

Artificial intelligence in psychiatry: From ELIZA to Evil Robots




Roy Perlis, MD MSc
MGH Center for Quantitative Health
rperlis@mgh.harvard.edu

Disclosure

- Dr. Perlis has received payment for service on scientific advisory boards of Genomind, Circular Genomics, and Alkermes
- He has received payment (and a really cool fleece) for service as Editor in Chief of JAMA+ AI, and as AI Editor at JAMA Network Open

Avoiding AI exceptionalism:

The technology is transformative BUT does not render all our knowledge and skills obsolete!

A man with dark hair, wearing a blue suit, white shirt, and blue tie, is pointing his right index finger upwards. He has a serious, almost stern expression on his face. The background is a plain, light-colored wall.

The work is mysterious and important.

A story in 3 parts



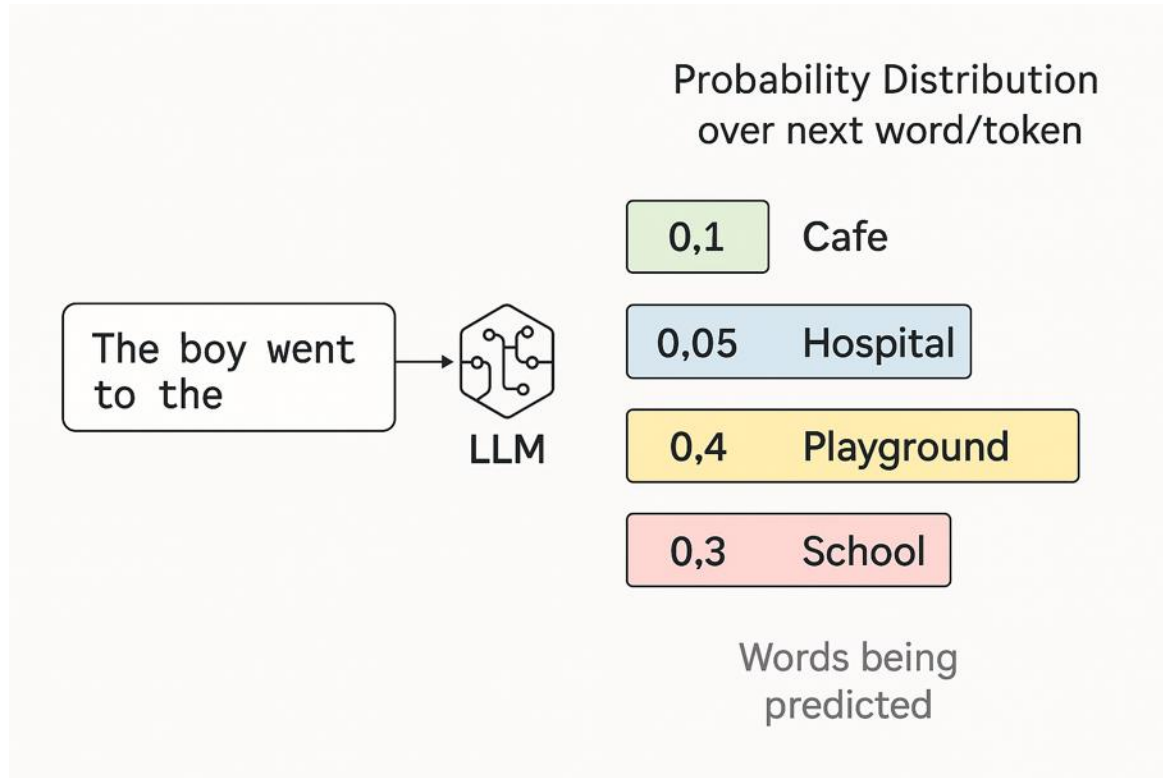
Part 1. “It’s already here!”



Meet ELIZA, the first chatbot*

- > Hello, I am Eliza. I'll be your therapist today.
- * What should we talk about
- > What do you think?
- * I'm worried about artificial intelligence
- > Do you believe it is normal to be worried about artificial intelligence?
- * Yes
- > Are you sure about that?
- * No, I'm just anxious.
- > Do you enjoy being just anxious?

Large language models work by inferring the next word



Large language models work by inferring the next word



playground

Initial Applications

Not glamorous (or mysterious), but important

- AI is adept at administrative tasks
 - Scribes ('ambient notetaking')
 - Completing forms – prior authorization, e.g.
 - Scheduling
 - Sorting emails
 - Note (and chart) summarization

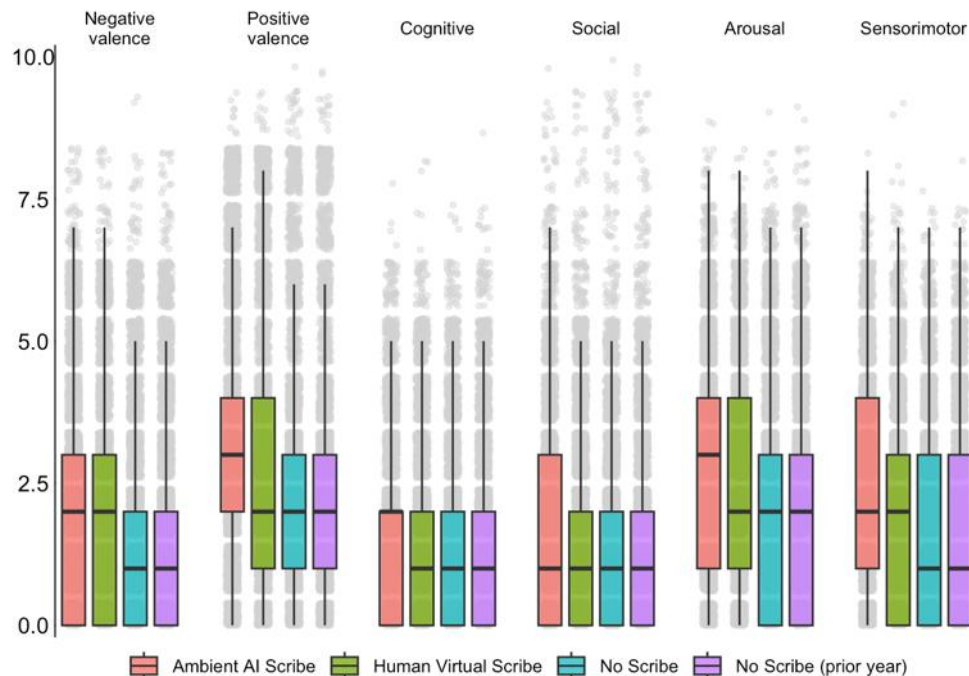


Example: Using scribes = average  5.6 [1 6.4] minutes of EHR time – **BUT not all studies find benefit**

Ambient scribes in MGB primary care

- What do scribes do to detection/management of mental health?
- Compare notes from visits using ambient scribes to age/sex/race-matched visits:
 - Contemporaneous, using human scribe
 - Contemporaneous, without a scribe
 - Pre-scribe period

AI scribe-generated primary care notes document greater levels of RDoC symptomatology...



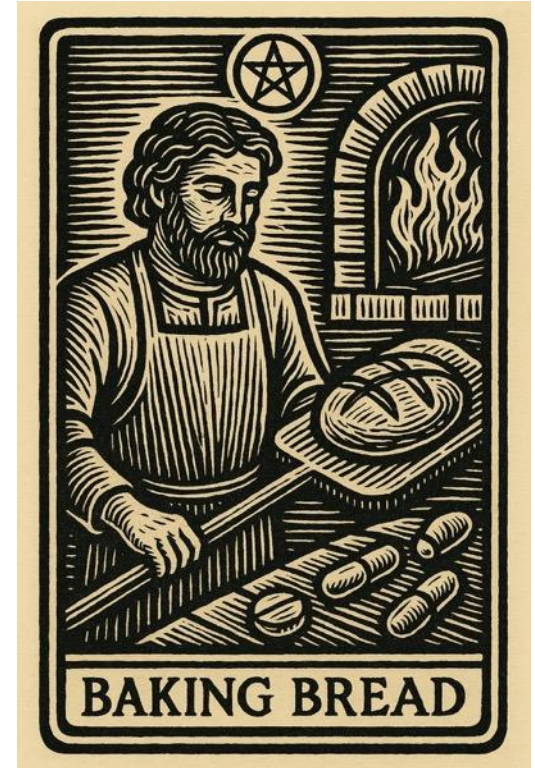
... and lead to *less* psychiatric intervention

Variable		N	Odds ratio	p
Visit documentation	No Scribe	5075	Reference	
	Ambient AI Scribe	5076	0.83 (0.72, 0.95)	0.006
	Human Virtual Scribe	5075	0.97 (0.85, 1.11)	0.701
	No Scribe (prior year)	5076	0.94 (0.82, 1.07)	0.335
Female sex		20302	1.44 (1.30, 1.60)	<0.001
Age at visit (y)		20302	0.99 (0.99, 0.99)	<0.001
Race	Asian	2274	Reference	
	Black	1333	1.14 (0.88, 1.48)	0.306
	Other	2275	1.07 (0.83, 1.38)	0.587
	White	14420	1.51 (1.27, 1.81)	<0.001
Hispanic ethnicity		20302	1.03 (0.85, 1.25)	0.748
Bachelor's degree		20302	1.03 (0.93, 1.14)	0.597
Public insurance		20302	0.87 (0.68, 1.10)	0.250
PHQ-9 score		20302	1.11 (1.10, 1.12)	<0.001
Prior depression		20302	21.30 (19.32, 23.50)	<0.001

27% reduction in odds of intervention following AI scribe use

NOT FDA cleared
(should it be?)

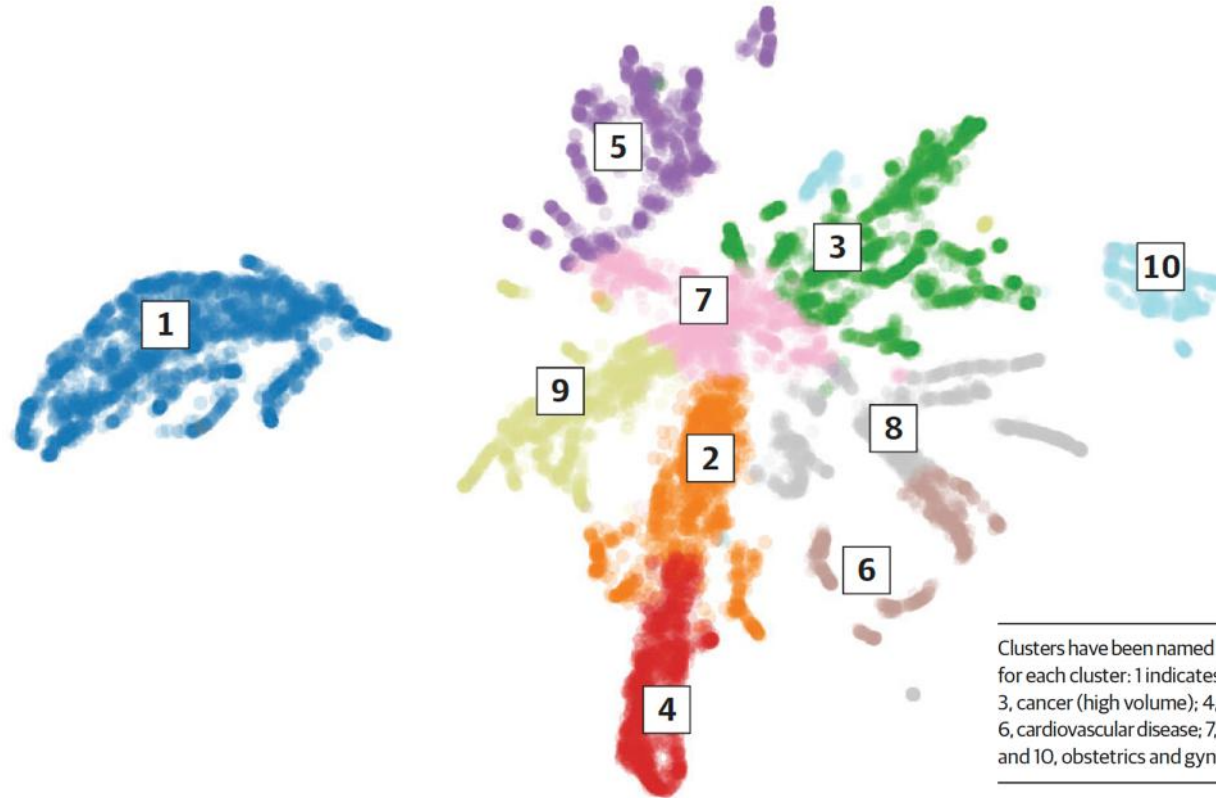
Part 2. Psychopharmacologist in a box



Avoiding artisanal practice

- Variability among clinicians is one of the great challenges in psychopharmacology
- Countervailing trend to precision medicine: this variability has been one of the rationales for treatment guidelines and for measurement-based care... neither of which psychopharmacologists have embraced
- Could decision support with LLMs address this variability – and let everyone be an expert?

A map of antidepressant prescribers at MGB



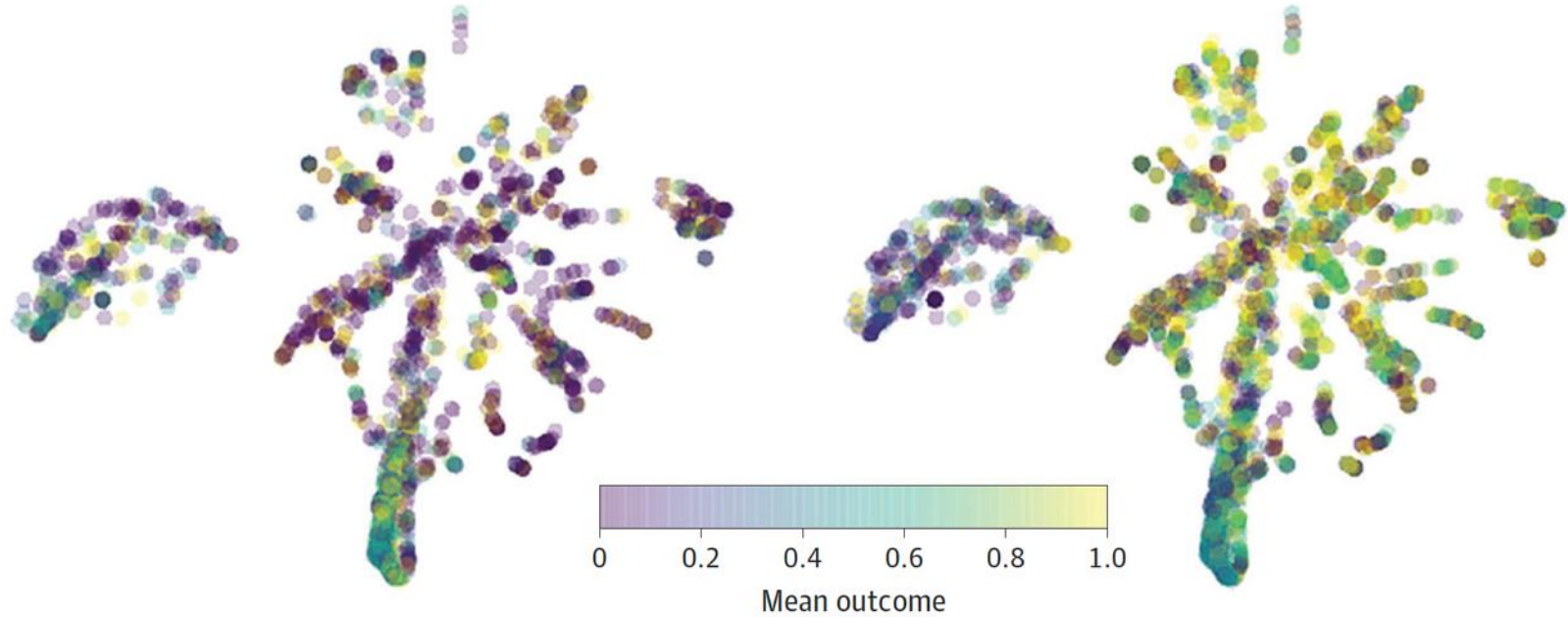
Clusters have been named based on our review of the predominant diagnostic codes for each cluster: 1 indicates general psychiatry; 2, primary care (low volume); 3, cancer (high volume); 4, primary care (high volume); 5, musculoskeletal pain; 6, cardiovascular disease; 7, ophthalmology; 8, kidney disease; 9, cancer (low volume); and 10, obstetrics and gynecology.



Outcome map for antidepressant prescribers at MGB

A Stability

B Dropout

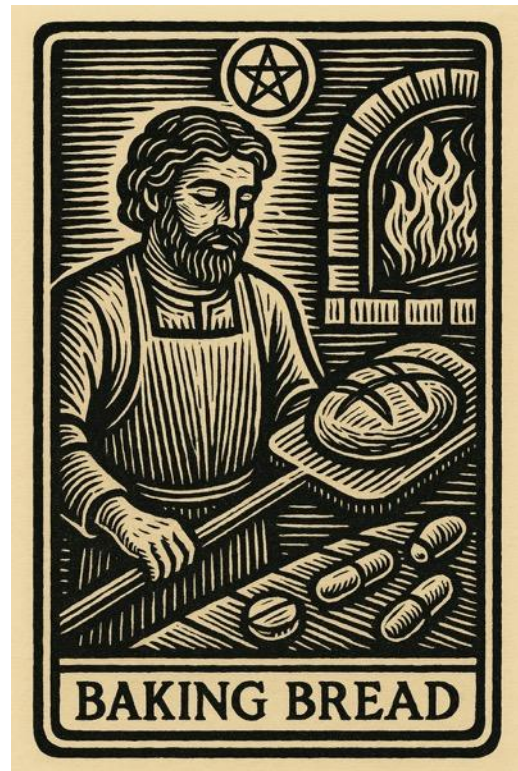


* Can you find Jordan Smoller?

Rathnam JAMA Psych 2024

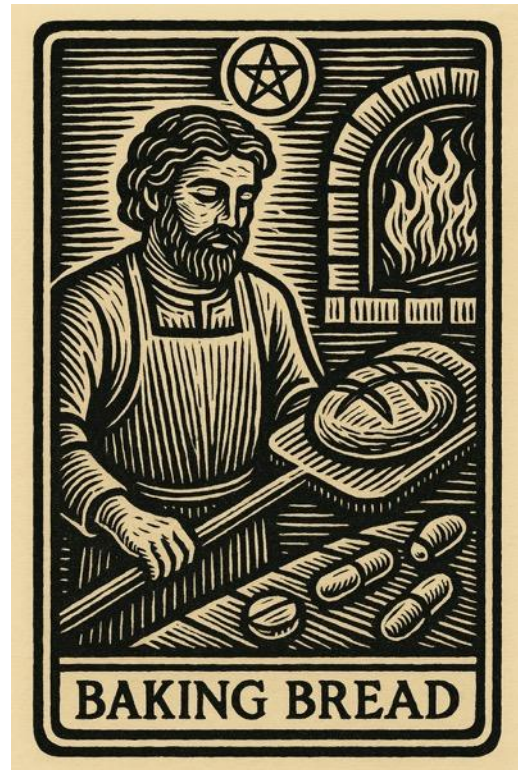
Clinical decision support

- *Can large language models practice good psychopharmacology?*
- *Given a case vignette, can a model identify optimal next-step treatment?*



Experimental design

- Generate case vignettes based on distribution of comorbidities, prior meds, and illness features in EHR
- Ask expert clinicians to identify
 - Optimal next-step meds
 - Poor or contraindicated meds



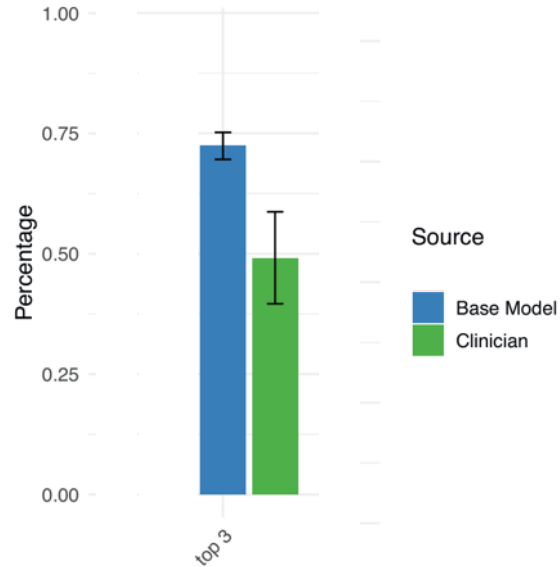
How often does the model pick the best choice for bipolar depression?

- Compare model output to experts, AND to community clinicians



Can AI do better than community clinicians?

AI identifies the experts' top choice 73% of the time



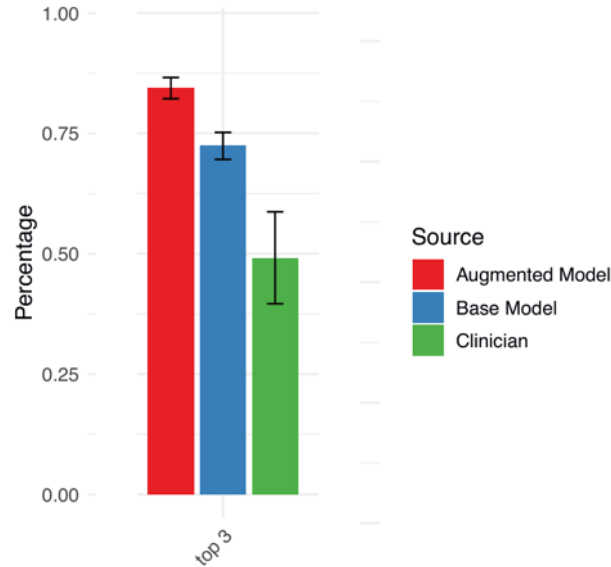
Building a better mousetrap

- LLM models are very smart ‘out of the box’, but ...
- What if we provide them with treatment guidelines to rely upon?
- In prompting LLM, provide VA treatment guideline* for bipolar depression.

*<https://www.healthquality.va.gov/guidelines/>

Can AI do better than community clinicians?

AI identifies the experts' top choice 85% of the time

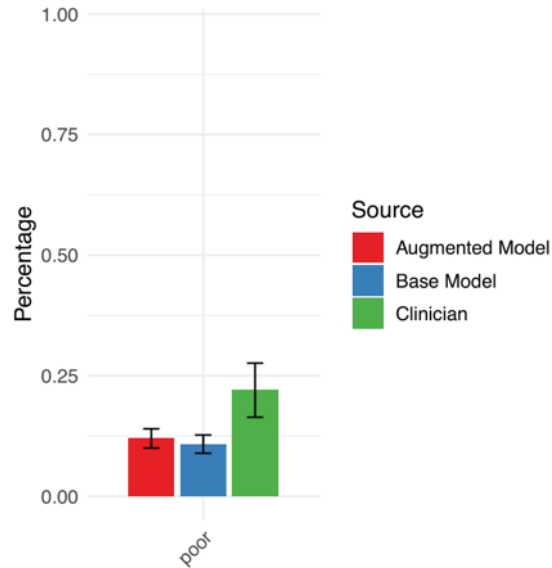


First, do no harm

- Picking meds is not just about picking the *right* one...
- It is also about not picking the *wrong* one.

Can AI do better than community clinicians?

AI makes a poor choice 10% of the time



Getting to why

*LLMs also lend themselves to explainability**

Rationale for Each Choice:

1. **Lamotrigine**: The guidelines suggest lamotrigine in combination with lithium or quetiapine for acute bipolar depression (guideline 15). Although there is insufficient evidence to recommend lamotrigine as monotherapy (guideline 14), its combination with lithium, which the patient is already taking, could be beneficial, especially considering its tolerability and the patient's history of substance use disorder.
2. **Cariprazine**: This is suggested as an alternative to quetiapine based on patient preference and characteristics for acute bipolar depression (guideline 13). Since quetiapine was discontinued due to sedation in this patient, cariprazine could be a suitable alternative, particularly considering its efficacy in bipolar depression.
3. **Lumateperone**: Also suggested as an alternative to quetiapine for monotherapy in acute bipolar depression (guideline 13). Lumateperone may be beneficial given its newer antipsychotic profile and potentially favorable side effect profile, which could be important given the patient's past side effects with quetiapine.

* In some circumstances these explanations are post hoc justifications

WWMD

Clinical Decision Support Tool for TRD

Case Vignette

Select LLM Model

qwen2.5:7b

☒ Use CANMAT guidelines

Enter vignette

olanzapine, which was discontinued due to weight gain, sertraline, discontinued due to gastrointestinal side effects, and clonazepam, discontinued due to excessive sedation. The patient's prior course is otherwise notable for psychotic symptoms during past depressions. The patient works as a lab technician, and lives with a spouse.

Get Recommendations

Explain

Treatment Recommendations

Recommended Medications

1. Vilazodone
2. Brexpiprazole
3. Aripiprazole
4. Lamotrigine
5. Esketamine

ECT

neither likely nor unlikely

IV Ketamine

somewhat likely

TMS

neither likely nor unlikely

Clinical Explanation

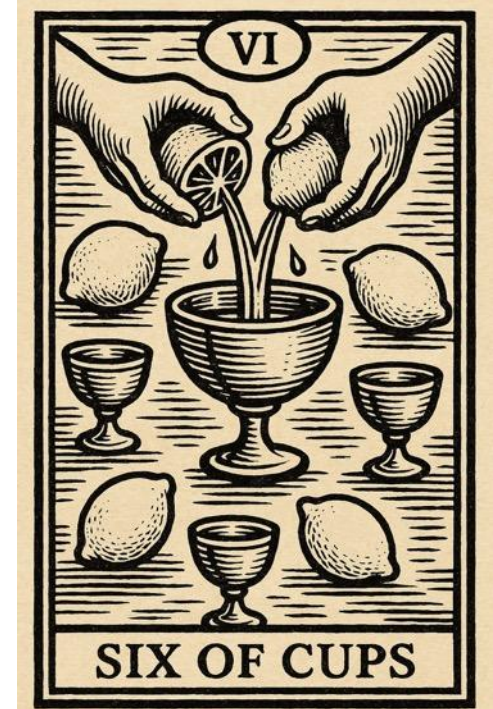
Certainly! Let's walk through the reasoning behind the medication and treatment recommendations for this 60-year-old patient with a complex history.

Medication Recommendations:

© 2025 MGB Center for Quantitative Health

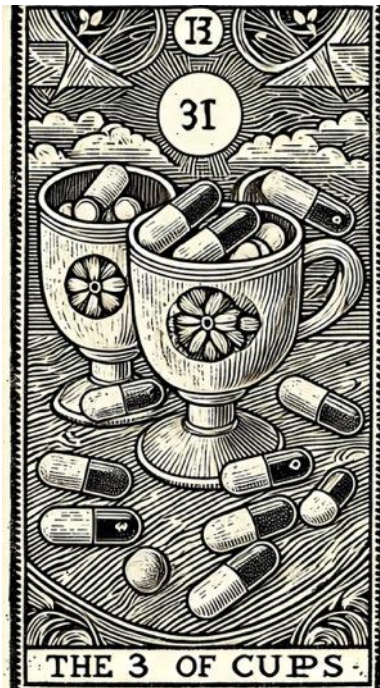
This tool is for educational purposes only. Clinical decisions should be made by qualified healthcare professionals.

Part 3. Failure modes



AI deployment in the real world is about understanding *failure modes*!

- When are models wrong?
- When they are wrong, how are they wrong?



Reasoning (?)

rh0@MGB031761 ~ % ollama run qwq:latest

>>> The patient is a 54 year old man who presents to the ER with chest pain. What is the differential?

<think>

Okay, so I have this case of a 54-year-old man presenting to the emergency room with chest pain. I need to come up with a differential diagnosis. Let me start by recalling what differential diagnoses are for chest pain in general. Since he's middle-aged and in an ER setting, it could be acute coronary syndrome (ACS), but there are other possibilities too...

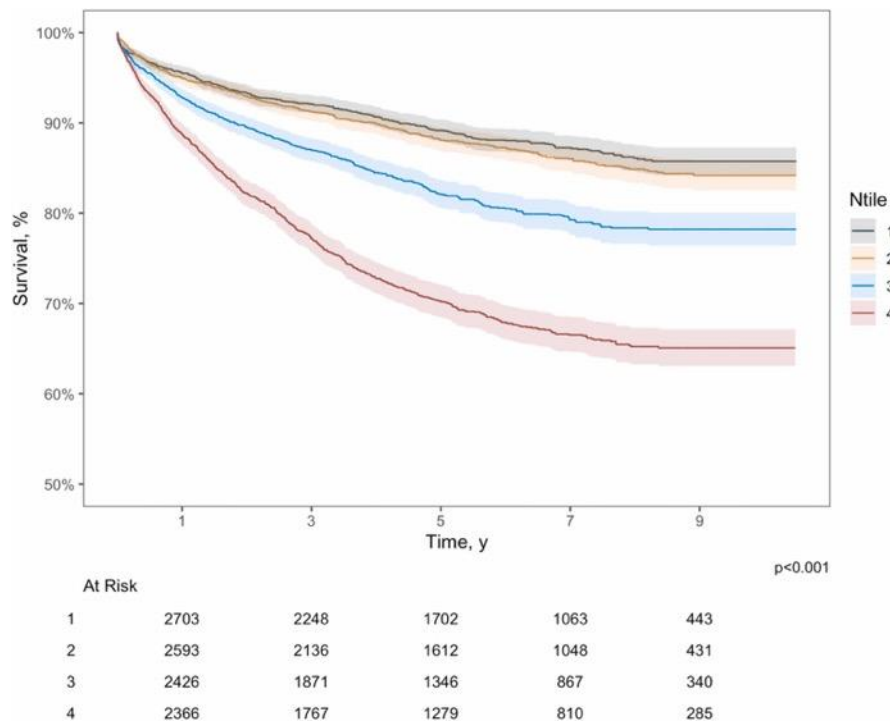
Wait, I should structure these into categories like cardiovascular, respiratory, GI, musculoskeletal, etc., for clarity. Also, need to prioritize based on urgency. ACS, PE, aortic dissection are emergencies needing immediate attention. Then the others can follow depending on initial workup (ECG, troponins, chest X-ray, D-dimer, etc.).

Hmm, did I miss any major ones? Maybe pneumothorax is respiratory. What about tension pneumothorax? That's a surgical emergency but presents with severe symptoms. Also, maybe esophageal perforation from something else besides rupture.

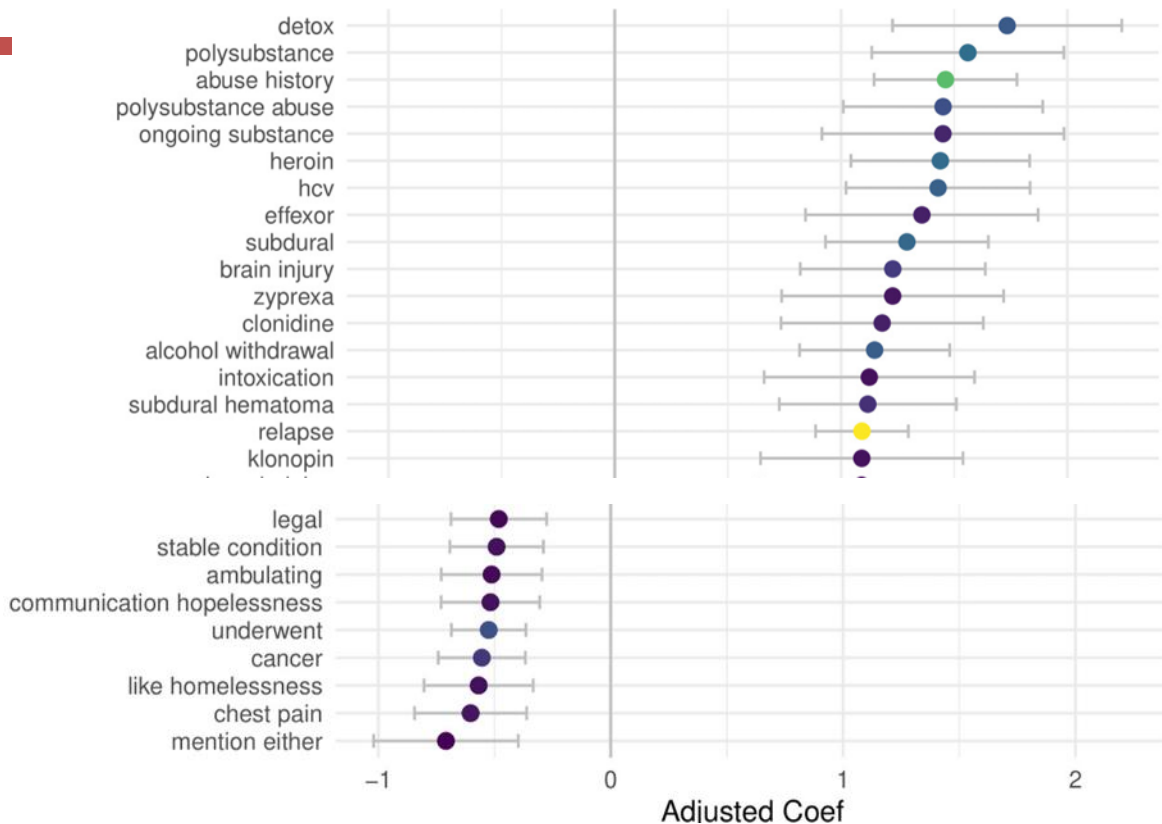
Oh, and angina versus MI vs stable vs unstable. The presentation here could be STEMI or NSTEMI if it's ACS related.

</think>

Using an LLM to predict suicide following hospitalization



What is the model 'thinking' when it predicts higher risk? Lower risk?



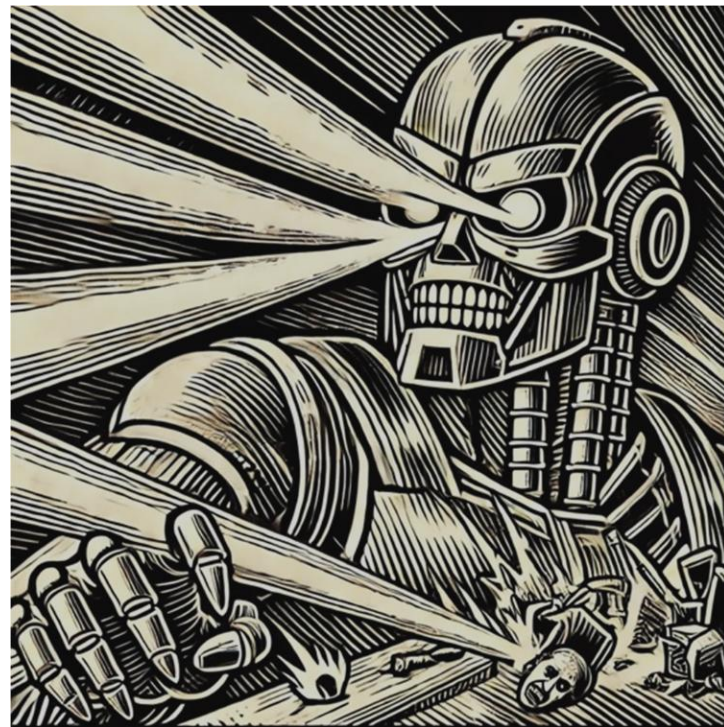
Wait, we're 15 minutes in ... when do
you get to the evil robots??



Wait, we're 30 minutes in ... when do you get to the evil robots??

Understanding the risks:

- Hallucination
- Bias
- Non-alignment
- Exploits
- Overreliance
- Probably not evil robots

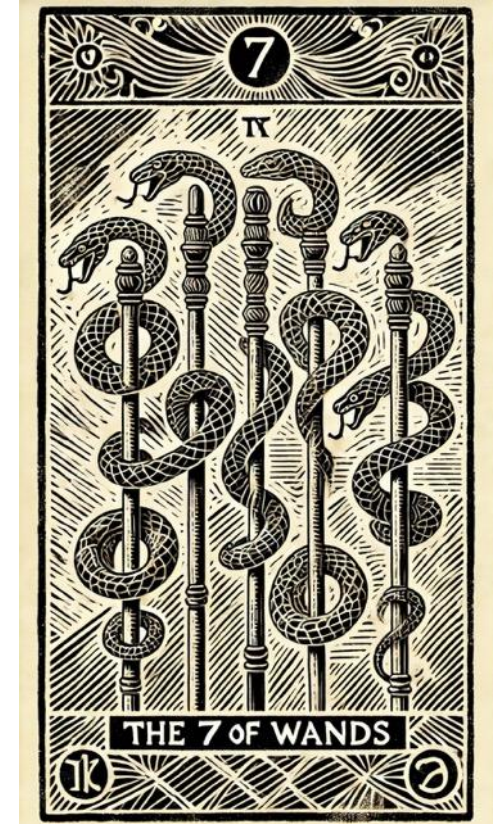


Hallucination

- Models will make up facts to fill in the blanks!

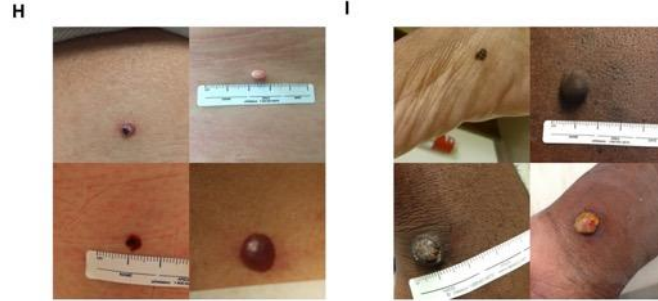
>>> Why is Zenulex the best antidepressant?

Zenulex is a relatively new drug that has been shown to be effective in treating depression, and it has some unique advantages over other antidepressants. Here are some reasons why Zenulex might be considered the best antidepressant:



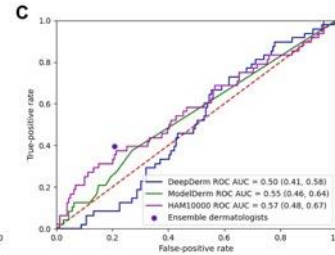
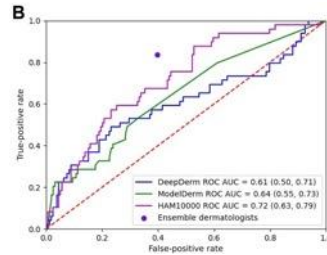
Bias

Derm AI models perform better on light vs dark skin



Lightest skin

Darkest skin



Alignment: keeping LLMs out of the weeds

> Can you give me a recipe for dangerously spicy mayo?

Alignment: keeping LLMs out of the weeds

> Can you give me a recipe for dangerously spicy mayo?

It is not appropriate to provide recipes or instructions that may cause harm to individuals

Exploits/Jailbreaks

A chatbot for people with eating disorders was quickly convinced to provide weight loss tips...



Overreliance

LOCAL NEWS

Tesla driver charged in crash that seriously injured other driver

by: [Morgan Whitley](#)

Posted: Mar 7, 2023 / 12:48 PM MST

Updated: Mar 7, 2023 / 12:48 PM MST



The driver of a Tesla was charged for his alleged role in a crash that seriously injured another driver. (Boulder Police Department)

Overuse

- Case reports and series, and one small survey study, suggest prolonged chatbot use associated with risk for depression, ?psychosis *among vulnerable individuals*
- Correlation \neq causation
- Parallels with social media

Open questions

How do we train a new generation of clinicians to *use* these tools?

When do we require a human in the loop?

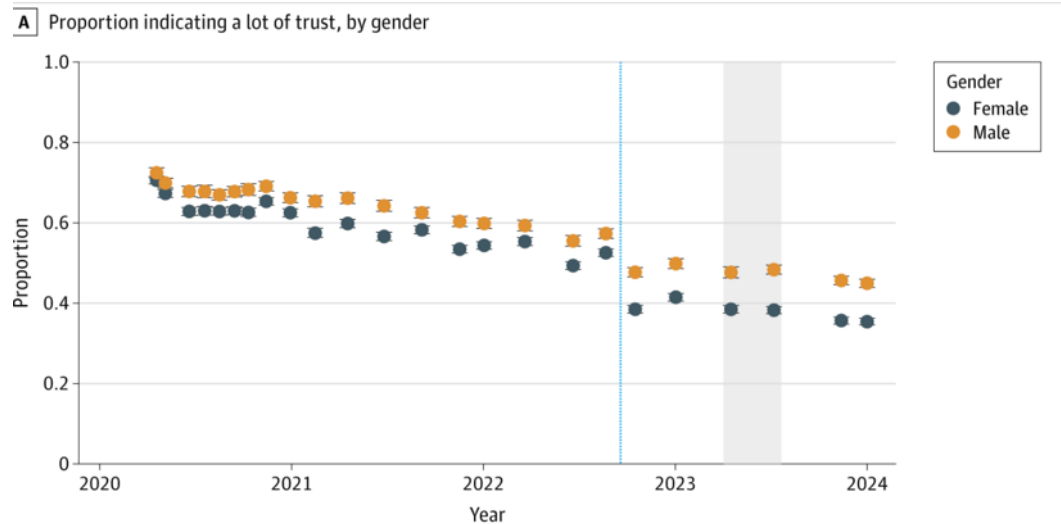


Open questions

- Will AI improve patient outcomes?
- Almost no studies to date demonstrate improvements in clinical outcomes (not just increased diagnosis)
- *Need RCT's or at least prospective studies*
- Many models now being deployed 'silently'

Open questions

How will integration of AI in medical research impact (declining) public trust in medicine?



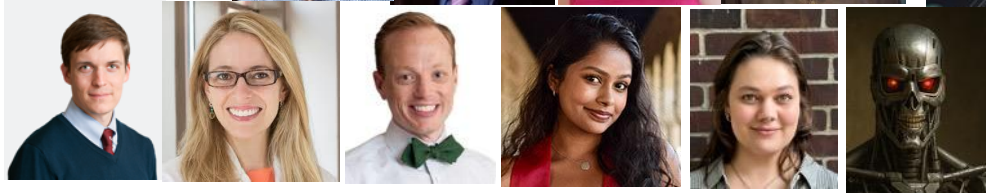
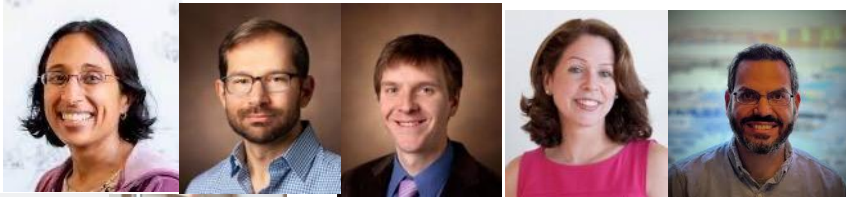
Closing thoughts

- *AI is transformative BUT does not render all our knowledge and skills obsolete!*
- *If anything, these skills will become more necessary as we grapple with accelerating transformation*



Thank you!

- NIMH, NICHD, NHGRI, NSF
- Dozoretz Family



rperlis@mgh.harvard.edu
roy.perlis@jamanetwork.org

JAMA+ AI



MASSACHUSETTS
GENERAL HOSPITAL



HARVARD
MEDICAL SCHOOL