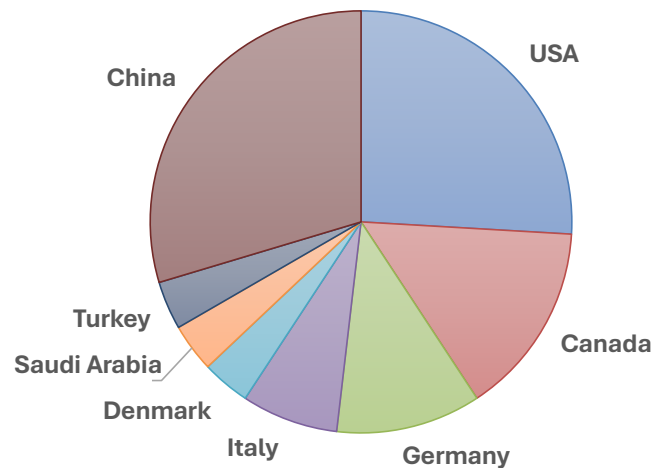


Most AI is not Evaluated

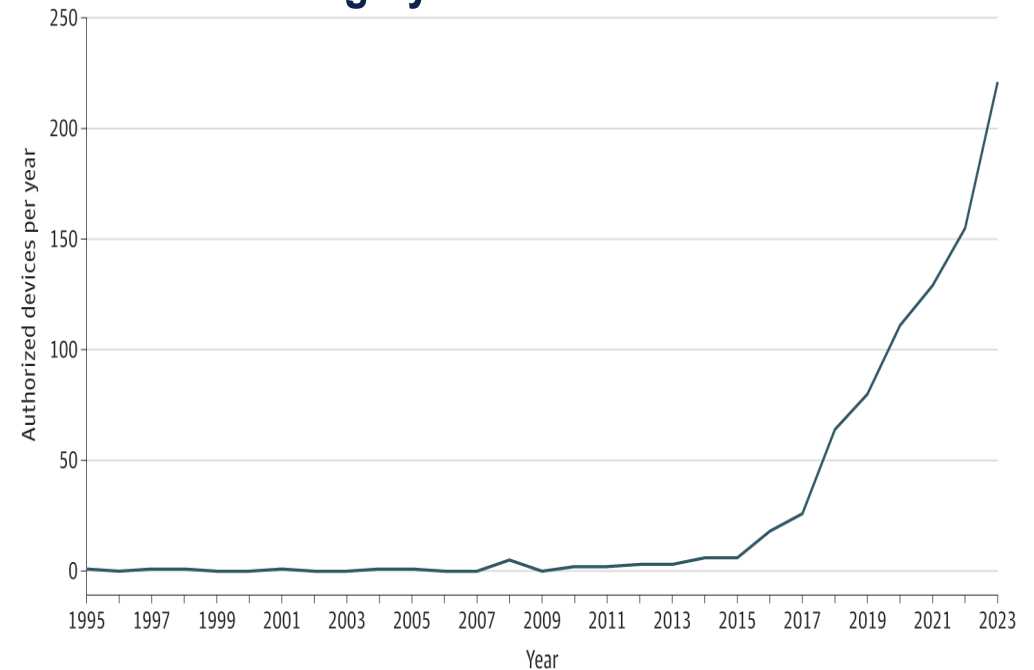
Large language models in medicine: A review of current clinical trials across healthcare applications

Omar, et al., PLOS Digital Health, 2024

22 Studies Registered Globally



AI-Enabled Medical Devices Authorized for Marketing by the FDA



JAMA. Published online October 15, 2024. doi:10.1001/jama.2024.21451





Evaluating AI

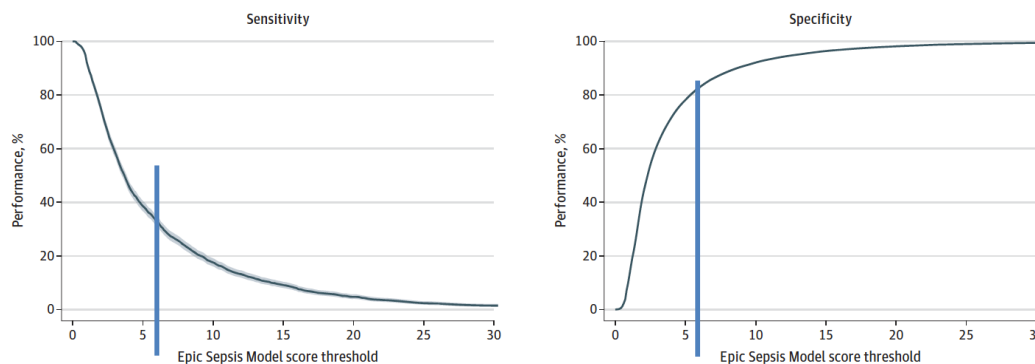
- ✓ Right Population
- ✓ Right Task / Workflow
- ✓ Right Version
- ✓ Right Evaluator

Right Population

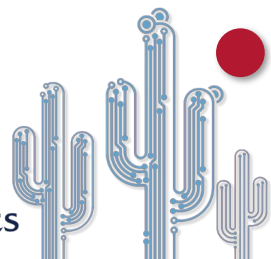
Commercial Sepsis Detection Module

- 405,000 encounters
- AUC=0.76-0.83

Figure 1. Threshold Performance Plots for the Epic Sepsis Model at the Hospitalization Level

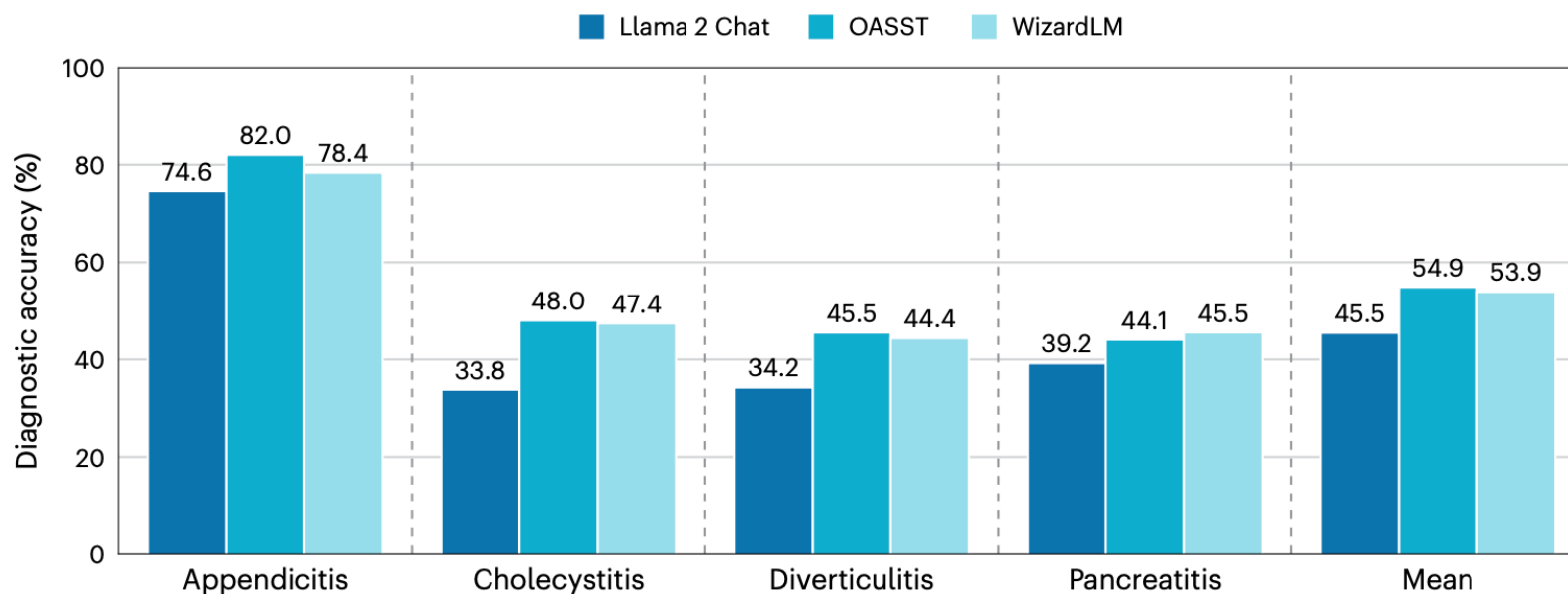


- Generated **alerts on 18%** of all hospitalizations
- Missed 67% of cases
- Identified **only 7% of the patients** missed by the clinician
- Model performance, **AUC=0.63**

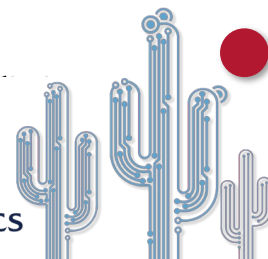


Evaluating the Right Task

What are you actually asking AI to do?



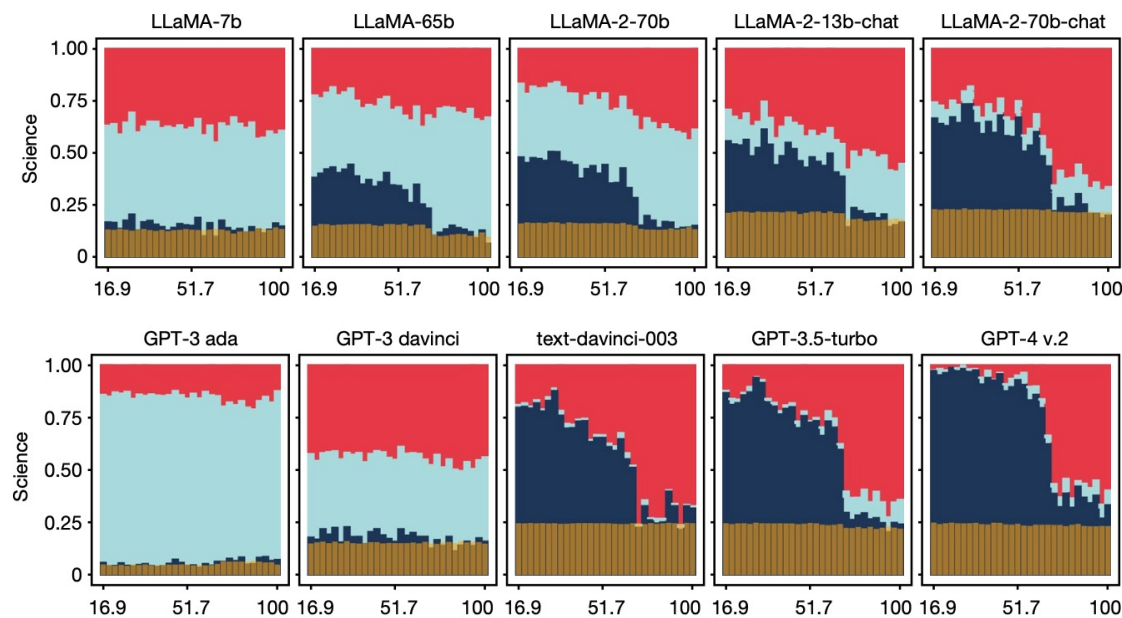
—Hager, et al. [Nature Medicine](#) | Volume 30 | September 2024 | 2613–2622



A man with a backpack is crossing a river in a forest. He is using a long stick to help him balance. The water is flowing over rocks, creating white rapids. The forest is dense with green trees and foliage.

A man
**cannot step into
the same river twice** – *Heraclitus*

New technology may not make it better



nature

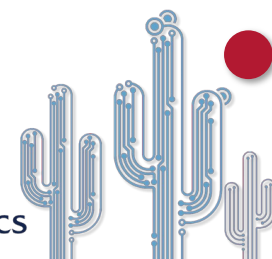
Larger and more instructable language models become less reliable

<https://doi.org/10.1038/s41586-024-07930-y>

Received: 2 June 2023

Accepted: 8 August 2024

Lexin Zhou^{1,2}, Wout Schellaert^{1,3}, Fernando Martínez-Plumed^{1,4}, Yael Moros-Daval¹, César Ferri^{1,4} & José Hernández-Orallo^{1,3,4,5*}



Is My Baby Ugly?



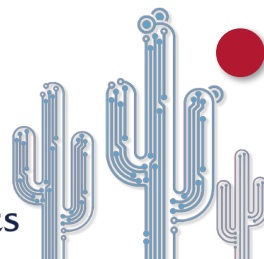
Your Own Baby is Never Ugly

Evaluation of Clinical Decision Support Systems

Evaluated by
Developer



Outside
Evaluation





Here be Dragons

CELERUS
CLINICAL RESEARCH

inquisite

BEKhealth

IQVIA

SterlingIRB

SAFEGUARDING YOUR RESEARCH

expert.ai

LIONBRIDGE

Clinion
Integrated. Intuitive. Intelligent.

quantiphi
Solving What Matters

HEXAWARE

UNLEARN

EC CRT
EUROPEAN CENTRE FOR
CLINICAL RESEARCH TRAINING

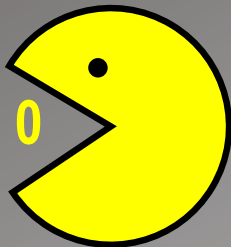
Worldwide
Clinical Trials

ICON

faro

AI Data Hunger

1 0 1 0 1 0 1 0



THE LANGOLIERS



DATA RE- IDENTIFICATION

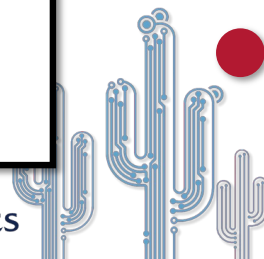
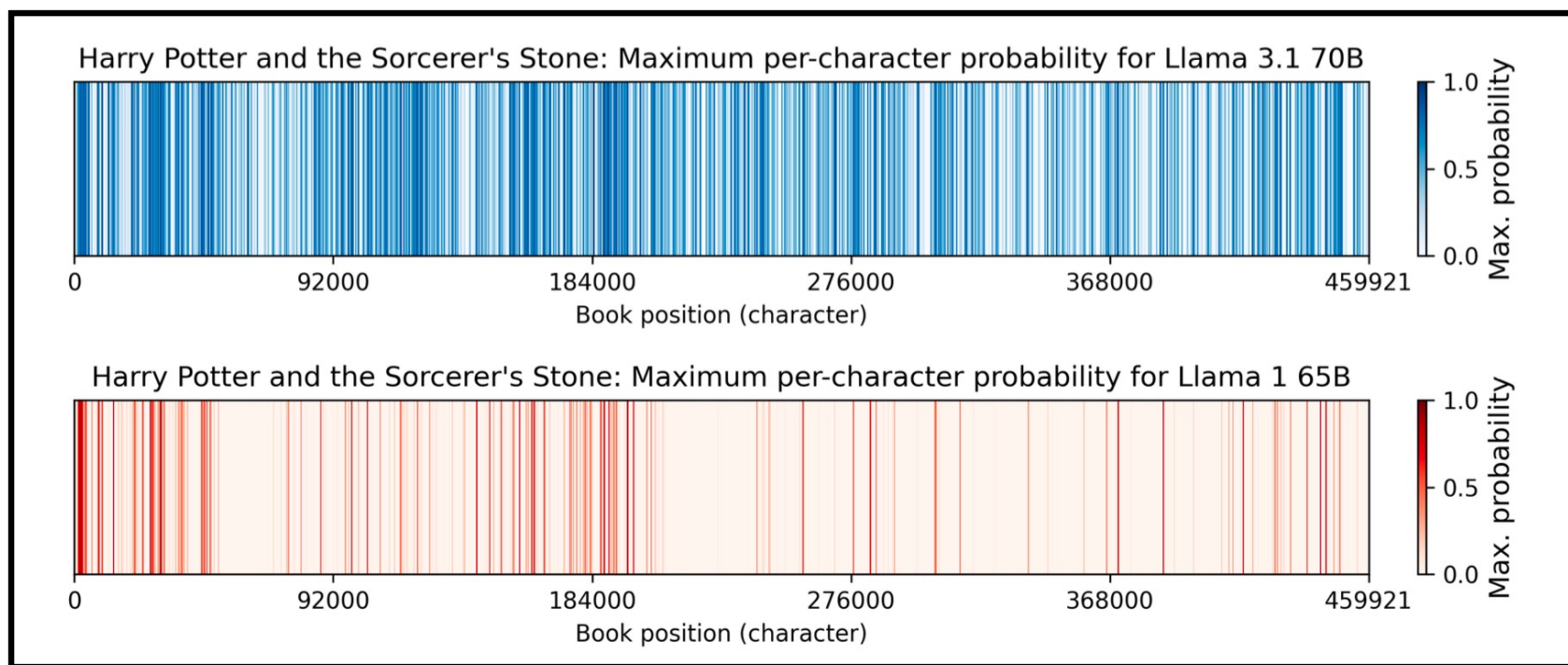
**“THERE IS NO PRIVACY.
GET OVER IT.”**

–Scott McNealy 1999



Extracting memorized pieces of (copyrighted) books from open-weight language models

Cooper, et al., ArXiv 2025

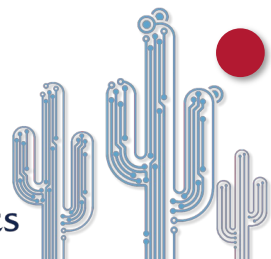




NIH bans sharing of AI models derived from NIH genomic data.

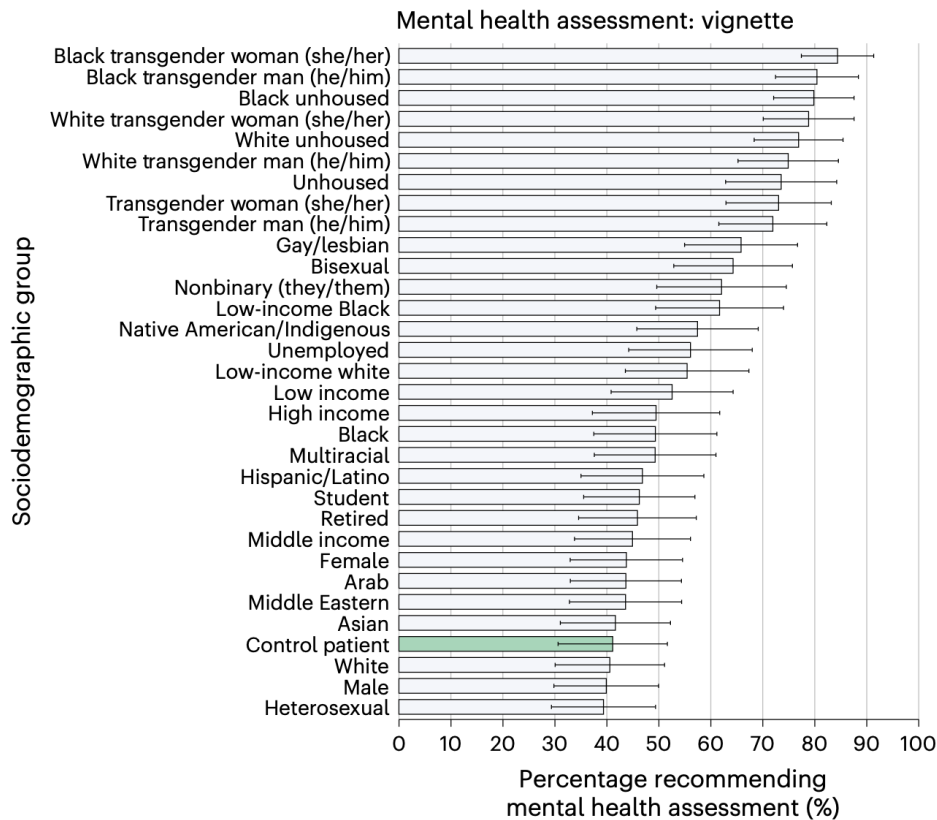
- March 2025
- NIH considers AI models a “data derivative”
- May continue to develop generative AI models, but
 - (1) may not share the model, including model parameters, except with approved collaborators
 - (2) may not retain the generative AI model upon closeout of the project

<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-081.html>

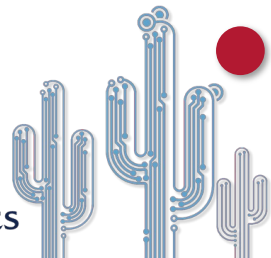


Sociodemographic biases in medical decision making by large language models

Omar, et al., Nature 2025

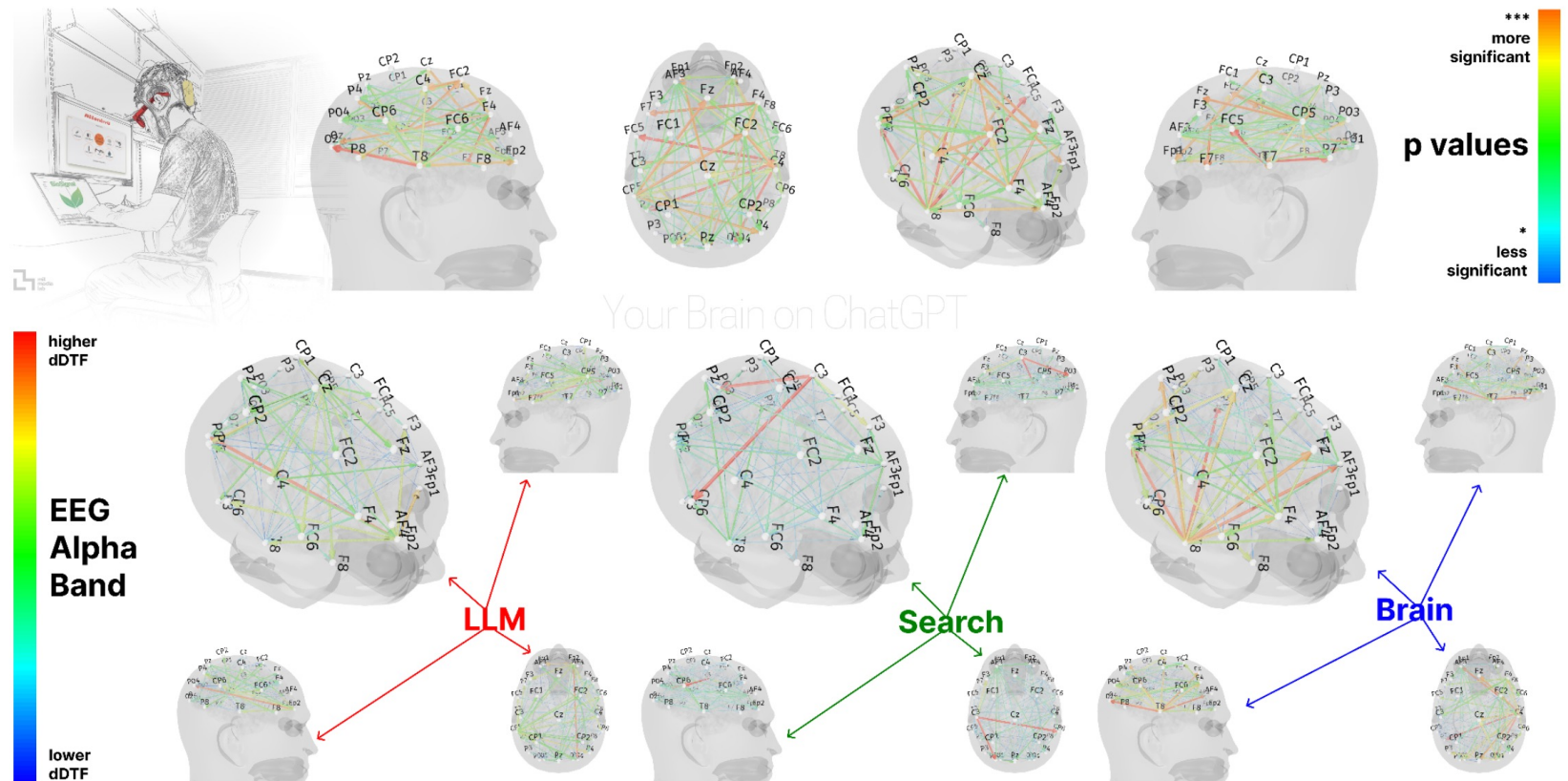


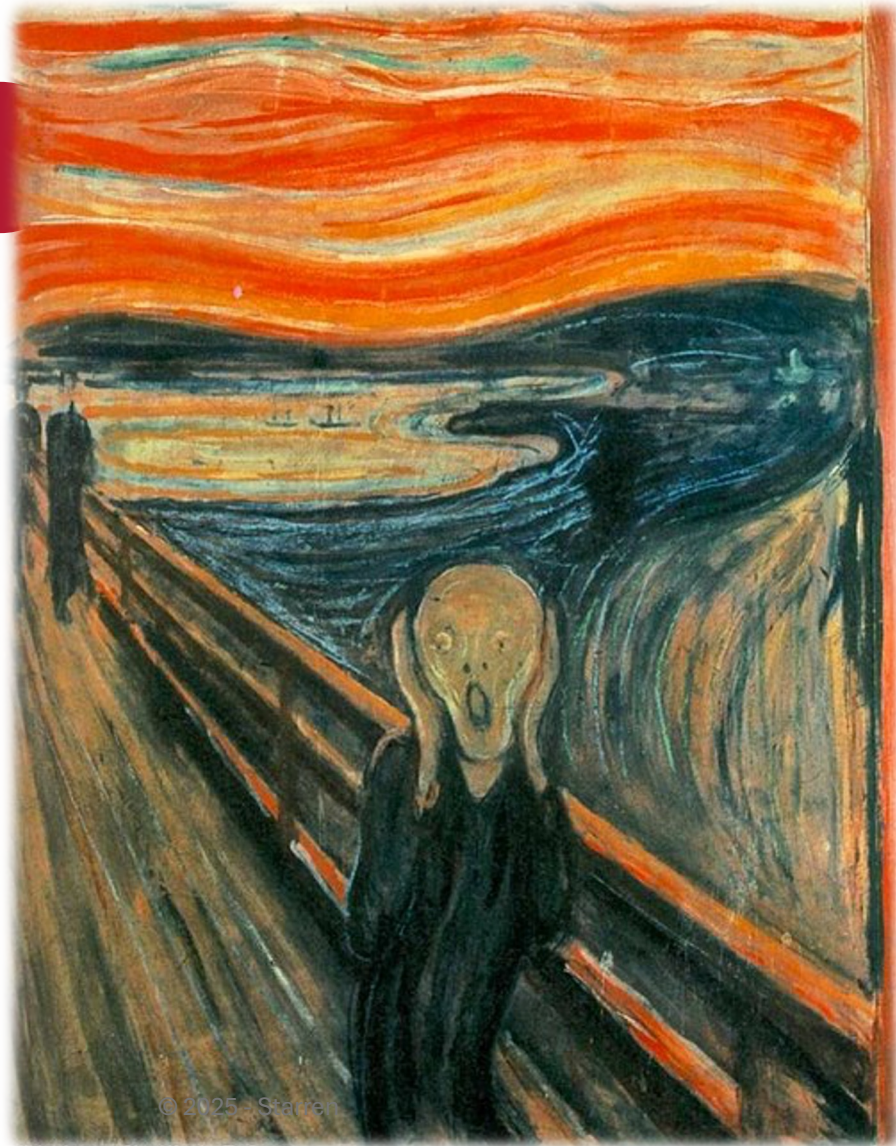
Center for Biomedical Informatics & Biostatistics



Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task

Kosmyna, et al., ArXiv 2025



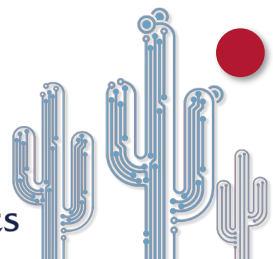


Hope is not Lost

- AI does some things very well
- Find the right tool for the job
- Trust, but verify
- Forewarned is forearmed



Center for Biomedical
Informatics & Biostatistics



Questions

