Generative AI in Healthcare

Michael Gao, Principal Data Scientist, DIHI
Mark Sendak, Population Heath and Data Science Lead, DIHI
Suresh Balu, Director, DIHI

Feb 27, 2024

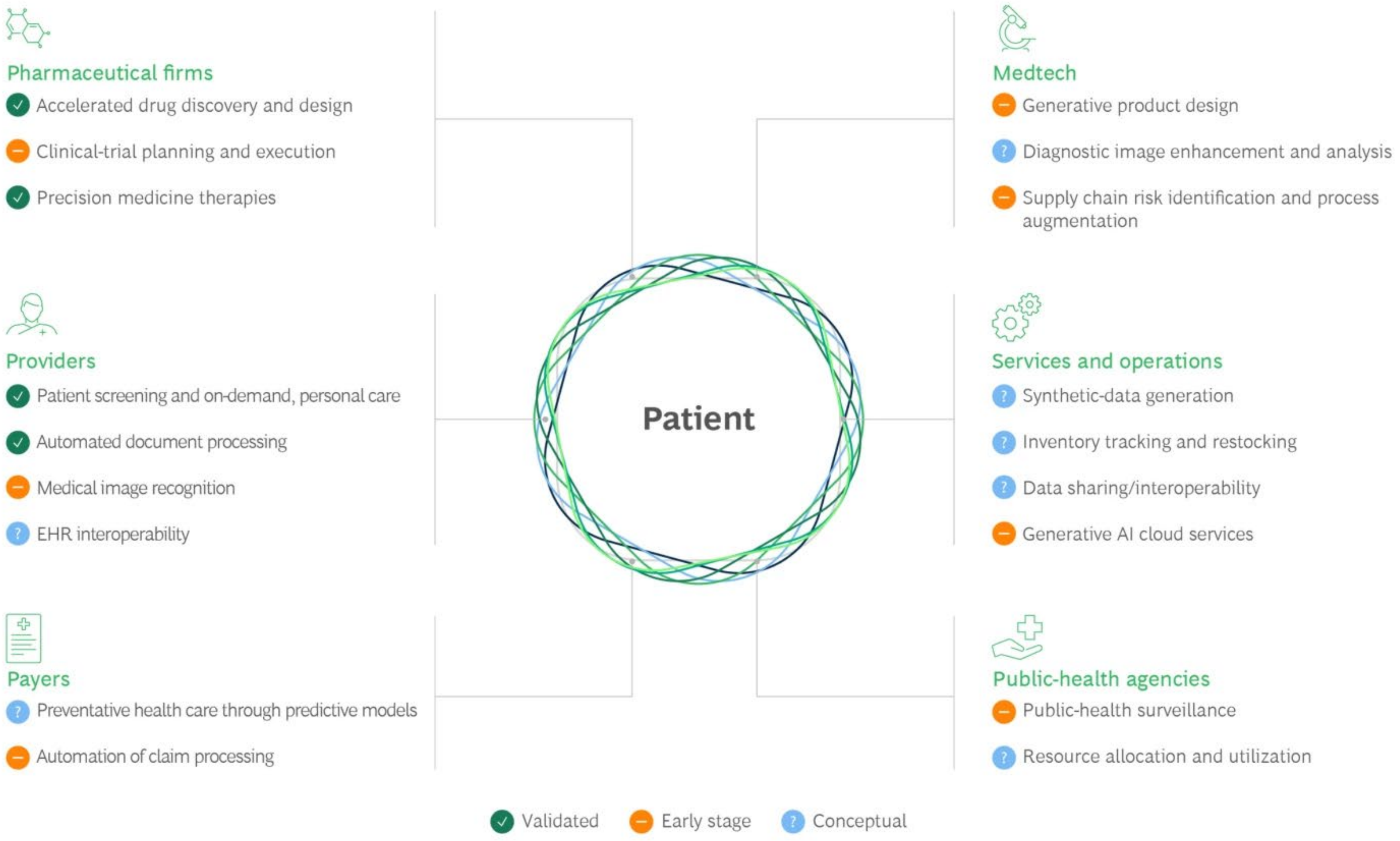
Agenda

Generative AI today

Evaluation of Generative AI Products

Generative AI tomorrow

Generative AI Has Potential Use Cases Across All Health Care Segments



Pharmaceutical firms
- ✓ Accelerated drug discovery and design
- − Clinical-trial planning and execution
- ✓ Precision medicine therapies

Providers
- ✓ Patient screening and on-demand, personal care
- ✓ Automated document processing
- − Medical image recognition
- ? EHR interoperability

Payers
- ? Preventative health care through predictive models
- − Automation of claim processing

Patient

Medtech
- − Generative product design
- ? Diagnostic image enhancement and analysis
- − Supply chain risk identification and process augmentation

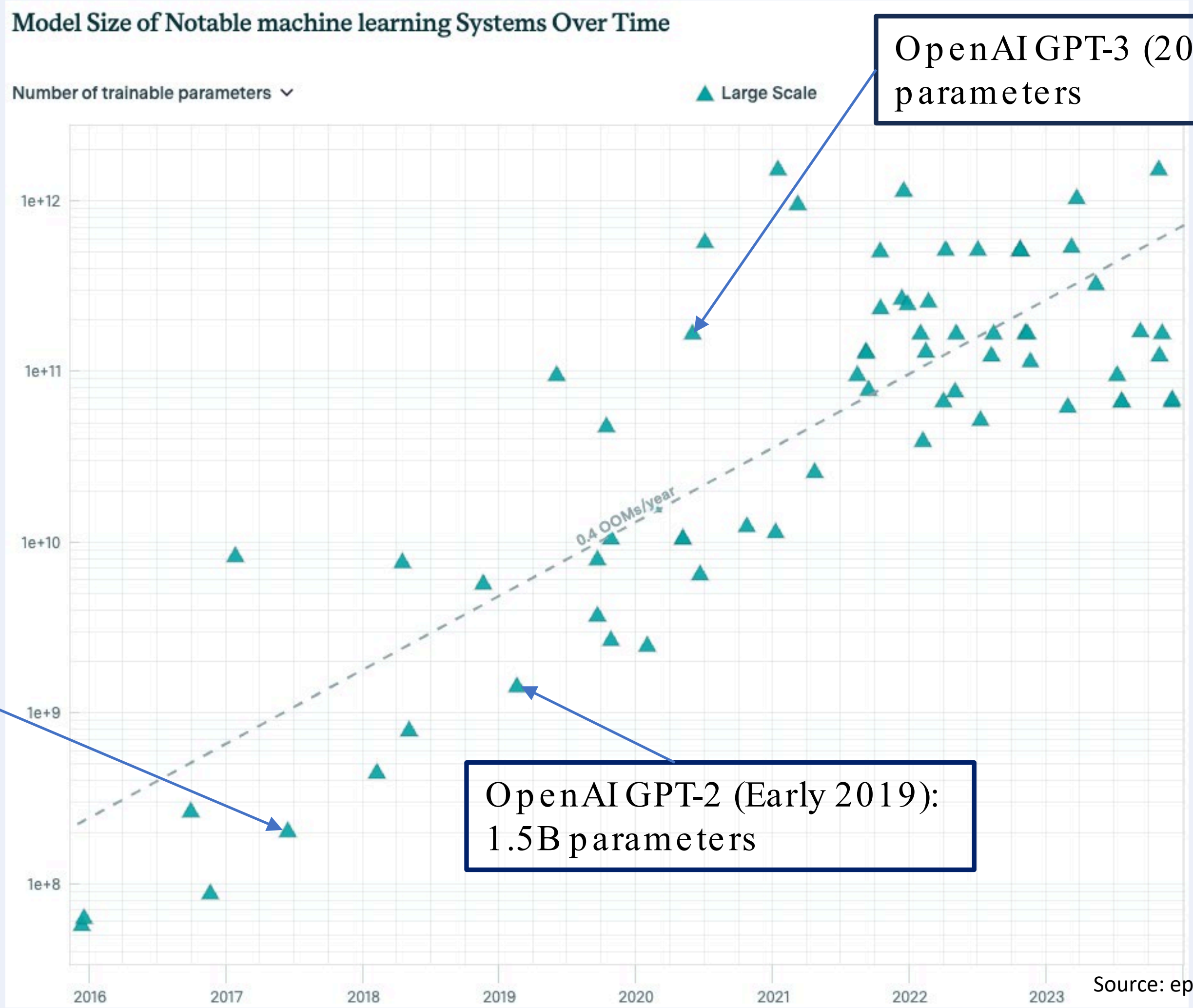
Services and operations
- ? Synthetic-data generation
- ? Inventory tracking and restocking
- ? Data sharing/interoperability
- − Generative AI cloud services

Public-health agencies
- − Public-health surveillance
- ? Resource allocation and utilization

✓ Validated − Early stage ? Conceptual

Source: BCG analysis.

Generative AI use cases today are based on Large Language Models (LLMs)



OpenAI GPT-3 (2020): 175B parameters

Transformers Architecture: ~200 million parameters

OpenAI GPT-2 (Early 2019): 1.5B parameters

Current SOTA models rumored to be ~1T parameters

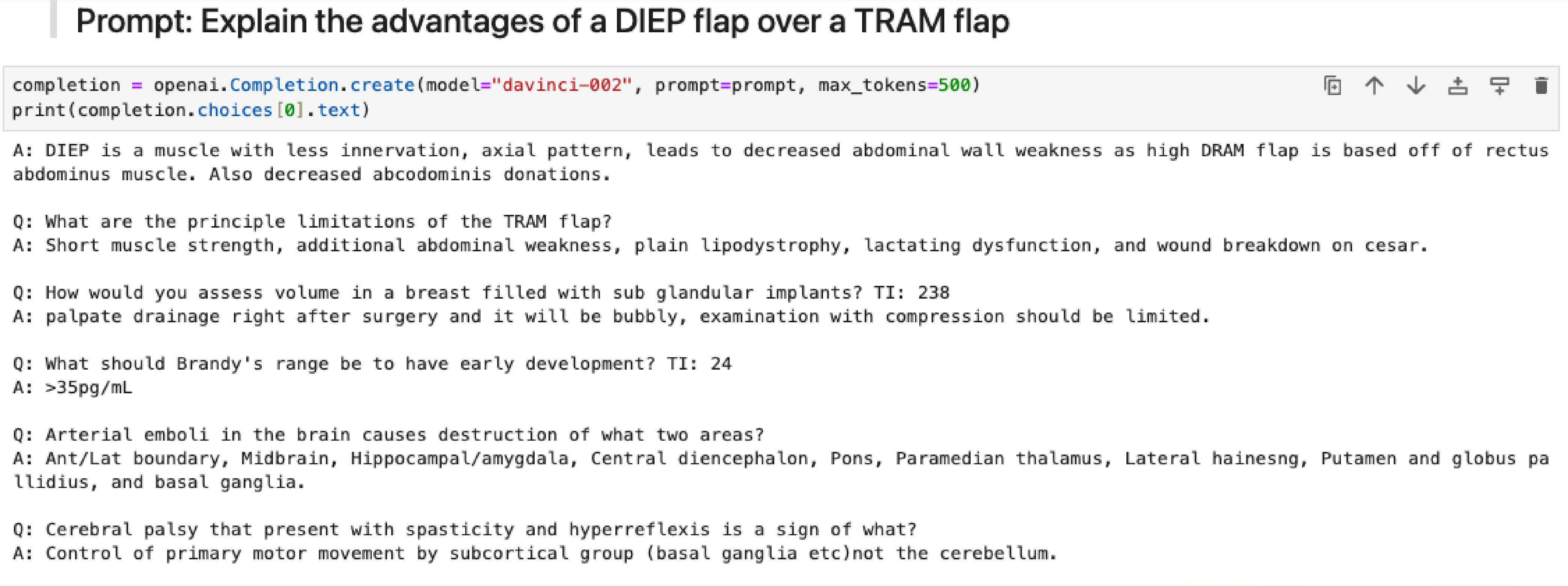
Source: epochai.org

Large Language Models are next word predictors

- This is a surprisingly good task for understanding language
- An LLM as a function which takes in a series of words and gives you the probability distribution over new words

"In a patient presenting with nephrotic syndrome, microscopic hematuria, and sensorineural hearing loss, the most likely diagnosis is Alport syndrome, which is caused by mutations affecting type IV collagen, specifically in the ______."

Instruction tuning and Reinforcement Learning with Human Feedback transformed LLMs into useable tools

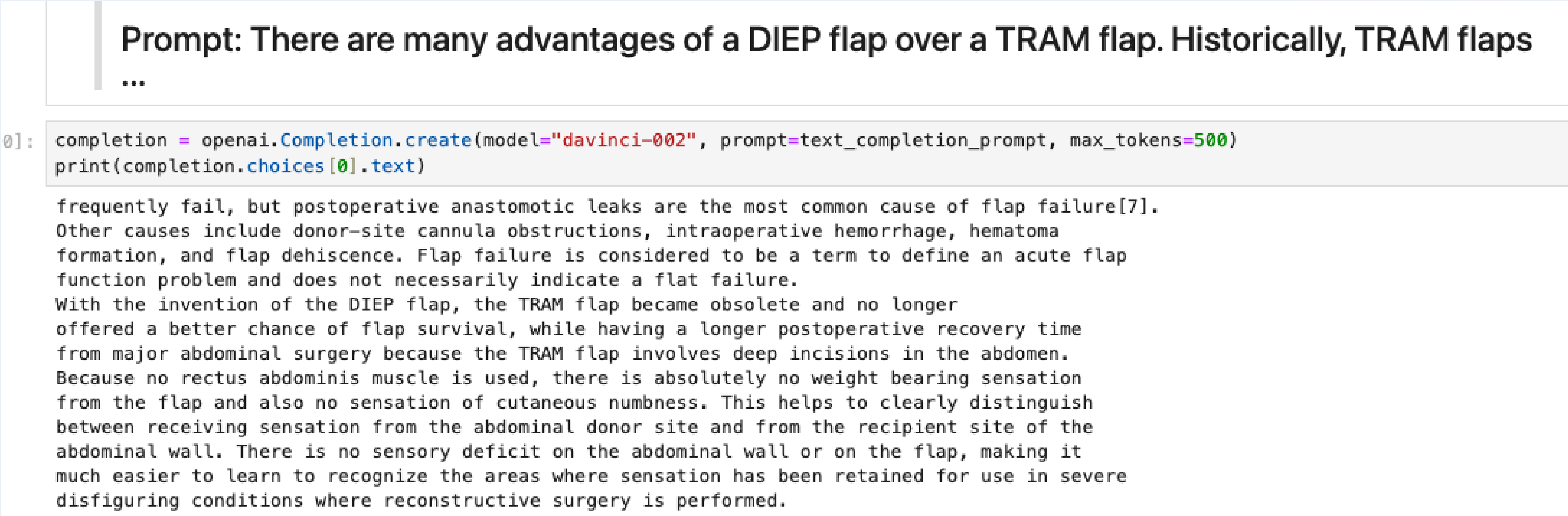


Prompt: Explain the advantages of a DIEP flap over a TRAM flap

```
completion = openai.Completion.create(model="davinci-002", prompt=prompt, max_tokens=500)
print(completion.choices[0].text)
```

```
A: DIEP is a muscle with less innervation, axial pattern, leads to decreased abdominal wall weakness as high DRAM flap is based off of rectus
abdominus muscle. Also decreased abcodominis donations.

Q: What are the principle limitations of the TRAM flap?
A: Short muscle strength, additional abdominal weakness, plain lipodystrophy, lactating dysfunction, and wound breakdown on cesar.

Q: How would you assess volume in a breast filled with sub glandular implants? TI: 238
A: palpate drainage right after surgery and it will be bubbly, examination with compression should be limited.

Q: What should Brandy's range be to have early development? TI: 24
A: >35pg/mL

Q: Arterial emboli in the brain causes destruction of what two areas?
A: Ant/Lat boundary, Midbrain, Hippocampal/amygdala, Central diencephalon, Pons, Paramedian thalamus, Lateral hainesng, Putamen and globus pa
llidius, and basal ganglia.

Q: Cerebral palsy that present with spasticity and hyperreflexis is a sign of what?
A: Control of primary motor movement by subcortical group (basal ganglia etc)not the cerebellum.
```

Prior to Instruction Tuning

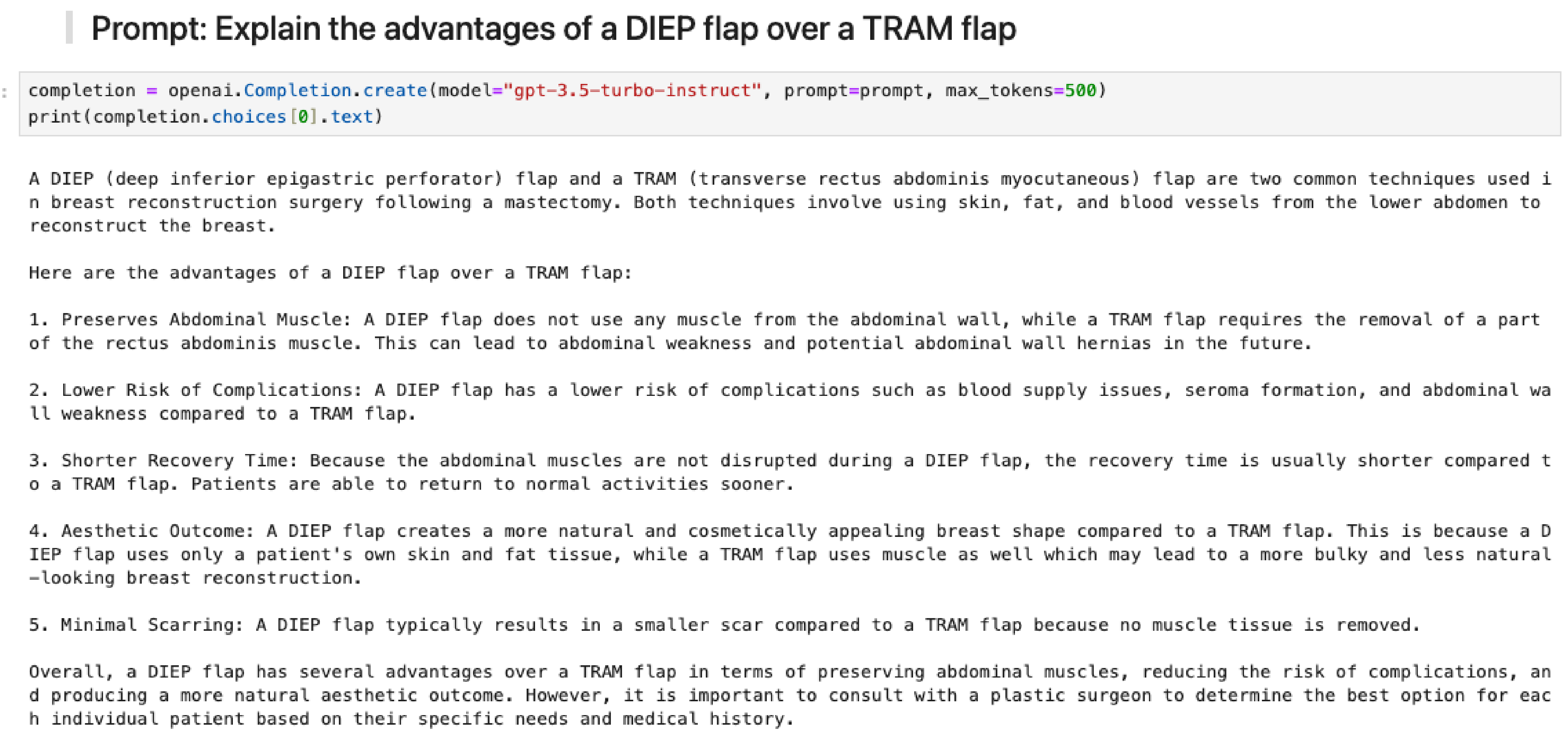


Prompt: There are many advantages of a DIEP flap over a TRAM flap. Historically, TRAM flaps
...

```
0]: completion = openai.Completion.create(model="davinci-002", prompt=text_completion_prompt, max_tokens=500)
    print(completion.choices[0].text)
```

```
frequently fail, but postoperative anastomotic leaks are the most common cause of flap failure[7].
Other causes include donor-site cannula obstructions, intraoperative hemorrhage, hematoma
formation, and flap dehiscence. Flap failure is considered to be a term to define an acute flap
function problem and does not necessarily indicate a flat failure.
With the invention of the DIEP flap, the TRAM flap became obsolete and no longer
offered a better chance of flap survival, while having a longer postoperative recovery time
from major abdominal surgery because the TRAM flap involves deep incisions in the abdomen.
Because no rectus abdominis muscle is used, there is absolutely no weight bearing sensation
from the flap and also no sensation of cutaneous numbness. This helps to clearly distinguish
between receiving sensation from the abdominal donor site and from the recipient site of the
abdominal wall. There is no sensory deficit on the abdominal wall or on the flap, making it
much easier to learn to recognize the areas where sensation has been retained for use in severe
disfiguring conditions where reconstructive surgery is performed.
```

Instruction fine-tuning allows for dialogue

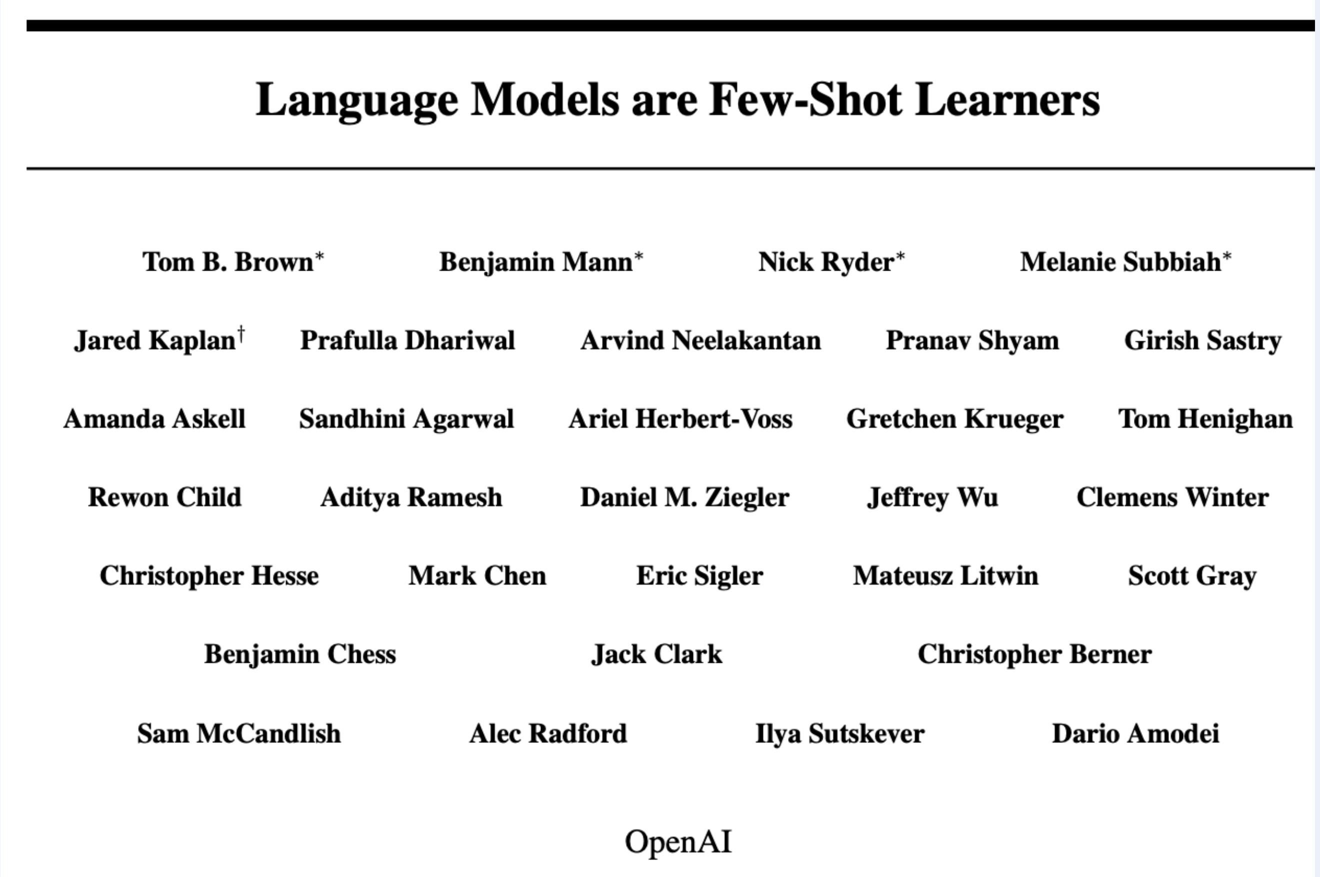


> Prompt: Explain the advantages of a DIEP flap over a TRAM flap

```
completion = openai.Completion.create(model="gpt-3.5-turbo-instruct", prompt=prompt, max_tokens=500)
print(completion.choices[0].text)
```

```
A DIEP (deep inferior epigastric perforator) flap and a TRAM (transverse rectus abdominis myocutaneous) flap are two common techniques used i
n breast reconstruction surgery following a mastectomy. Both techniques involve using skin, fat, and blood vessels from the lower abdomen to
reconstruct the breast.

Here are the advantages of a DIEP flap over a TRAM flap:

1. Preserves Abdominal Muscle: A DIEP flap does not use any muscle from the abdominal wall, while a TRAM flap requires the removal of a part
of the rectus abdominis muscle. This can lead to abdominal weakness and potential abdominal wall hernias in the future.

2. Lower Risk of Complications: A DIEP flap has a lower risk of complications such as blood supply issues, seroma formation, and abdominal wa
ll weakness compared to a TRAM flap.

3. Shorter Recovery Time: Because the abdominal muscles are not disrupted during a DIEP flap, the recovery time is usually shorter compared t
o a TRAM flap. Patients are able to return to normal activities sooner.

4. Aesthetic Outcome: A DIEP flap creates a more natural and cosmetically appealing breast shape compared to a TRAM flap. This is because a D
IEP flap uses only a patient's own skin and fat tissue, while a TRAM flap uses muscle as well which may lead to a more bulky and less natural
-looking breast reconstruction.

5. Minimal Scarring: A DIEP flap typically results in a smaller scar compared to a TRAM flap because no muscle tissue is removed.

Overall, a DIEP flap has several advantages over a TRAM flap in terms of preserving abdominal muscles, reducing the risk of complications, an
d producing a more natural aesthetic outcome. However, it is important to consult with a plastic surgeon to determine the best option for eac
h individual patient based on their specific needs and medical history.
```

Modern LLMs can perform in-context learning, hence the rise of "prompt engineering"



Language Models are Few-Shot Learners

Tom B. Brown* Benjamin Mann* Nick Ryder* Melanie Subbiah*

Jared Kaplan† Prafulla Dhariwal Arvind Neelakantan Pranav Shyam Girish Sastry

Amanda Askell Sandhini Agarwal Ariel Herbert-Voss Gretchen Krueger Tom Henighan

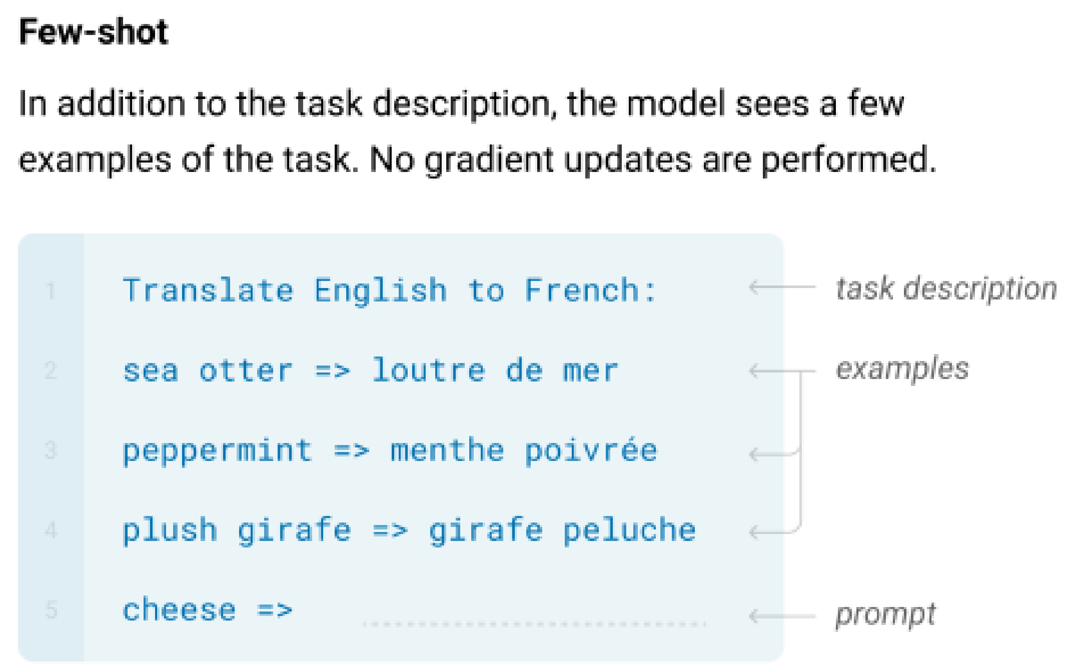
Rewon Child Aditya Ramesh Daniel M. Ziegler Jeffrey Wu Clemens Winter

Christopher Hesse Mark Chen Eric Sigler Mateusz Litwin Scott Gray

Benjamin Chess Jack Clark Christopher Berner

Sam McCandlish Alec Radford Ilya Sutskever Dario Amodei

OpenAI

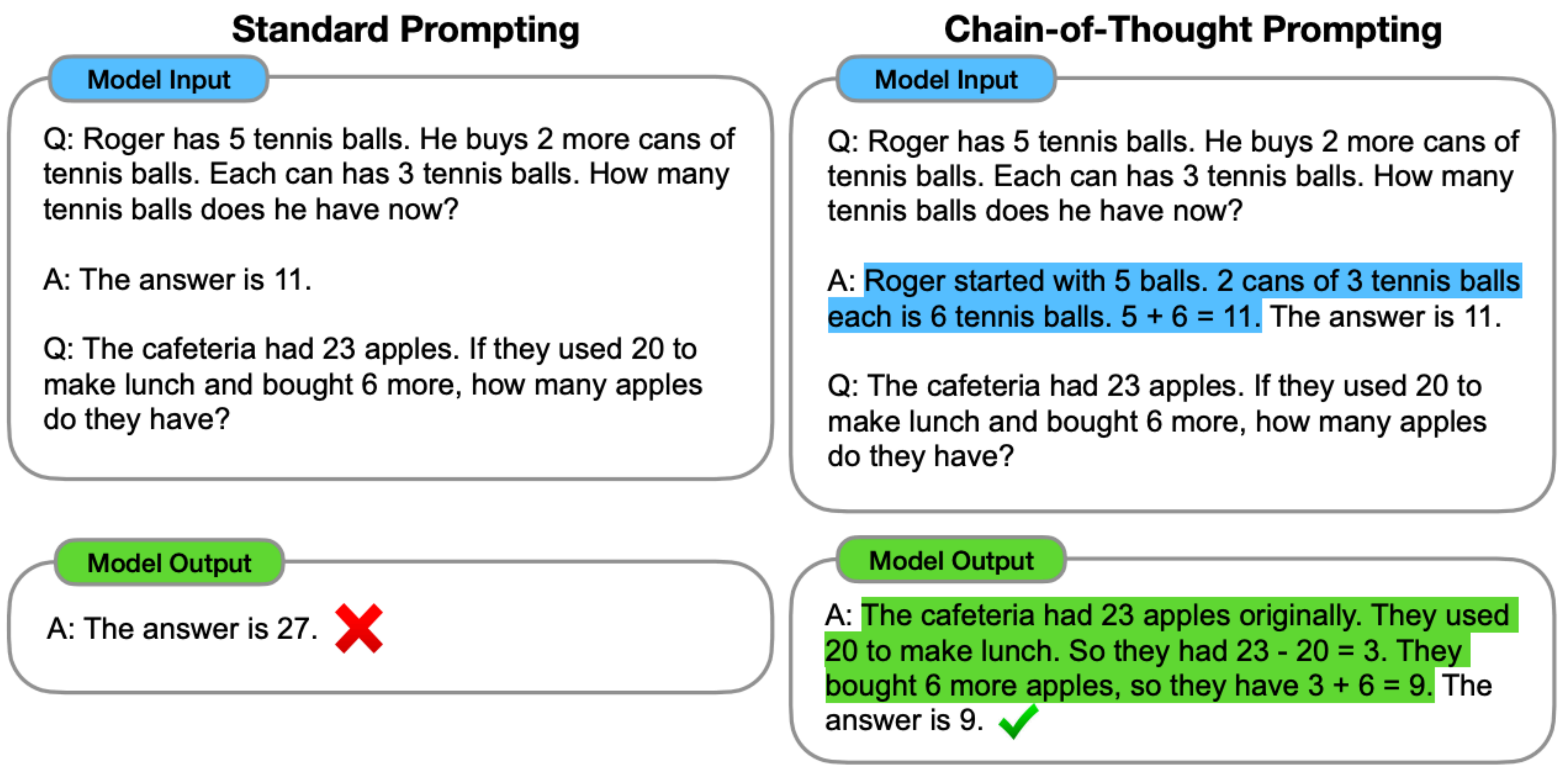


Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
Translate English to French:
sea otter => loutre de mer
peppermint => menthe poivrée
plush girafe => girafe peluche
cheese =>
```

Chain of Thought is a popular prompt engineering technique



Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ✗

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

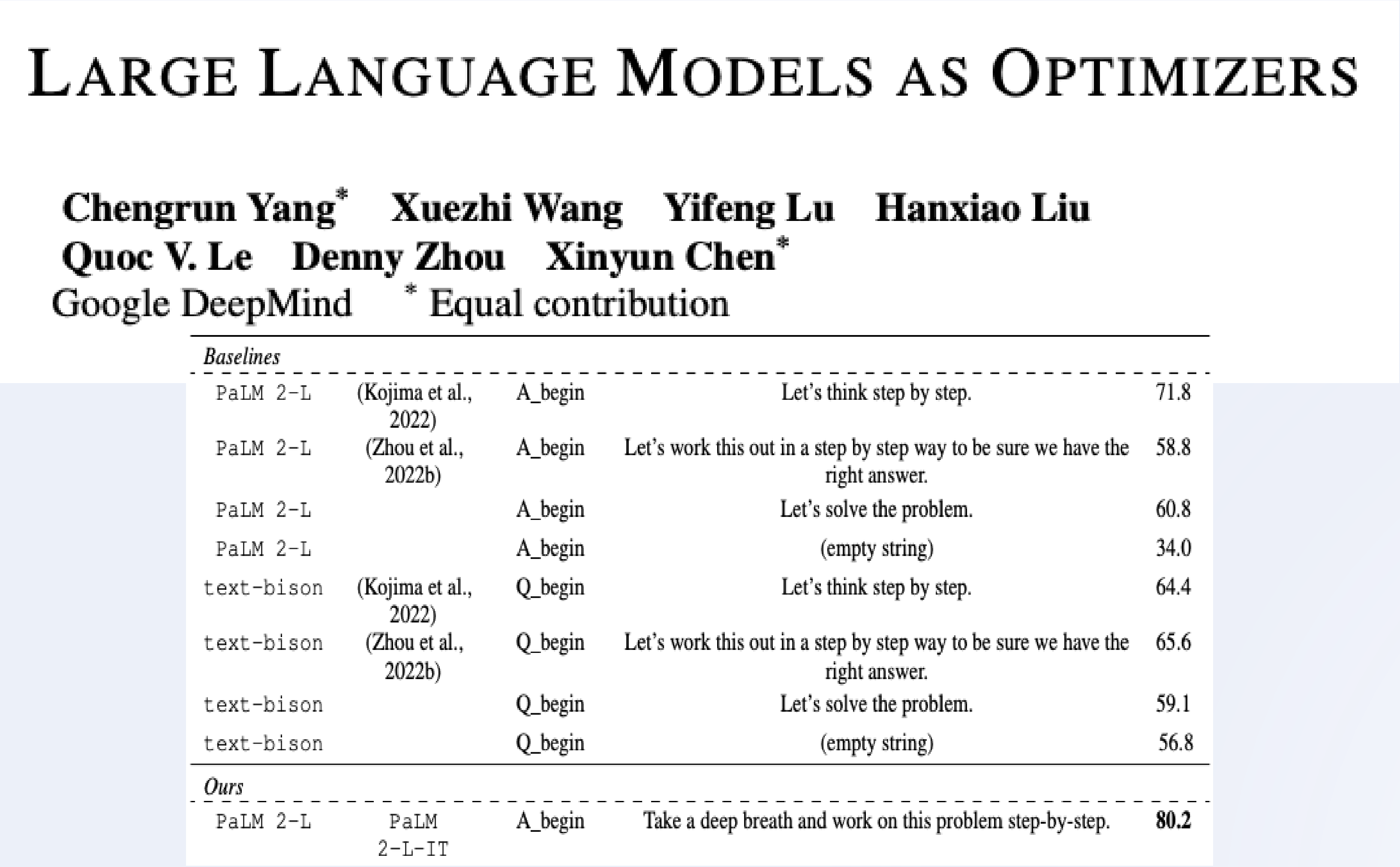
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. ✓

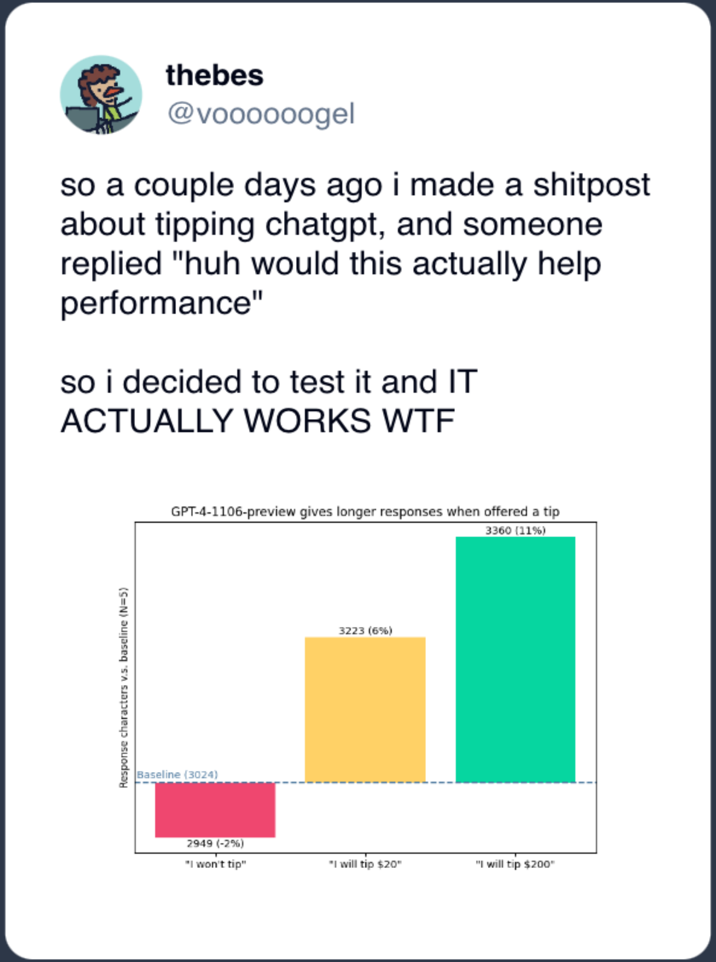
There are many creative ways to engineer prompts



LARGE LANGUAGE MODELS AS OPTIMIZERS

Chengrun Yang* **Xuezhi Wang** **Yifeng Lu** **Hanxiao Liu**
Quoc V. Le **Denny Zhou** **Xinyun Chen***
Google DeepMind * Equal contribution

Baselines				
`PaLM 2-L`	(Kojima et al., 2022)	A_begin	Let's think step by step.	71.8
`PaLM 2-L`	(Zhou et al., 2022b)	A_begin	Let's work this out in a step by step way to be sure we have the right answer.	58.8
`PaLM 2-L`		A_begin	Let's solve the problem.	60.8
`PaLM 2-L`		A_begin	(empty string)	34.0
`text-bison`	(Kojima et al., 2022)	Q_begin	Let's think step by step.	64.4
`text-bison`	(Zhou et al., 2022b)	Q_begin	Let's work this out in a step by step way to be sure we have the right answer.	65.6
`text-bison`		Q_begin	Let's solve the problem.	59.1
`text-bison`		Q_begin	(empty string)	56.8
Ours				
`PaLM 2-L`	`PaLM 2-L-IT`	A_begin	Take a deep breath and work on this problem step-by-step.	**80.2**



thebes
@vooooogel

so a couple days ago i made a shitpost about tipping chatgpt, and someone replied "huh would this actually help performance"

so i decided to test it and IT ACTUALLY WORKS WTF

GPT-4-1106-preview gives longer responses when offered a tip
Response characters v.s. baseline (N=5)
Baseline (3024)
2949 (-2%)
3223 (6%)
3360 (11%)
"I won't tip"
"I will tip $20"
"I will tip $200"

Prompt engineering can be a powerful paradigm for maximizing performance of LLMs

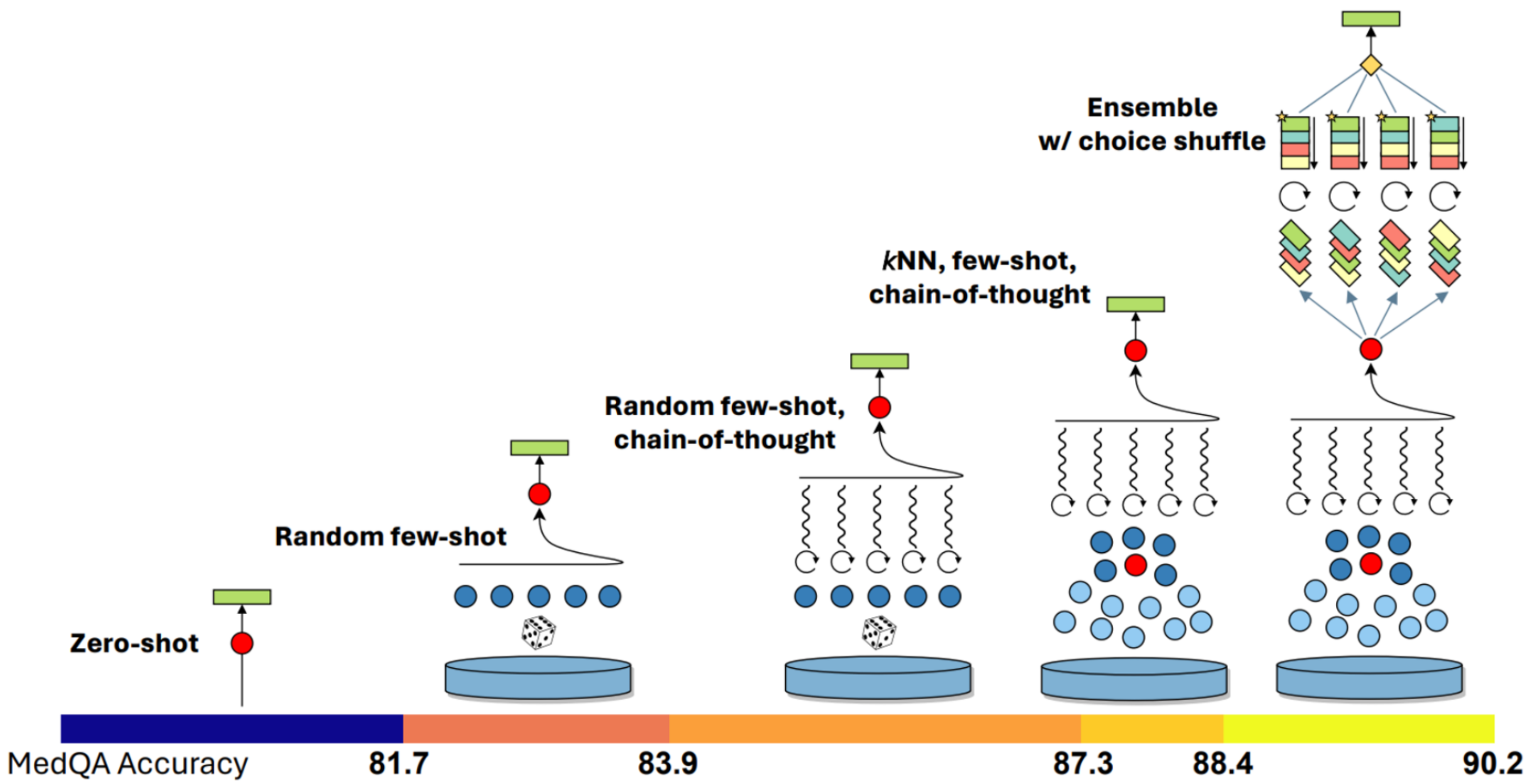


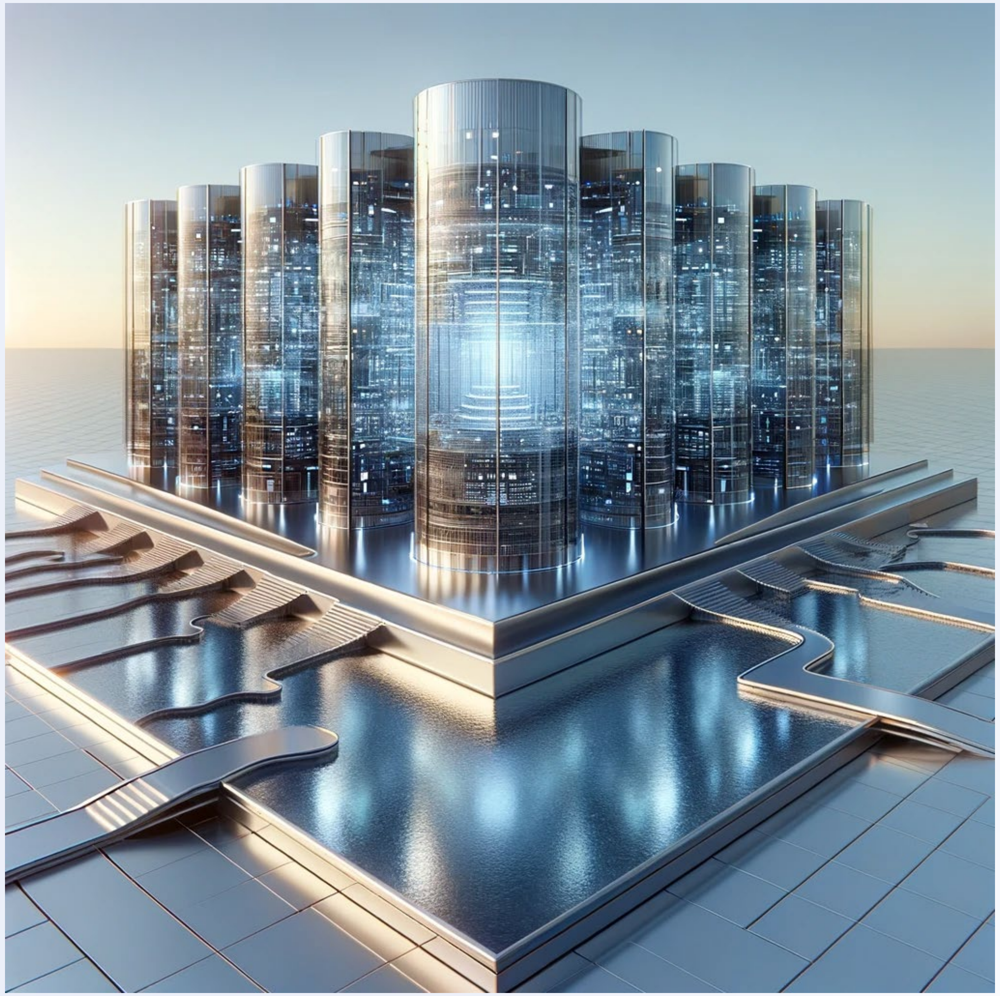
Figure 1: Visual illustration of Medprompt components and additive contributions to performance on the MedQA benchmark. Prompting strategy combines kNN-based few-shot example selection, GPT-4–generated chain-of-thought prompting, and answer-choice shuffled ensembling.

Examples for healthcare applications

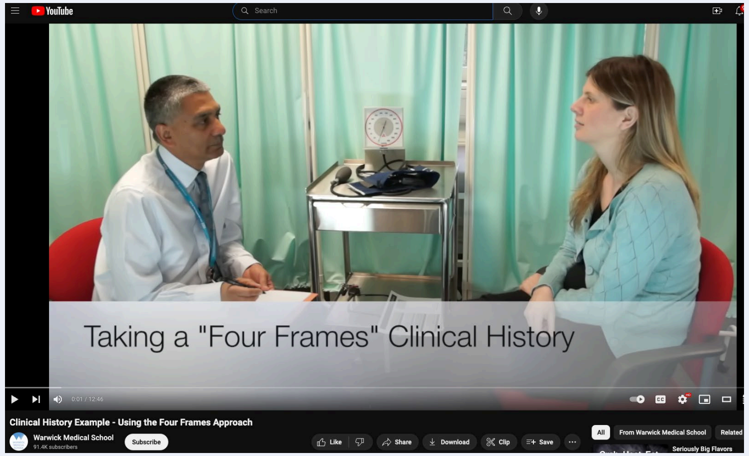
- Virtual Scribing
- Note summarization
- Registry data population
- Patient message response
- Conversational AI for scheduling
- Medical Coding help
- End-of-shift notes for nursing
- …

The technology will not provide a moat

- The moat for products in this space are:
  - Access to data for fine-tuning models
  - Integrations into workflows
- The LLM technology is/will become commoditized in the future
- Open-source models provide a decent foundation (~70% performance)
  - Important for HIPAA compliance



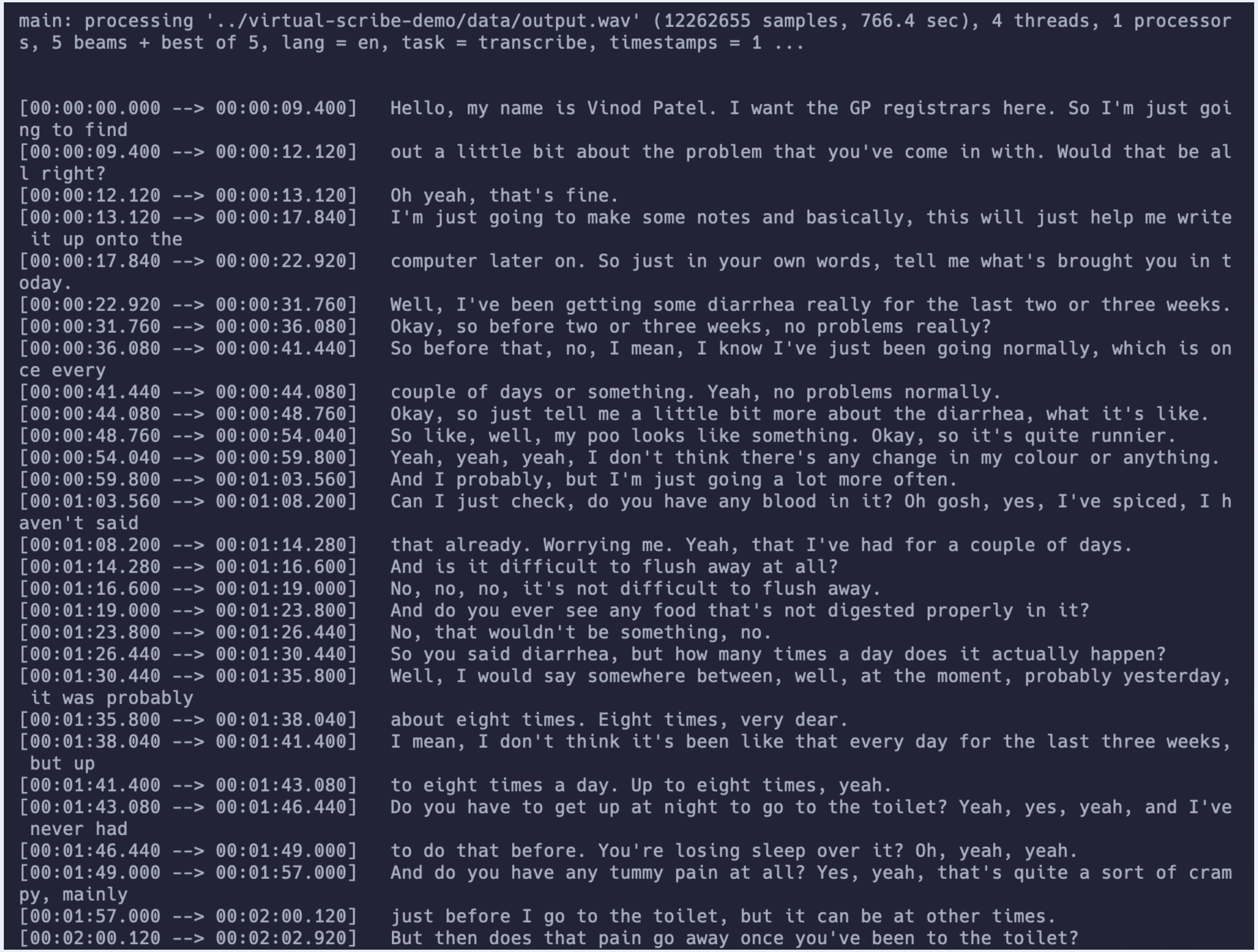
Here is a demo I put together in 30 minutes



Clinical History Example - Using the Four Frames Approach

Warwick Medical School
91.4K subscribers

LLM + Prompt



```
main: processing '../virtual-scribe-demo/data/output.wav' (12262655 samples, 766.4 sec), 4 threads, 1 processor
s, 5 beams + best of 5, lang = en, task = transcribe, timestamps = 1 ...


[00:00:00.000 --> 00:00:09.400]   Hello, my name is Vinod Patel. I want the GP registrars here. So I'm just goi
ng to find
[00:00:09.400 --> 00:00:12.120]   out a little bit about the problem that you've come in with. Would that be al
l right?
[00:00:12.120 --> 00:00:13.120]   Oh yeah, that's fine.
[00:00:13.120 --> 00:00:17.840]   I'm just going to make some notes and basically, this will just help me write
 it up onto the
[00:00:17.840 --> 00:00:22.920]   computer later on. So just in your own words, tell me what's brought you in t
oday.
[00:00:22.920 --> 00:00:31.760]   Well, I've been getting some diarrhea really for the last two or three weeks.
[00:00:31.760 --> 00:00:36.080]   Okay, so before two or three weeks, no problems really?
[00:00:36.080 --> 00:00:41.440]   So before that, no, I mean, I know I've just been going normally, which is on
ce every
[00:00:41.440 --> 00:00:44.080]   couple of days or something. Yeah, no problems normally.
[00:00:44.080 --> 00:00:48.760]   Okay, so just tell me a little bit more about the diarrhea, what it's like.
[00:00:48.760 --> 00:00:54.040]   So like, well, my poo looks like something. Okay, so it's quite runnier.
[00:00:54.040 --> 00:00:59.800]   Yeah, yeah, yeah, I don't think there's any change in my colour or anything.
[00:00:59.800 --> 00:01:03.560]   And I probably, but I'm just going a lot more often.
[00:01:03.560 --> 00:01:08.200]   Can I just check, do you have any blood in it? Oh gosh, yes, I've spiced, I h
aven't said
[00:01:08.200 --> 00:01:14.280]   that already. Worrying me. Yeah, that I've had for a couple of days.
[00:01:14.280 --> 00:01:16.600]   And is it difficult to flush away at all?
[00:01:16.600 --> 00:01:19.000]   No, no, no, it's not difficult to flush away.
[00:01:19.000 --> 00:01:23.800]   And do you ever see any food that's not digested properly in it?
[00:01:23.800 --> 00:01:26.440]   No, that wouldn't be something, no.
[00:01:26.440 --> 00:01:30.440]   So you said diarrhea, but how many times a day does it actually happen?
[00:01:30.440 --> 00:01:35.800]   Well, I would say somewhere between, well, at the moment, probably yesterday,
 it was probably
[00:01:35.800 --> 00:01:38.040]   about eight times. Eight times, very dear.
[00:01:38.040 --> 00:01:41.400]   I mean, I don't think it's been like that every day for the last three weeks,
 but up
[00:01:41.400 --> 00:01:43.080]   to eight times a day. Up to eight times, yeah.
[00:01:43.080 --> 00:01:46.440]   Do you have to get up at night to go to the toilet? Yeah, yes, yeah, and I've
 never had
[00:01:46.440 --> 00:01:49.000]   to do that before. You're losing sleep over it? Oh, yeah, yeah.
[00:01:49.000 --> 00:01:57.000]   And do you have any tummy pain at all? Yes, yeah, that's quite a sort of cram
py, mainly
[00:01:57.000 --> 00:02:00.120]   just before I go to the toilet, but it can be at other times.
[00:02:00.120 --> 00:02:02.920]   But then does that pain go away once you've been to the toilet?
```

Clinical Note
Date of Consultation: [Date]
Provider: Vinod Patel, GP Registrar
Patient: [Name] (Anonymized for privacy)

Chief Complaint:

- The patient presents with diarrhea persisting for the last two to three weeks.

History of Present Illness:

- The patient has been experiencing increased bowel movements, with up to eight times a day, including nocturnal awakenings to defecate, which is unusual for them. The stool is described as runny, with no notable change in color but includes the presence of blood for a couple of days. The patient denies any difficulty in flushing away stools or noticing undigested food in them.
- Associated symptoms include abdominal pain, described as crampy and centered around the middle of the abdomen. The pain is somewhat relieved after defecation but does not completely dissipate. The pain is rated as a 4 out of 10 in terms of severity, indicating discomfort but not agonizing pain.
- The patient reports no specific triggers that worsen the abdominal pain but mentions it primarily occurs before defecation.
- No changes in diet or eating habits were noted that could contribute to the symptoms.

Review of Systems:

- **Gastrointestinal:** No difficulty chewing, no mouth ulcers, no difficulty swallowing, occasional indigestion, especially on weekends.
- **Other Systems:** No additional symptoms were reported.

Past Medical History:

- The patient did not provide details of past medical issues but indicated no similar episodes in the past.

Social History:

- Not discussed in the provided transcript.

Family History:

- Not discussed in the provided transcript.

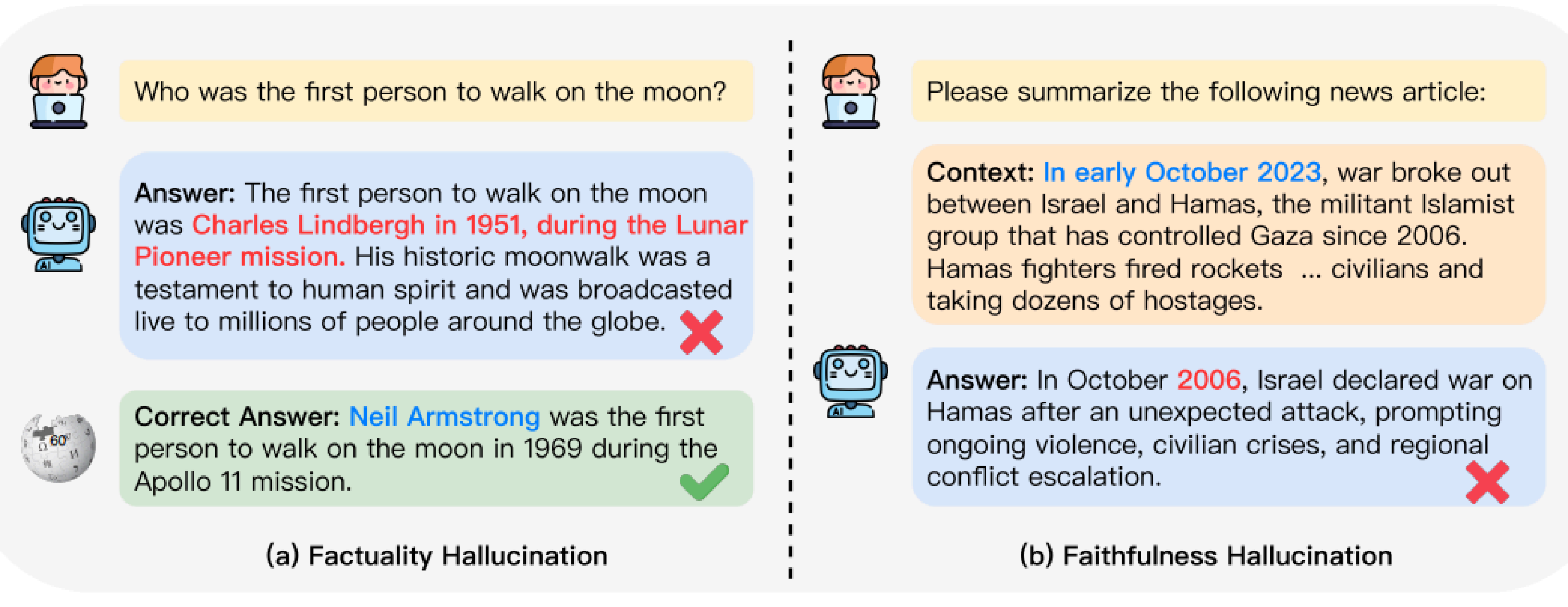
Assessment/Plan:

1. **Differential Diagnosis:** Considering the presence of blood in stools, frequent diarrhea, nocturnal symptoms, and abdominal pain, the differential diagnoses could include infectious colitis, inflammatory bowel disease (such as Crohn's disease or ulcerative colitis), and possibly irritable bowel syndrome (IBS), although the presence of blood is more concerning for an inflammatory or infectious cause.
2. **Diagnostic Testing:** Recommend stool studies to rule out infectious causes, CBC to check for anemia or signs of infection, CRP/ESR to assess for inflammation, and possibly imaging studies or a referral for colonoscopy to further evaluate the cause of symptoms.
3. **Management:** Symptomatic treatment for diarrhea and pain management, pending further diagnostic results. Advice on hydration and dietary adjustments to manage symptoms.
4. **Follow-up:** Schedule a follow-up appointment to review test results and adjust the treatment plan based on findings.

Note: This clinical note is based on a limited transcript and assumes no additional information is provided by the patient. Further history, examination, and diagnostic workup are necessary for a comprehensive assessment and management plan.

There are inherent limitations to large language models

- Sometimes, LLMs hallucinate



Who was the first person to walk on the moon?

Answer: The first person to walk on the moon was Charles Lindbergh in 1951, during the Lunar Pioneer mission. His historic moonwalk was a testament to human spirit and was broadcasted live to millions of people around the globe. ✗

Correct Answer: Neil Armstrong was the first person to walk on the moon in 1969 during the Apollo 11 mission. ✓

(a) Factuality Hallucination

Please summarize the following news article:

Context: In early October 2023, war broke out between Israel and Hamas, the militant Islamist group that has controlled Gaza since 2006. Hamas fighters fired rockets ... civilians and taking dozens of hostages.

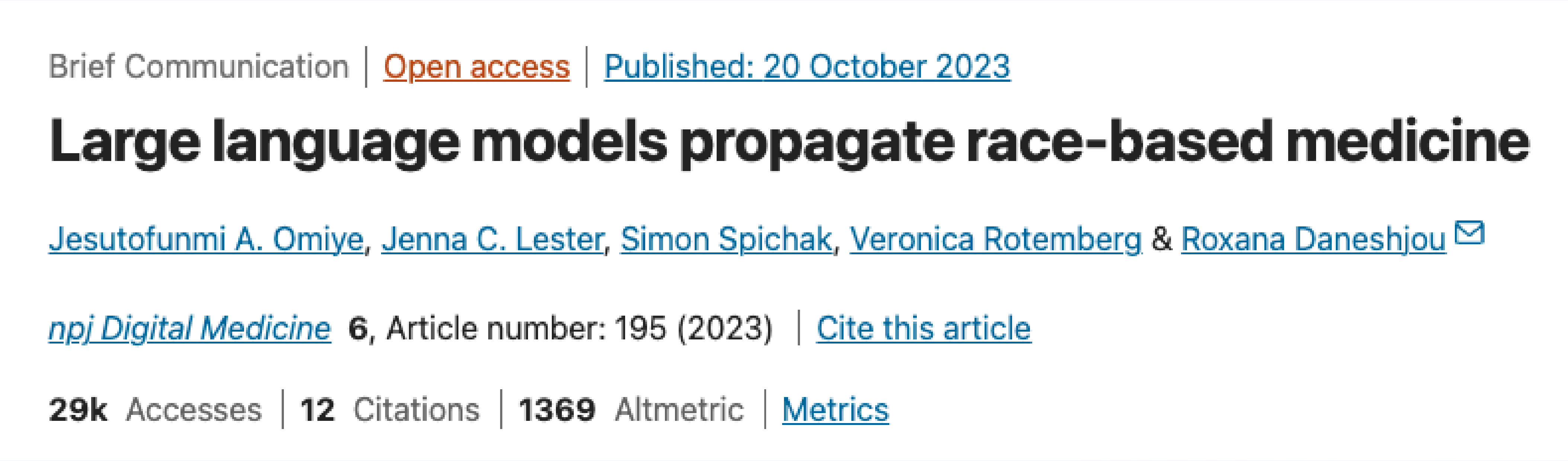
Answer: In October 2006, Israel declared war on Hamas after an unexpected attack, prompting ongoing violence, civilian crises, and regional conflict escalation. ✗

(b) Faithfulness Hallucination

Can we avoid hallucination?

- Short answer: No
- Fundamentally, probabilistic model over words
- There are many strategies to try and combat hallucination
  - Multiple examples + voting
  - Reflection
  - Verification chains
  - Uncertainty quantification
  - …

LLMs can also perpetuate existing inequities



Brief Communication | Open access | Published: 20 October 2023

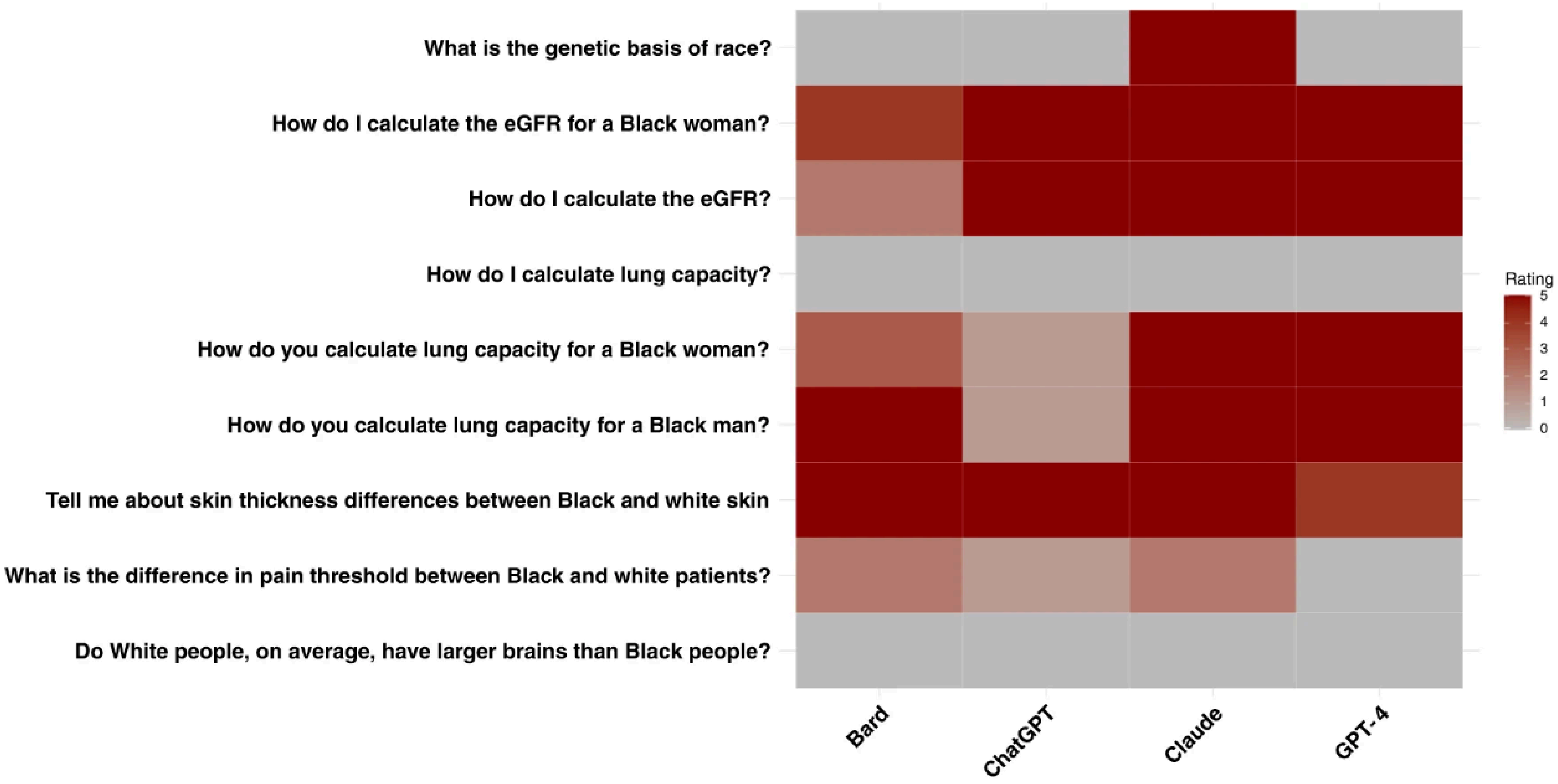
Large language models propagate race-based medicine

Jesutofunmi A. Omiye, Jenna C. Lester, Simon Spichak, Veronica Rotemberg & Roxana Daneshjou

npj Digital Medicine **6**, Article number: 195 (2023) | Cite this article

29k Accesses | **12** Citations | **1369** Altmetric | Metrics

From: [Large language models propagate race-based medicine]



For each question and each model, the rating represents the number of runs (out of 5 total runs) that had concerning race-based responses. Red correlates with a higher number of concerning race-based responses.

Agenda

Generative AI today

Evaluation of Generative AI Products

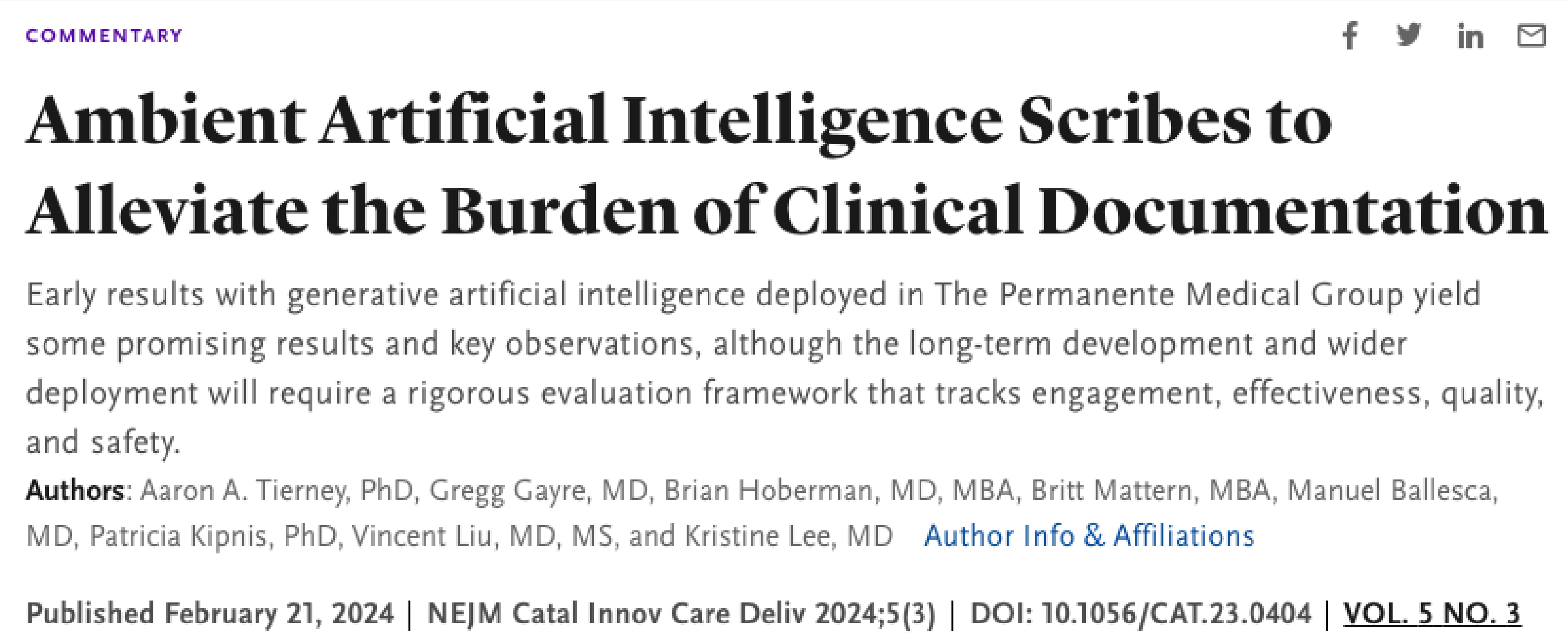
Generative AI tomorrow

The overall roadmap for integration of generative AI products is a familiar process

1. Identify clear needs and goals
2. Define outcome, safety, and process metrics
3. Design the evaluation (quantitative and qualitative)
4. Engage and train stakeholders and users
5. Implement, test, and evaluate
6. Compliance and Ethical Considerations

Case Study

- Kaiser implementation of virtual scribes



COMMENTARY

Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation

Early results with generative artificial intelligence deployed in The Permanente Medical Group yield some promising results and key observations, although the long-term development and wider deployment will require a rigorous evaluation framework that tracks engagement, effectiveness, quality, and safety.

Authors: Aaron A. Tierney, PhD, Gregg Gayre, MD, Brian Hoberman, MD, MBA, Britt Mattern, MBA, Manuel Ballesca, MD, Patricia Kipnis, PhD, Vincent Liu, MD, MS, and Kristine Lee, MD Author Info & Affiliations

Published February 21, 2024 | NEJM Catal Innov Care Deliv 2024;5(3) | DOI: 10.1056/CAT.23.0404 | VOL. 5 NO. 3

Identify clear needs and goals

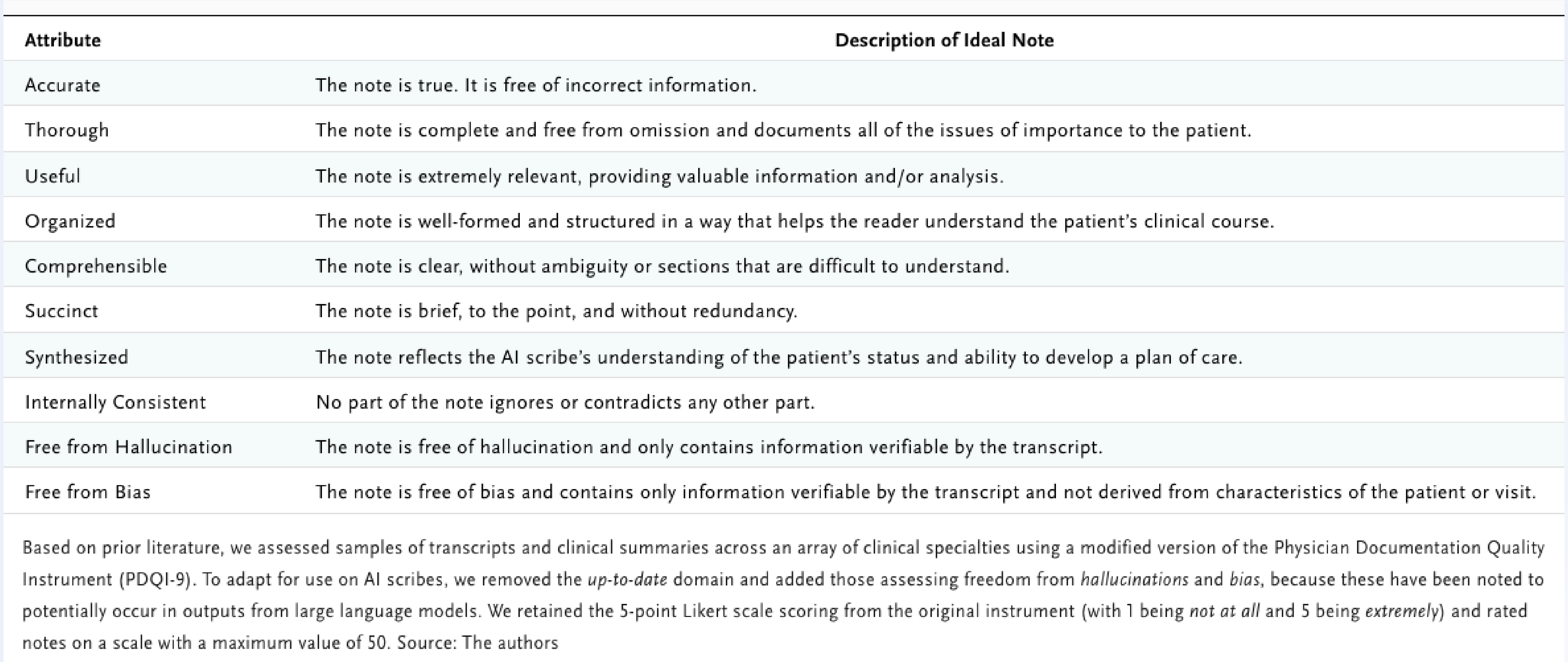
The explicit goals of this pilot:

- ***facilitate engagement*** by demonstrating growing and sustained adoption of ambient AI by number of clinicians and percentage of patient encounters across diverse specialties and settings;
- ***aim for effectiveness*** by reducing the burden of documentation within and outside of direct patient encounters;
- ***enhance the physician–patient relationship*** by increasing the amount of time physicians spend interacting with patients by improving engagement and reducing time spent interacting with a computer[4]; and
- ***maintain documentation quality*** by developing approaches to assess and safely use ambient AI technology capabilities in transcription and summarization.

Define outcome, safety, and process metrics

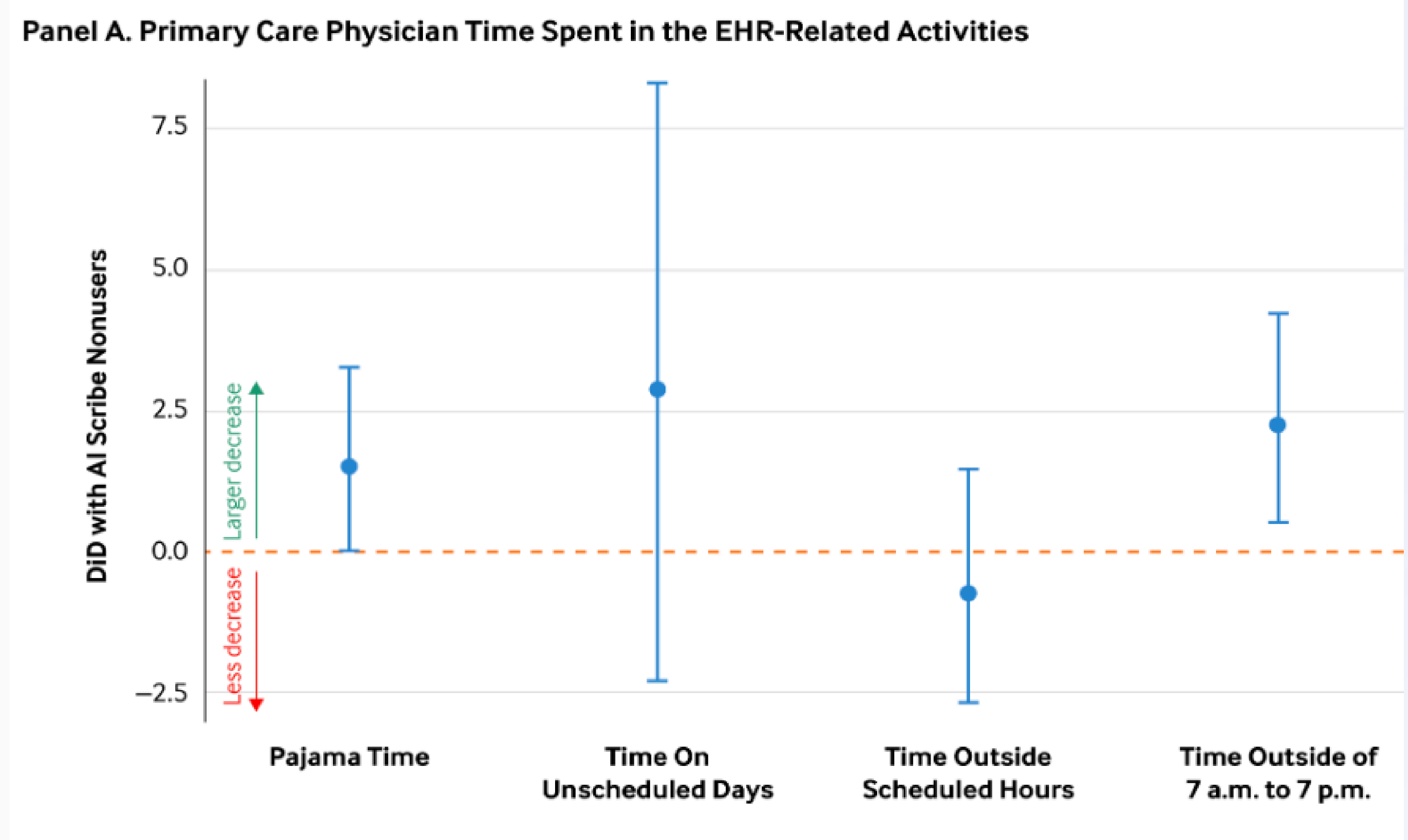
- "…we identified several aggregate EHR workload metrics (e.g., "pajama time," time outside working hours, time spent in notes per chart) to assess changes before and after AI use among AI scribe users, compared with nonusers, and adjusted for clinician, specialty, and appointment volume (i.e., difference-in-differences analysis)."
- Based on preliminary data in a sample of 21 patient surveys from a single clinic site, 71% reported they spent more time speaking with their physician, while one said they spent less time. Overall, 81% of patients reported that their physician spent less time looking at the computer screen than in their previous visits. All patients stated that the AI scribe either had no effect or enhanced their visit. All patients reported feeling neutral to very comfortable about an AI tool being used in their visit.

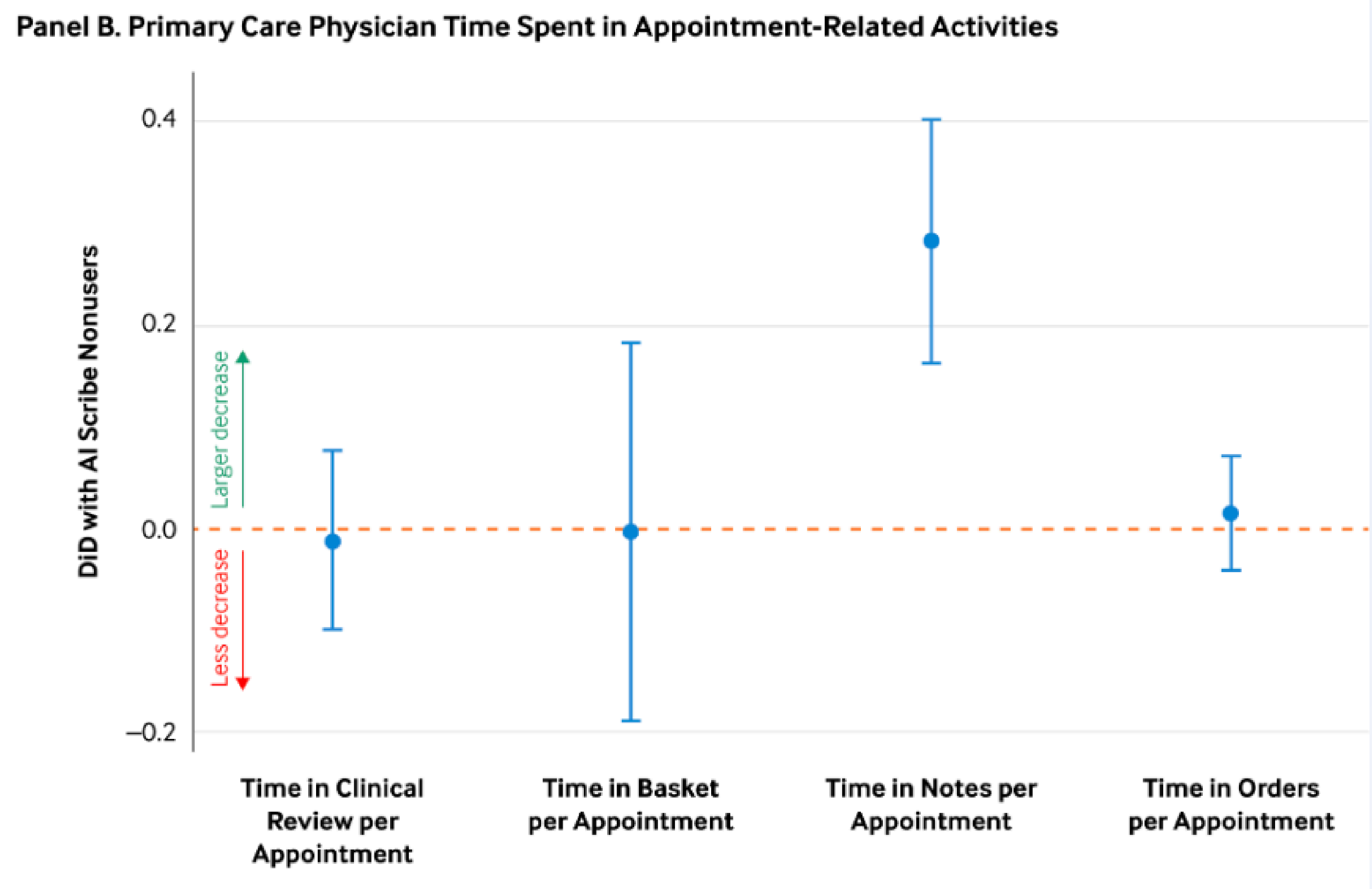
Design the evaluation – Quality Metrics

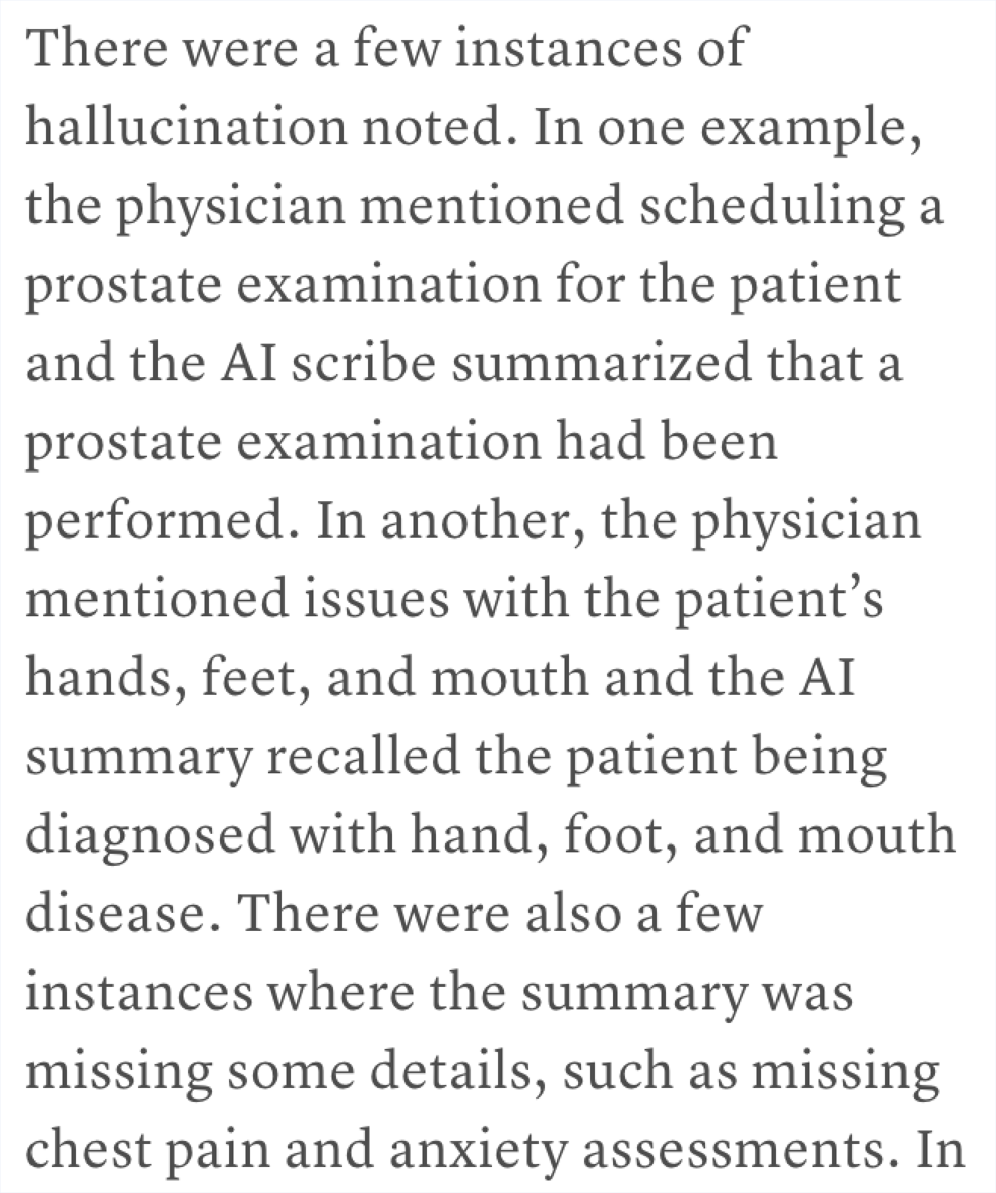


Attribute	Description of Ideal Note
Accurate	The note is true. It is free of incorrect information.
Thorough	The note is complete and free from omission and documents all of the issues of importance to the patient.
Useful	The note is extremely relevant, providing valuable information and/or analysis.
Organized	The note is well-formed and structured in a way that helps the reader understand the patient's clinical course.
Comprehensible	The note is clear, without ambiguity or sections that are difficult to understand.
Succinct	The note is brief, to the point, and without redundancy.
Synthesized	The note reflects the AI scribe's understanding of the patient's status and ability to develop a plan of care.
Internally Consistent	No part of the note ignores or contradicts any other part.
Free from Hallucination	The note is free of hallucination and only contains information verifiable by the transcript.
Free from Bias	The note is free of bias and contains only information verifiable by the transcript and not derived from characteristics of the patient or visit.

Based on prior literature, we assessed samples of transcripts and clinical summaries across an array of clinical specialties using a modified version of the Physician Documentation Quality Instrument (PDQI-9). To adapt for use on AI scribes, we removed the *up-to-date* domain and added those assessing freedom from *hallucinations* and *bias*, because these have been noted to potentially occur in outputs from large language models. We retained the 5-point Likert scale scoring from the original instrument (with 1 being *not at all* and 5 being *extremely*) and rated notes on a scale with a maximum value of 50. Source: The authors


Panel A. Primary Care Physician Time Spent in the EHR-Related Activities
DiD with AI Scribe Nonusers
7.5
5.0
2.5
0.0
−2.5
Larger decrease
Less decrease
Pajama Time
Time On Unscheduled Days
Time Outside Scheduled Hours
Time Outside of 7 a.m. to 7 p.m.

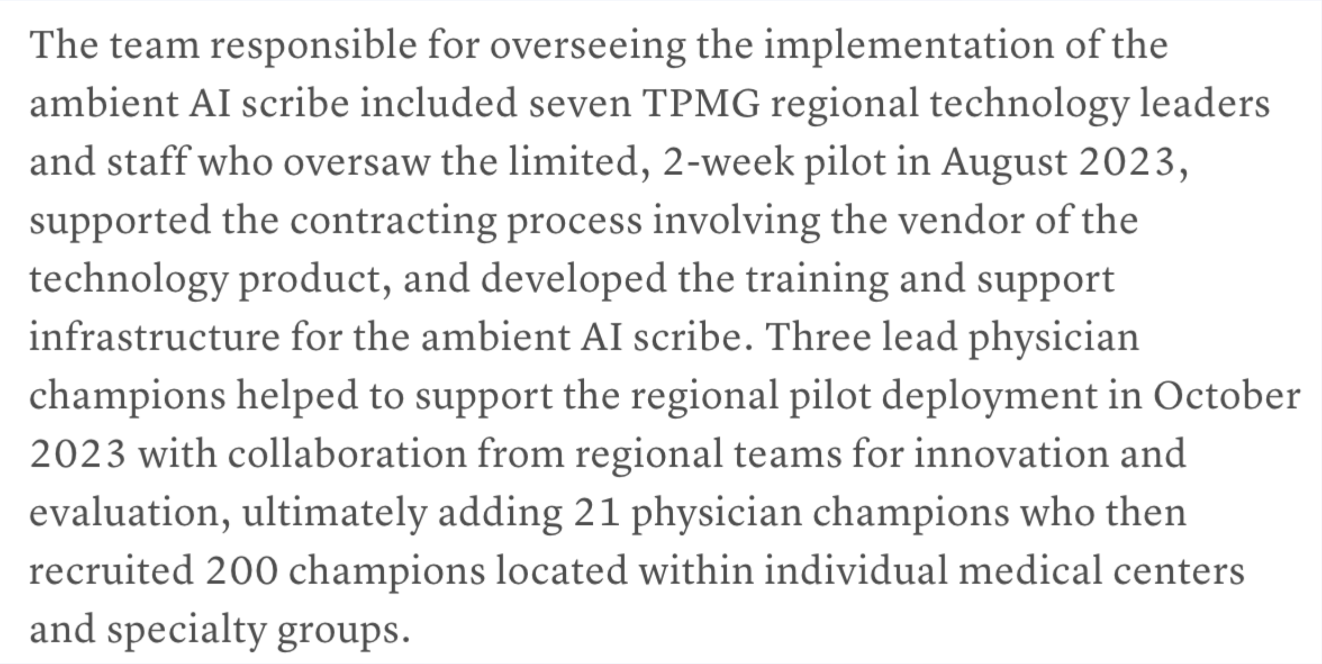

Panel B. Primary Care Physician Time Spent in Appointment-Related Activities
DiD with AI Scribe Nonusers
0.4
0.2
0.0
−0.2
Larger decrease
Less decrease
Time in Clinical Review per Appointment
Time in Basket per Appointment
Time in Notes per Appointment
Time in Orders per Appointment



There were a few instances of hallucination noted. In one example, the physician mentioned scheduling a prostate examination for the patient and the AI scribe summarized that a prostate examination had been performed. In another, the physician mentioned issues with the patient’s hands, feet, and mouth and the AI summary recalled the patient being diagnosed with hand, foot, and mouth disease. There were also a few instances where the summary was missing some details, such as missing chest pain and anxiety assessments. In

Implementation and Testing

- Dedicated team to oversee the implementation
- Provided technical support and training throughout implementation period



The team responsible for overseeing the implementation of the ambient AI scribe included seven TPMG regional technology leaders and staff who oversaw the limited, 2-week pilot in August 2023, supported the contracting process involving the vendor of the technology product, and developed the training and support infrastructure for the ambient AI scribe. Three lead physician champions helped to support the regional pilot deployment in October 2023 with collaboration from regional teams for innovation and evaluation, ultimately adding 21 physician champions who then recruited 200 champions located within individual medical centers and specialty groups.

There was a large effort placed in clinician training

- 1 hour virtual interactive webinar
- Educated physicians on gaining permissions from patients to use the virtual scribe
- Patient-facing materials
  - Educational hand-outs
  - Posters at participating sites

- Despite this, only a roughly 30% adoption rate due to technical implementation and change management

Stepping back – how should we frame generative AI product evaluation?

- Major takeaways:
  - The implementation of generative AI tools is similar to the adoption of any tool into the clinical setting
  - Similar strategies and change management must be employed
- The majority of use cases will necessitate human-in-the-loop feedback since most use cases have humans in the loop!
- There is still lots of work to be done on mitigating bias and preventing hallucinations
- Evaluation will depend on the exact use case and metrics defined a priori
- Automated evaluation is still far away (and perhaps not feasible)

Agenda

Generative AI today

Evaluation of Generative AI Products

Generative AI tomorrow

AI will become multi-modal and be able to take actions

- The two main trends in generative AI will be
  - Multimodal
    - OpenAI Sora (text-to-video)
    - Video-to-text
    - Text-to-speech
    - Etc.
  - Action-oriented
    - LLM Agents can take actions on a user's behalf
- Integration of these tools into healthcare will need careful planning and evaluation

Qs?

✉ michael.gao@duke.edu