Daniel A. Donoho, MD 10/25/2023

Artificial Intelligence for Practicing Pediatricians

Claiming CME Credit

The INOVA CME management system changed effective June 2023. What does this mean for you?

1. All providers must create an account on the new platform, visit: cme.inova.org.
2. Once you have an account, credit for this session can be claimed in one of two ways:
    1. Text today's session code **VUTCAG** to 703-260-9391.
    2. Visit cme.inova.org/code to enter today's session code ("**VUTCAG**") on the website.

CME credit must be claimed within 30 days of the presentation date.

Disclosures

NIH - NIBIB K23
Internal Funding
dotphrase.ai – cofounder, unvalued
Nonprofit SDSC. - founder, nonprofit
& tech stack at the end …

REAL DISCLOSURES:
I am not an AI enthusiast nor an AI expert
AI is underhyped
Replacement and doomer narratives are overhyped
Impact of AI will be both jarring and imperceptible


TODAY
Artificial Intelligence
Add Topic +
Will AI replace your doctor? As a physician, I'm worried new tech will hurt patient care.

Digital Neurosurgery 2023

Artificial Intelligence x Neurosurgery in Palo Alto, CA
Computer Vision Seminar on October 12, 2023
Main Meeting on October 13-15, 2023

REGISTER

Agenda →

DigitalNSGY.com

Aims of this Talk

1. **Explain** why AI/ML hype has intensified
2. **Define** commonly used terminology
3. **Provide a framework** to consider ”AI products”
4. **Demonstrative examples**

Core messages of AI in healthcare

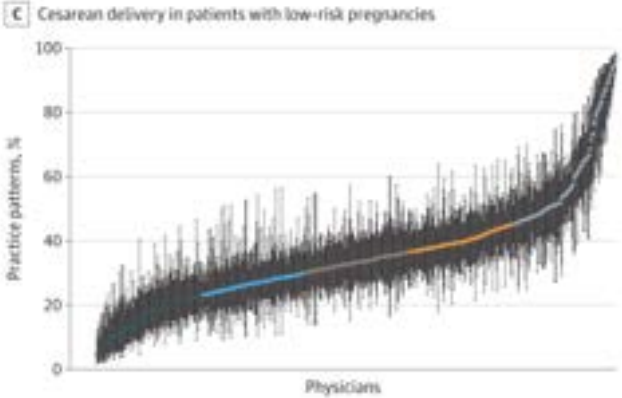
Digital transformation is **inevitable**

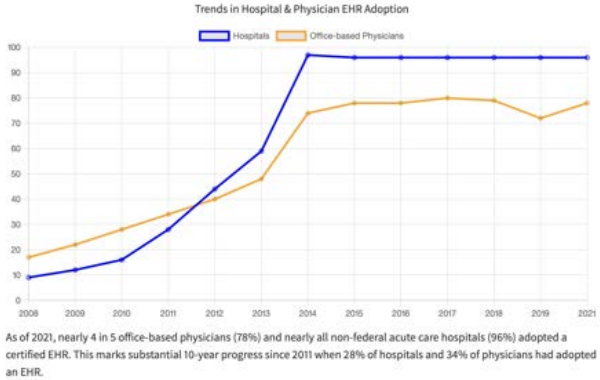
Capabilities increase **rapidly**

AI is not **magical**, particularly in medicine

A "competent user" makes **all the difference**

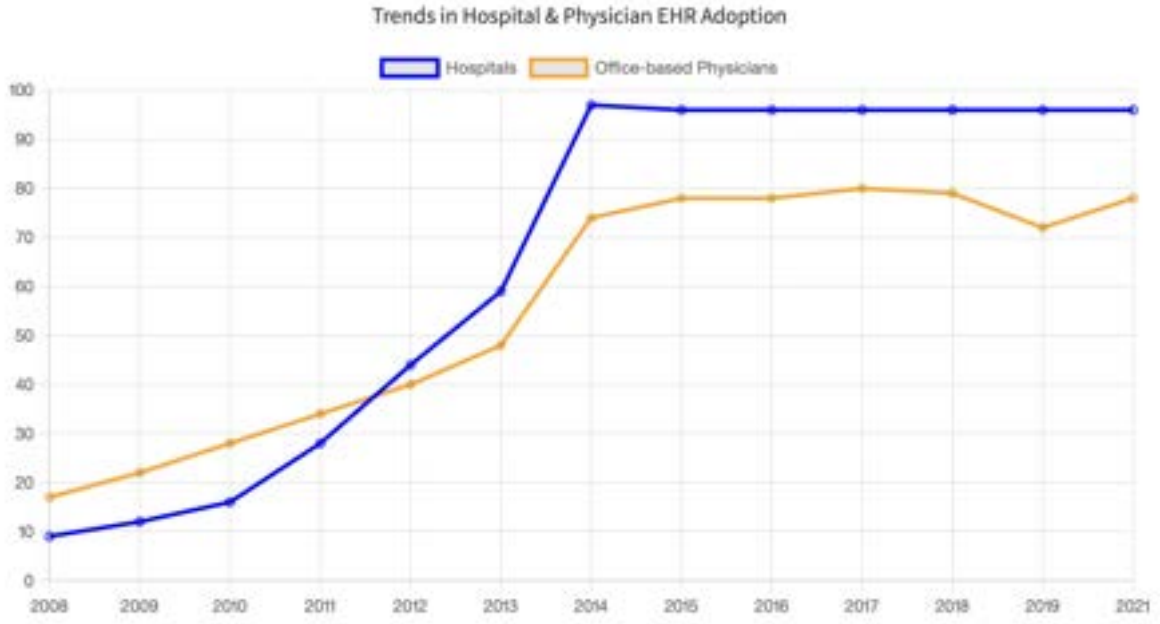
Counterarguments That We All Know





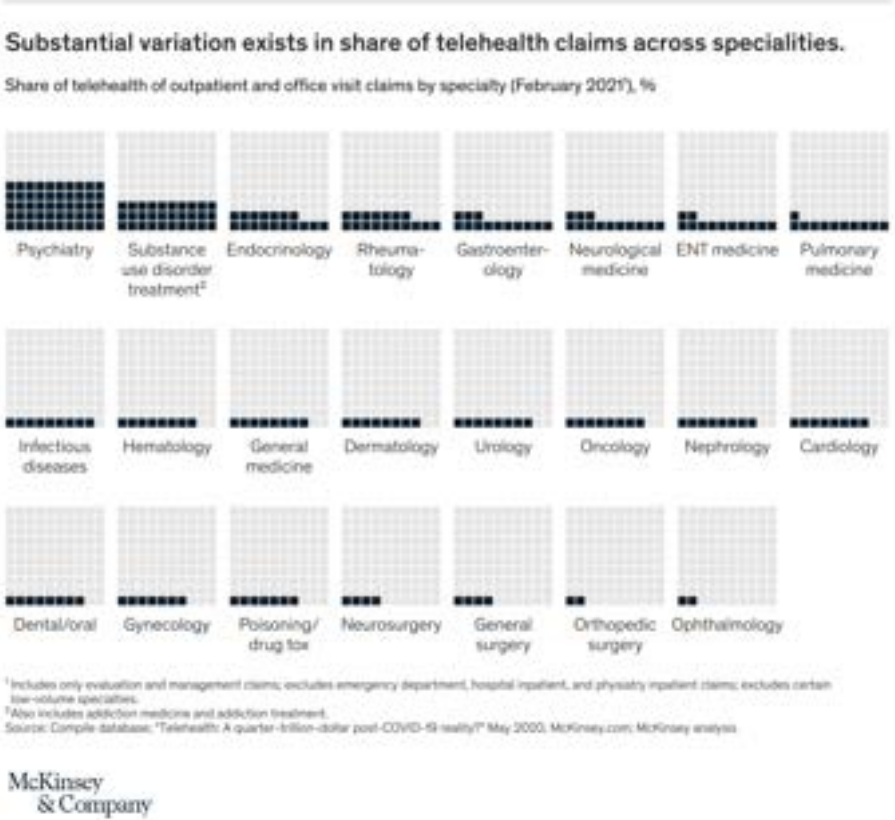
As of 2021, nearly 4 in 5 office-based physicians (78%) and nearly all non-federal acute care hospitals (96%) adopted a certified EHR. This marks substantial 10-year progress since 2011 when 28% of hospitals and 34% of physicians had adopted an EHR.

V1.0 of digitization of medicine: EHR / eRx



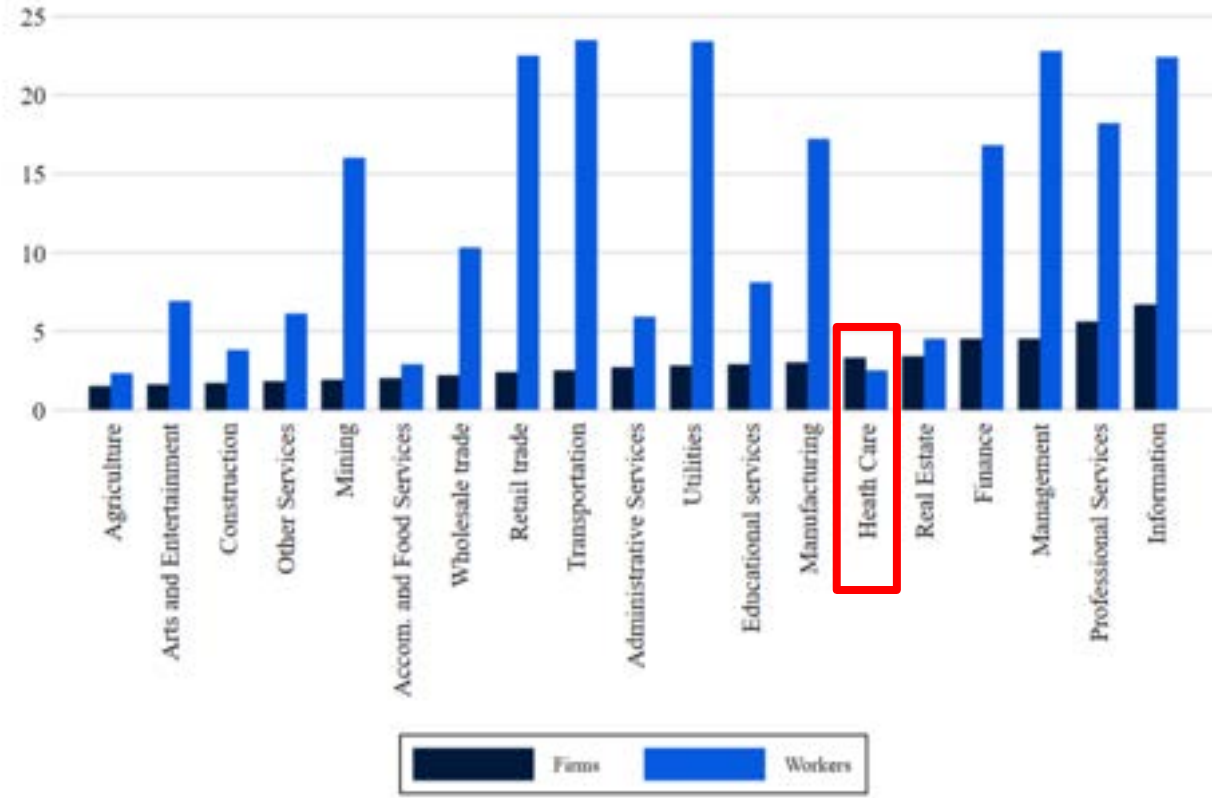
As of 2021, nearly 4 in 5 office-based physicians (78%) and nearly all non-federal acute care hospitals (96%) adopted a certified EHR. This marks substantial 10-year progress since 2011 when 28% of hospitals and 34% of physicians had adopted an EHR.

We lived through v2.0



Little has changed … but much is changing

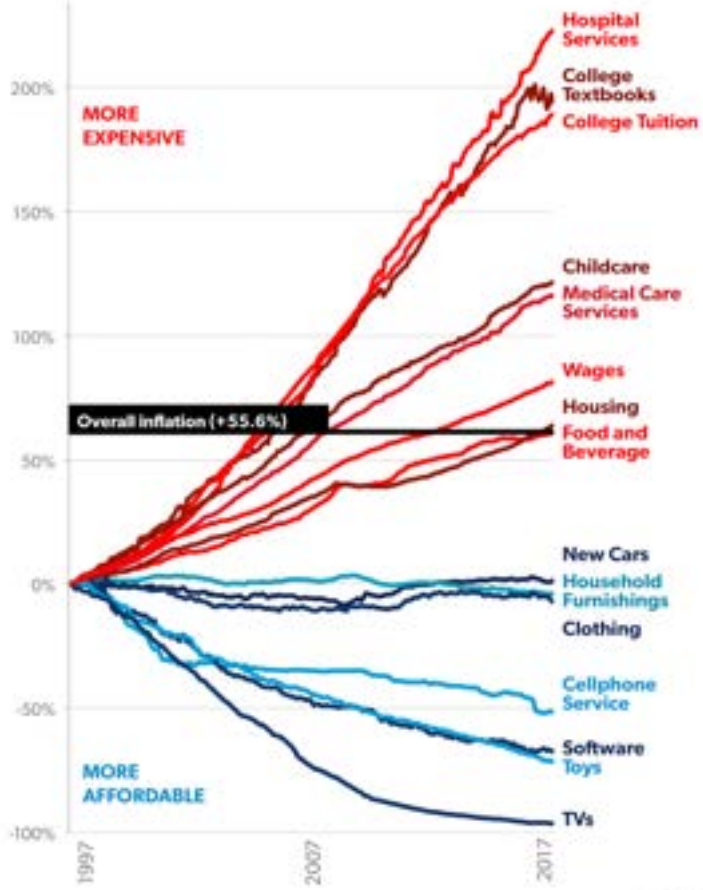
The third wave of digital disruption in medicine



Few Healthcare firms and workers use AI/ML…

Price changes (Jan. 1997–Dec. 2017)

Selected US Consumer Goods and Services, and Wages



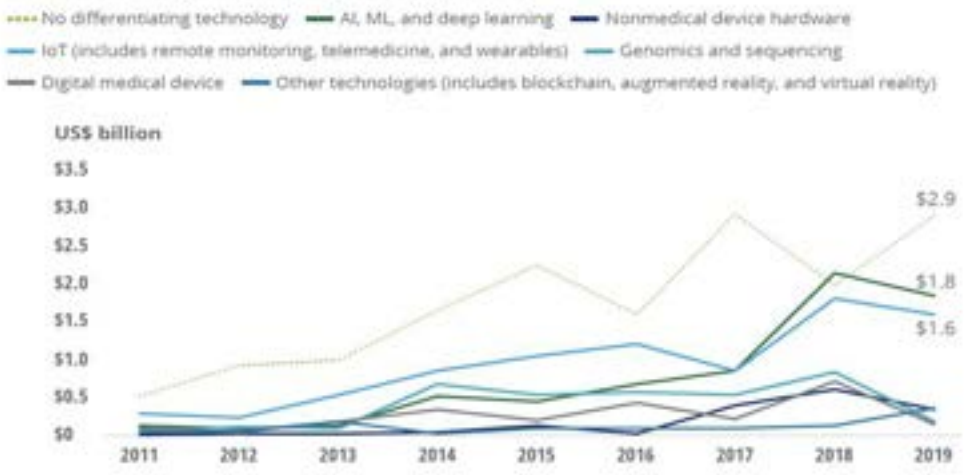
Source: BLS

Carpe Diem AEI

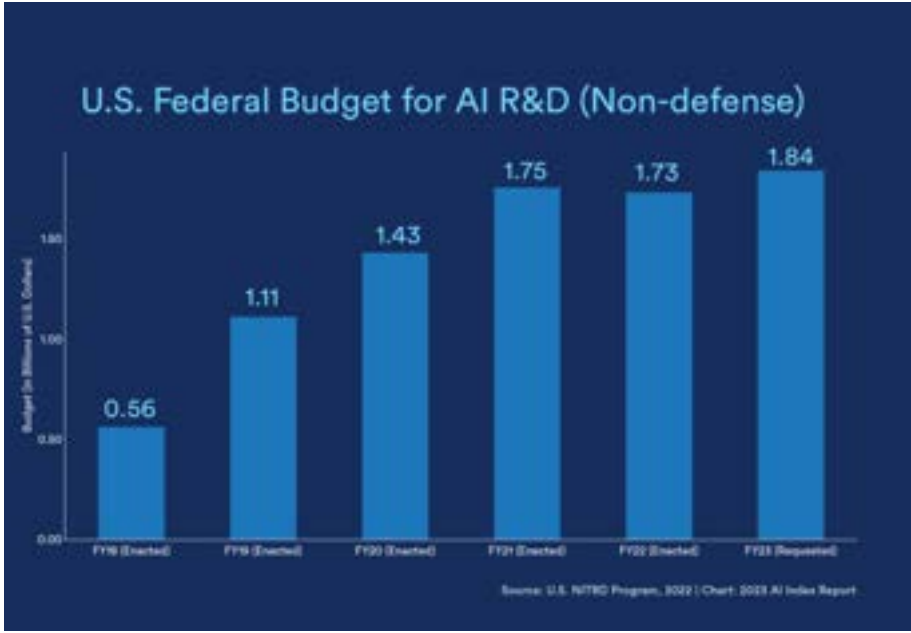
In 2022, the AI focus area with the most investment was medical and healthcare ($6.1 billion); followed by data management,

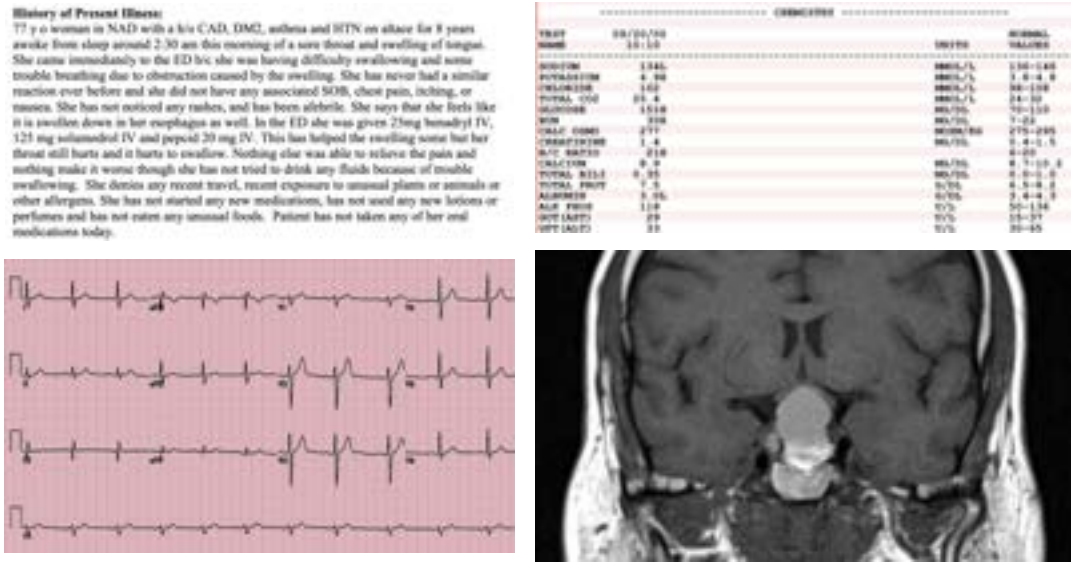
Of innovators that are using differentiating technologies, those leveraging AI, ML, and deep learning received the most funding in recent years

Venture funding, by differentiating technology, 2011–19



Source: Rock Health Digital Health Funding Database and Deloitte analysis.



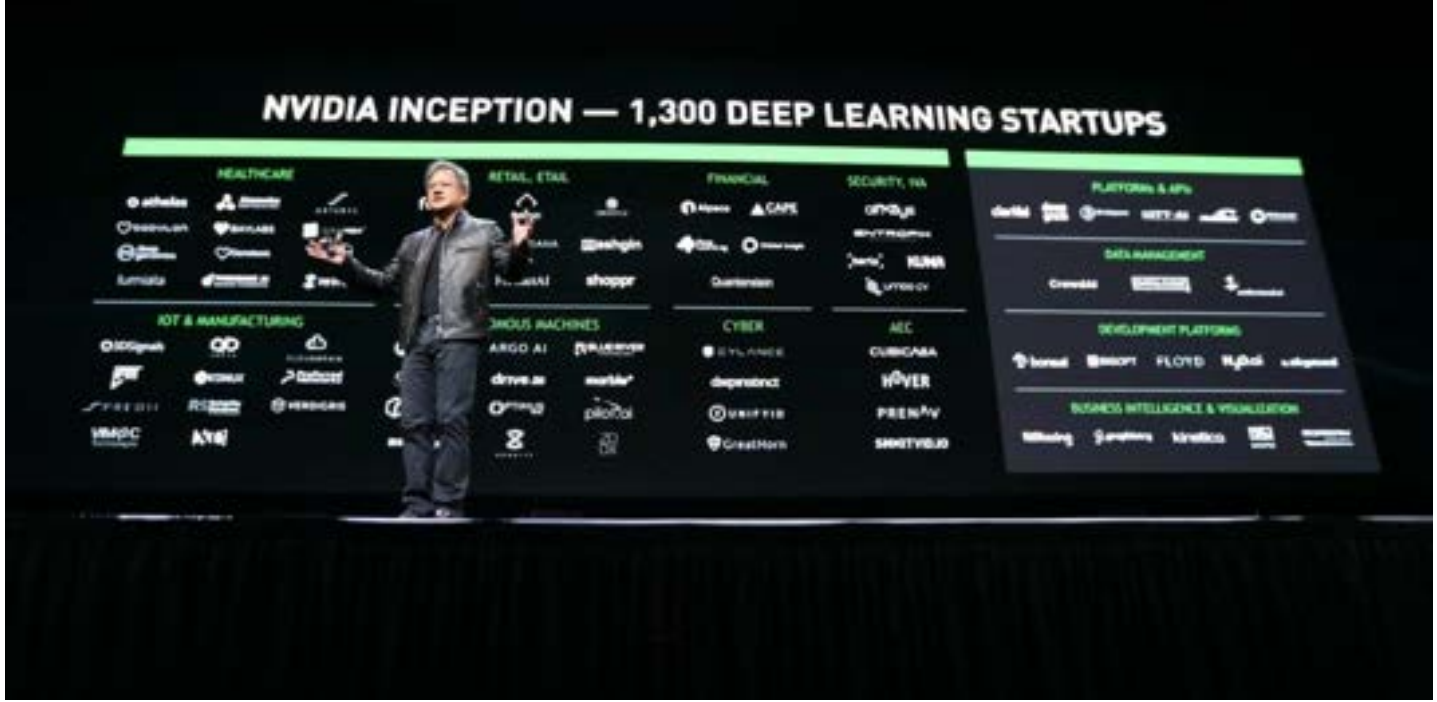


History of Present Illness:
77 y o woman in NAD with a h/o CAD, DM2, asthma and HTN on altace for 8 years awoke from sleep around 2:30 am this morning of a sore throat and swelling of tongue. She came immediately to the ED b/c she was having difficulty swallowing and some trouble breathing due to obstruction caused by the swelling. She has never had a similar reaction ever before and she did not have any associated SOB, chest pain, itching, or nausea. She has not noticed any rashes, and has been afebrile. She says that she feels like it is swollen down in her esophagus as well. In the ED she was given 25mg benadryl IV, 125 mg solumedrol IV and pepcid 20 mg IV. This has helped the swelling some but her throat still hurts and it hurts to swallow. Nothing else was able to relieve the pain and nothing make it worse though she has not tried to drink any fluids because of trouble swallowing. She denies any recent travel, recent exposure to unusual plants or animals or other allergens. She has not started any new medications, has not used any new lotions or perfumes and has not eaten any unusual foods. Patient has not taken any of her oral medications today.

	A	B	C	D	E
1		Sales in Each Quarter			
2	Product Name	Jan'2018	April'2018	July'2018	October'2018
3	ABC Mutton	$ 2,667.60	$ 4,013.10	$ 4,836.00	$ 6,087.90
4	Crab Meat	$ 1,768.41	$ 1,978.00	$ 4,412.32	$ 1,656.00
5	Camembert Pierrot	$ 3,182.40	$ 4,683.50	$ 9,579.50	$ 3,060.00
6	Ipoh Coffee	$ 1,398.40	$ 4,496.50	$ 1,196.00	$ 3,979.00
7	Hot Pepper Sauce	$ 1,347.36	$ 2,750.69	$ 1,375.62	$ 3,899.51
8	Hot Spiced Okra	$ 1,509.60	$ 530.40	$ 68.00	$ 850.00
9	Mozzarella di Giovanni	$ 1,390.00	$ 4,488.20	$ 3,027.60	$ 2,697.00
10	Sir Rodney's Scones	$ 1,462.00	$ 644.00	$ 1,733.00	$ 1,434.00
11	Steeleye Stout	$ 1,310.40	$ 1,368.00	$ 1,323.00	$ 1,273.50
12	Veggie-spread	$ 1,202.87	$ 263.40	$ 842.88	$ 2,590.10
13	Grand Total	$ 19,239.04			
14					

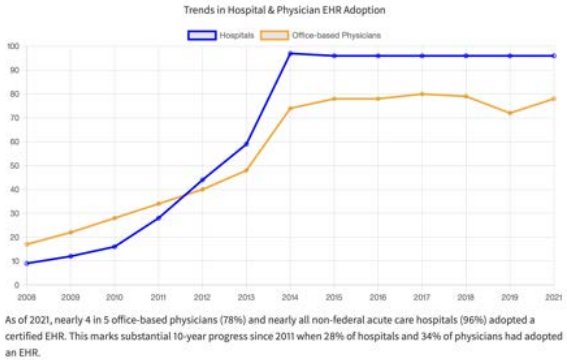
Unstructured* Unlabelled/sparsely labelled data

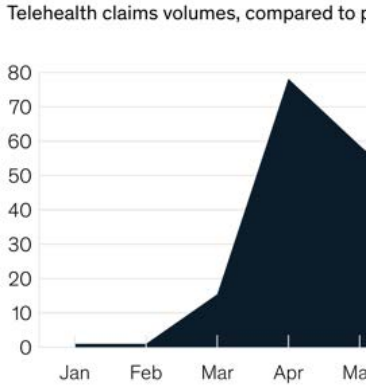
Structured, Labelled Data



The “deep learning revolution” in AI/ML: systems that model relationships in **massive quantities of unlabelled**, **unstructured data**

What's Different about this third "digitization" hype Cycle?





Broad consumer trends (we saw this before…)
Technological disruption of daily work (ditto)
"Novel" technology … with far reaching implications … and unknown limitations

Commonly Used Terminology

Artificial Intelligence - phenomenon and family

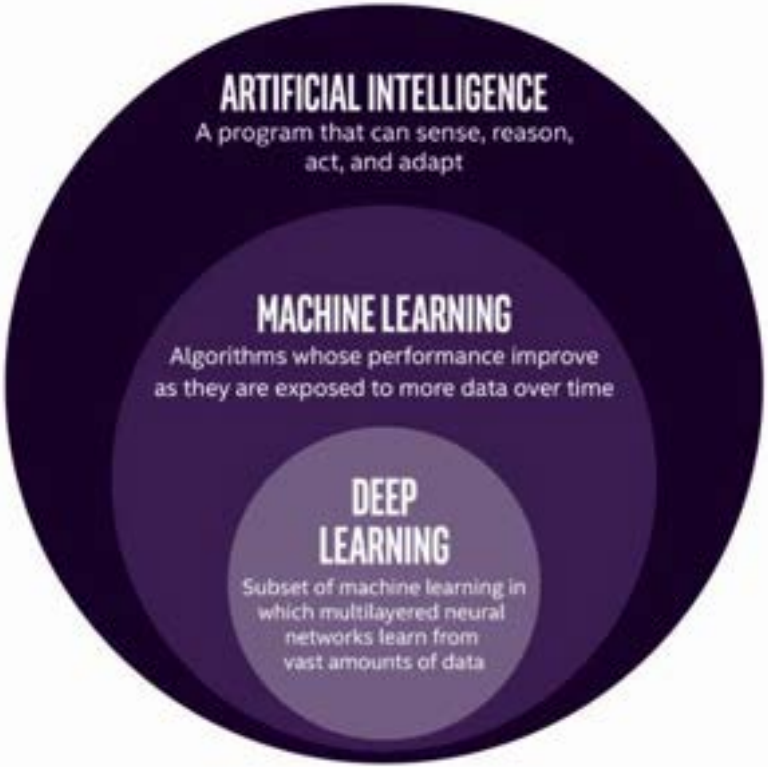
- Weak/Narrow AI (ML) vs. Strong/General AI (“AI”)
- **Generative AI**

Machine Learning - without explicit instruction

- Non-Neural Network: Good old regression
- Neural Network

Deep Learning - multiple layers

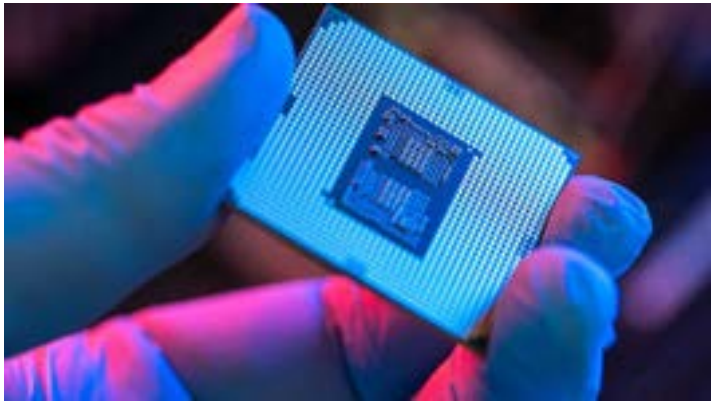
- Supervised Learning
- Semi- / Unsupervised Learning
- Reinforcement Learning

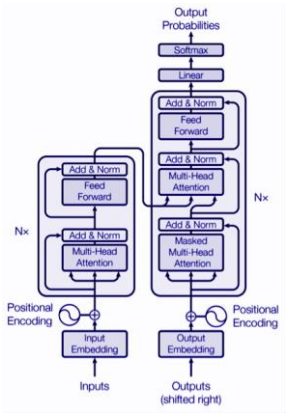


Is AI new? No …

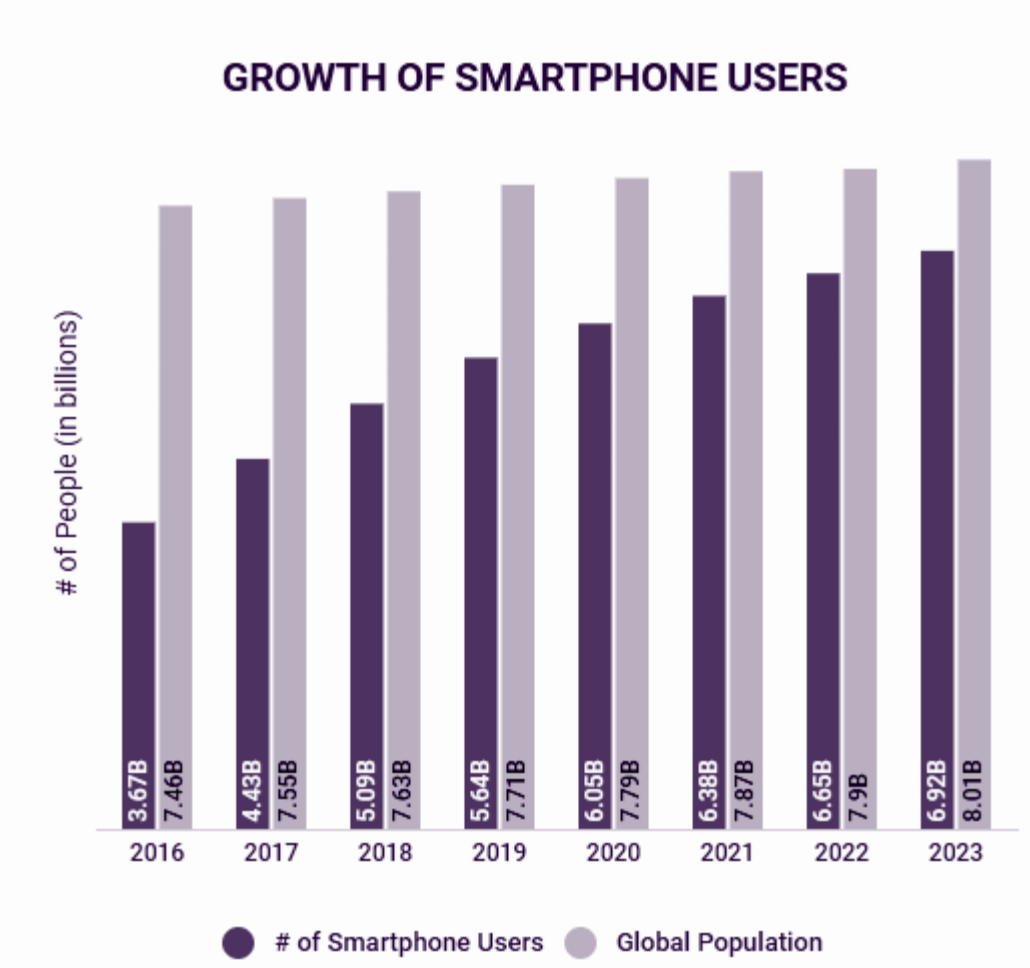
We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

Why now?




Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)


4:43
FaceTime
Tuesday
24
Calendar
Photos
Camera
Mail
Clock
Maps
Weather
Reminders
Notes
Stocks
News
Books
App Store
Podcasts
TV
Health
Home
Wallet
Settings

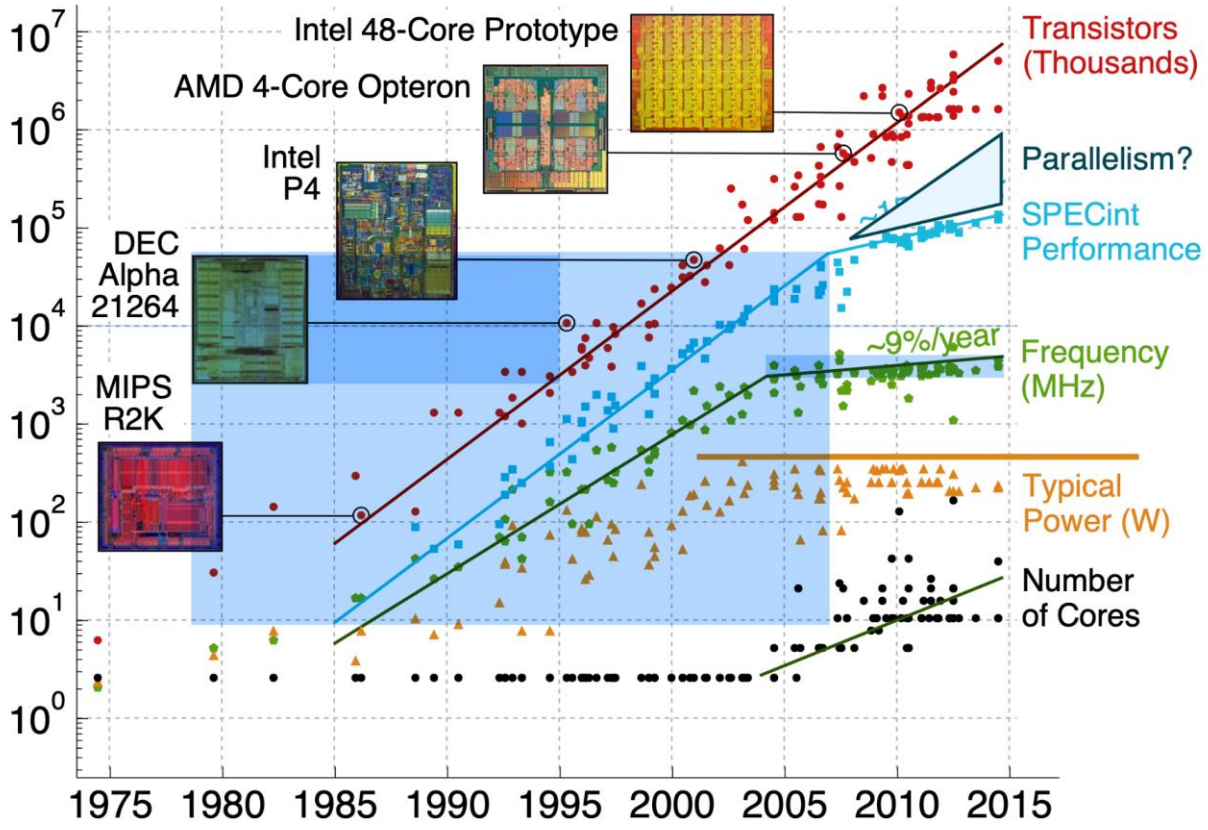

GROWTH OF SMARTPHONE USERS
# of People (in billions)
3.67B
7.46B
4.43B
7.55B
5.09B
7.63B
5.64B
7.71B
6.05B
7.79B
6.38B
7.87B
6.65B
7.9B
6.92B
8.01B
2016
2017
2018
2019
2020
2021
2022
2023
# of Smartphone Users
Global Population

Global Data Creation is About to Explode

Actual and forecast amount of data created worldwide 2010-2035 (in zettabytes)

1 zettabyte is equal to **1 billion terabytes**.

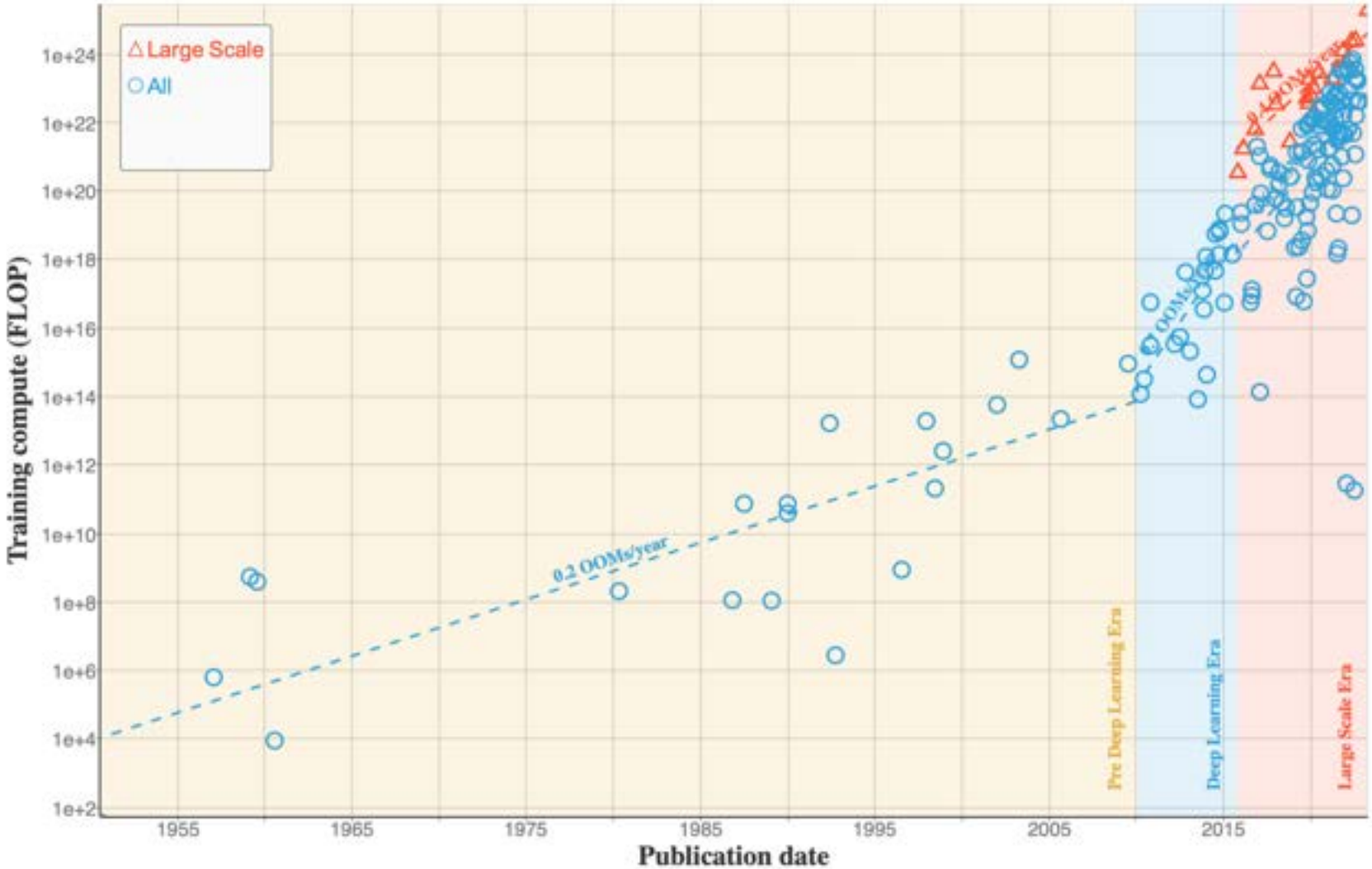
Trend 3: The Multicore "Hail Mary Pass"

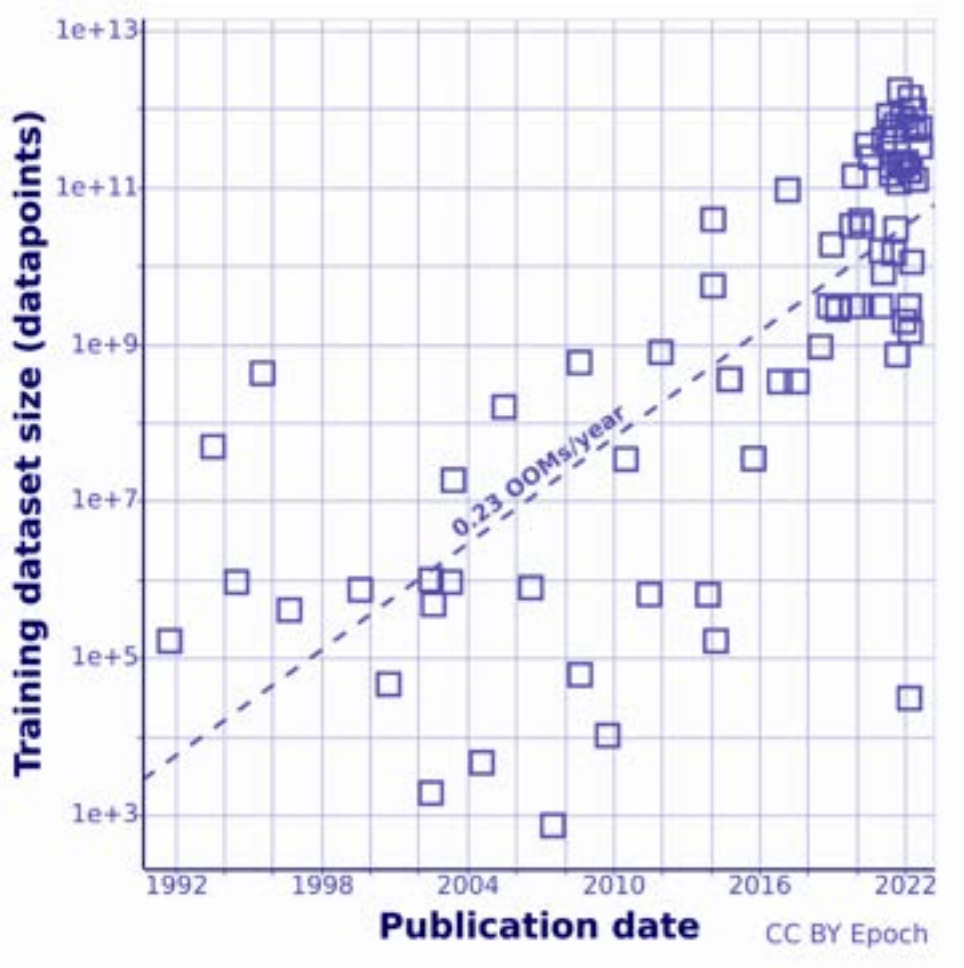


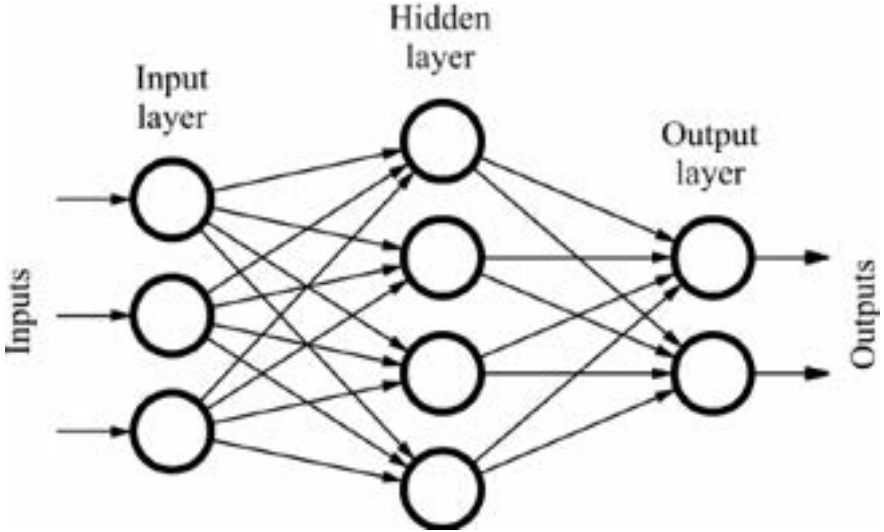
Another possibility is that a select few of us will be able to catch today's risky Hail Mary pass. Perhaps only multimedia apps such as video games can exploit data-level parallelism and take advantage of the increasing number of cores. In that case, the microprocessors of 2020 may look more like the GPUs from Nvidia, Advanced Micro Devices, and Intel than the traditional microprocessors of today. That is, the GPU will be promoted from a sideshow to the main event. It's unclear whether such applications by themselves will be able to sustain the growth of the information technology industry as a whole.

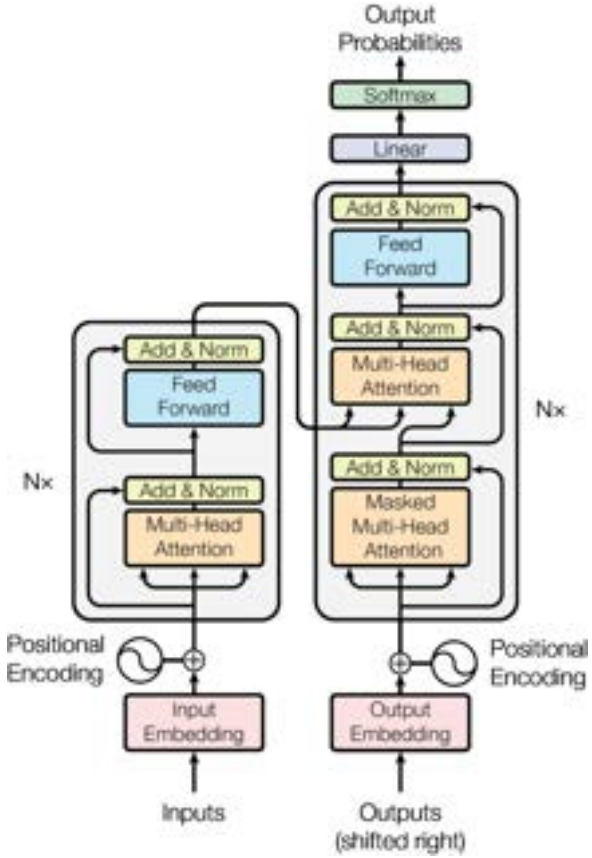
IEEE Spectrum | The Trouble With Multicore

BY DAVID PATTERSON | 30 JUN 2010


Large Scale
All
Training compute (FLOP)
1e+24
1e+22
1e+20
1e+18
1e+16
1e+14
1e+12
1e+10
1e+8
1e+6
1e+4
1e+2
1955
1965
1975
1985
1995
2005
2015
Publication date
0.2 OOMs/year
Pre Deep Learning Era
Deep Learning Era
Large Scale Era


Training dataset size (datapoints)
1e+13
1e+11
1e+9
1e+7
1e+5
1e+3
0.23 OOMs/year
1992
1998
2004
2010
2016
2022
Publication date
CC BY Epoch





Children's National

And one ingredient that puts all of these innovations together…

Something fundamental to computation-based research really has changed in the last ten years. In certain fields, progress is simply dramatically more rapid than previously. Researchers in affected fields are living through a period of profound transformation, as the fields undergo a transition to frictionless reproducibility (FR). This transition markedly changes the rate of spread of ideas and practices, affects mindsets, and erases memories of much that came before.

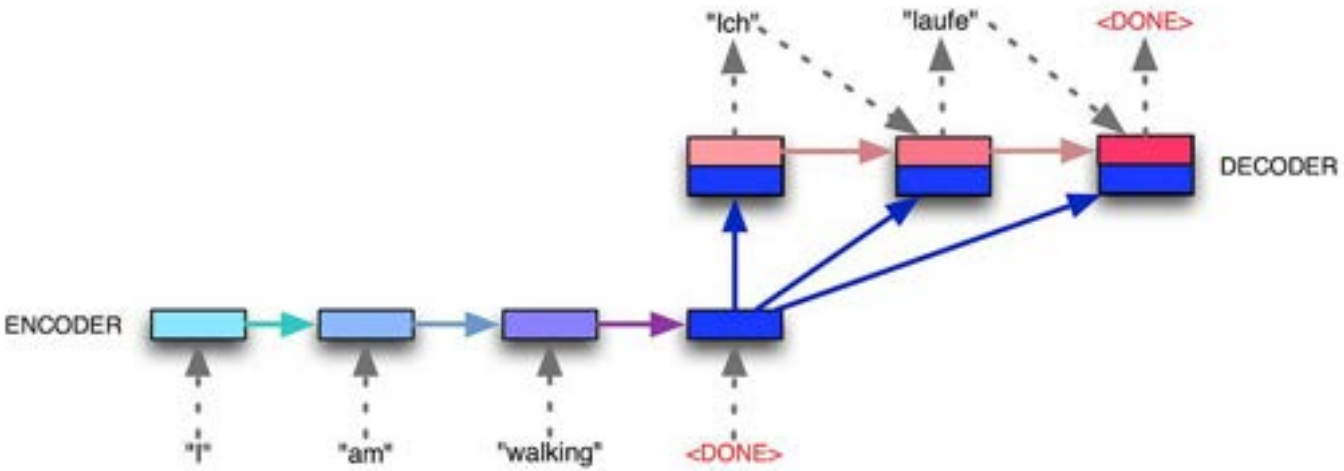
The emergence of frictionless reproducibility follows from the maturation of 3 data science principles that came together after decades of work by many technologists and numerous research communities. The mature principles involve data sharing, code sharing, and competitive challenges, *however* implemented in the particularly strong form of frictionless open services. Empirical Machine Learning (EML) is today's leading adherent field, and its consequent rapid changes are responsible for the AI progress we see. Still, other fields can and do benefit when they adhere to the same principles.

Data Science at the Singularity
Version 1.00

David Donoho *

October 3, 2023

“Generative” models : the power of seq2seq



What if you could…

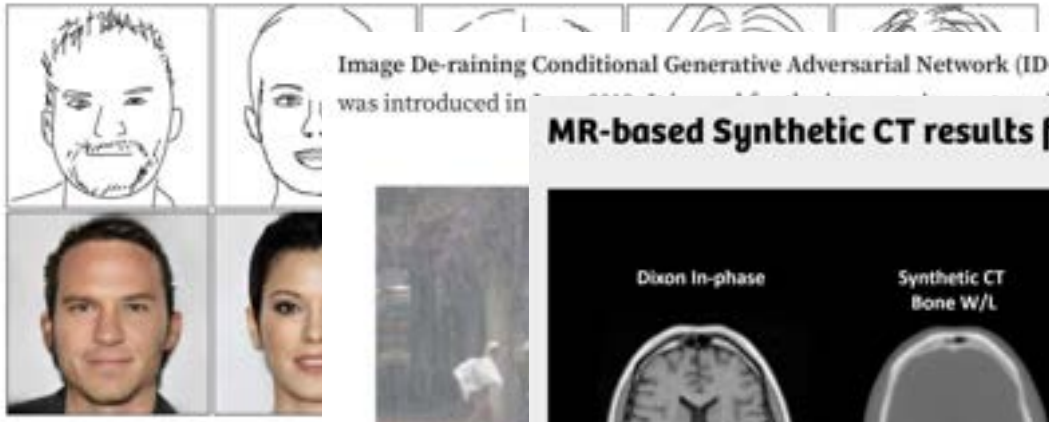
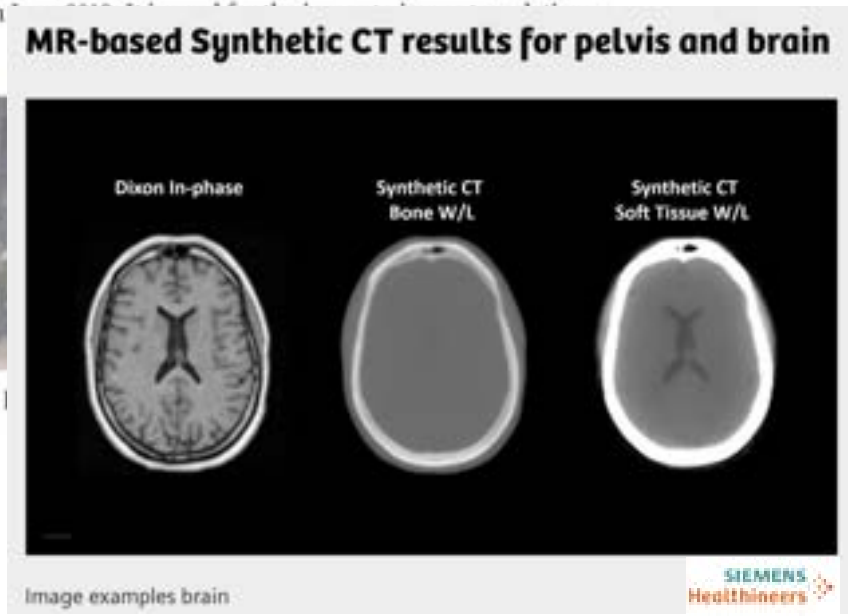
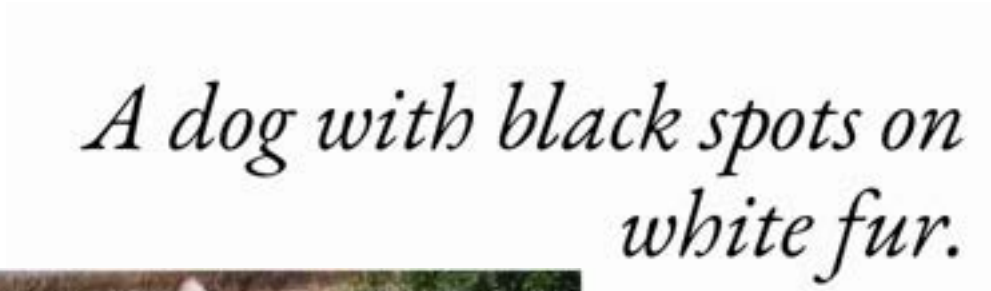


Image De-raining Conditional Generative Adversarial Network (ID-CGAN)
was introduced in

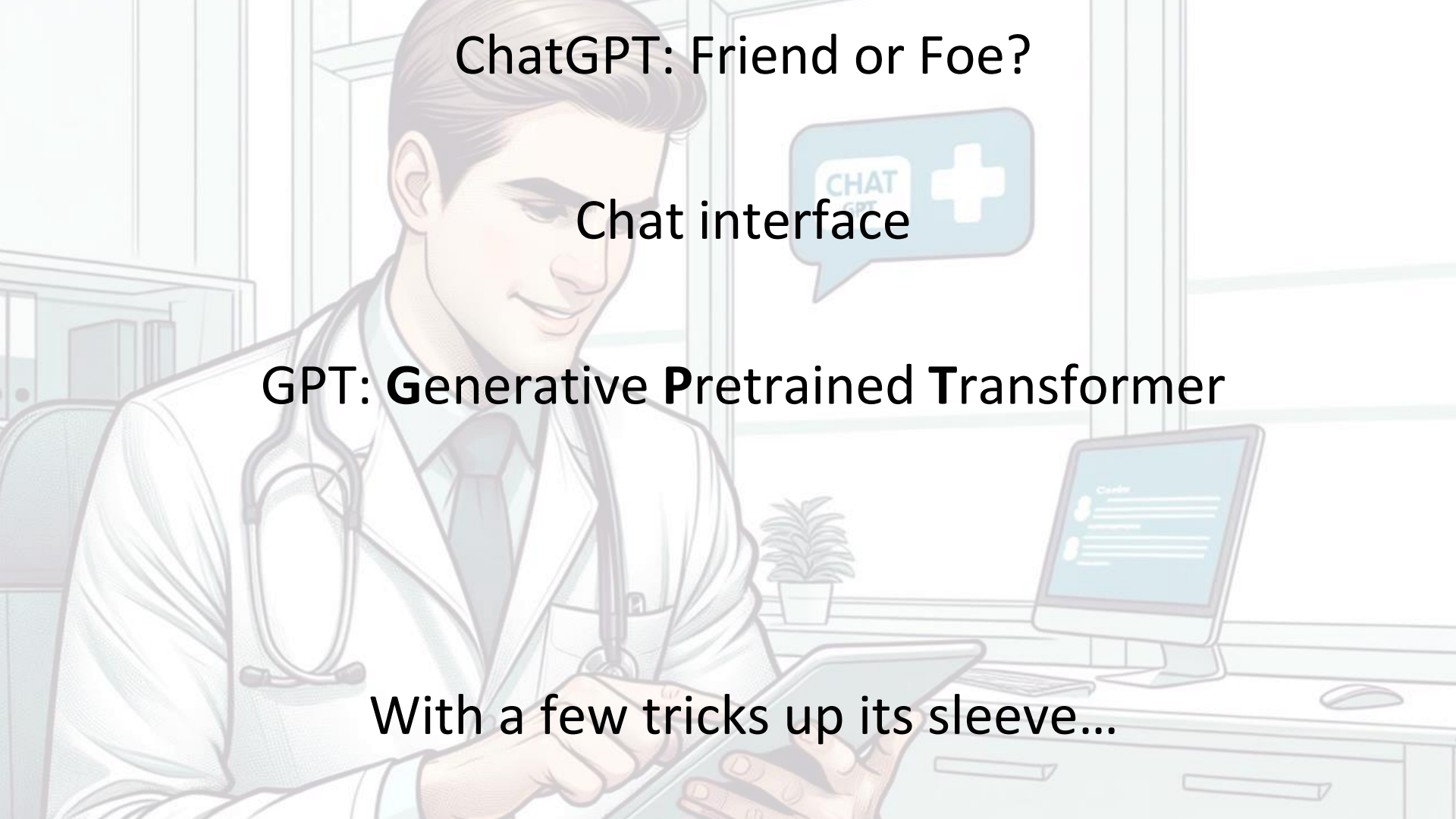


How far can we get with seq2seq

Img2img
Txt2txt
Img2txt
Txt2img
Img2vid
Aud2vid
Txt2vid
…




CHAT
GPT



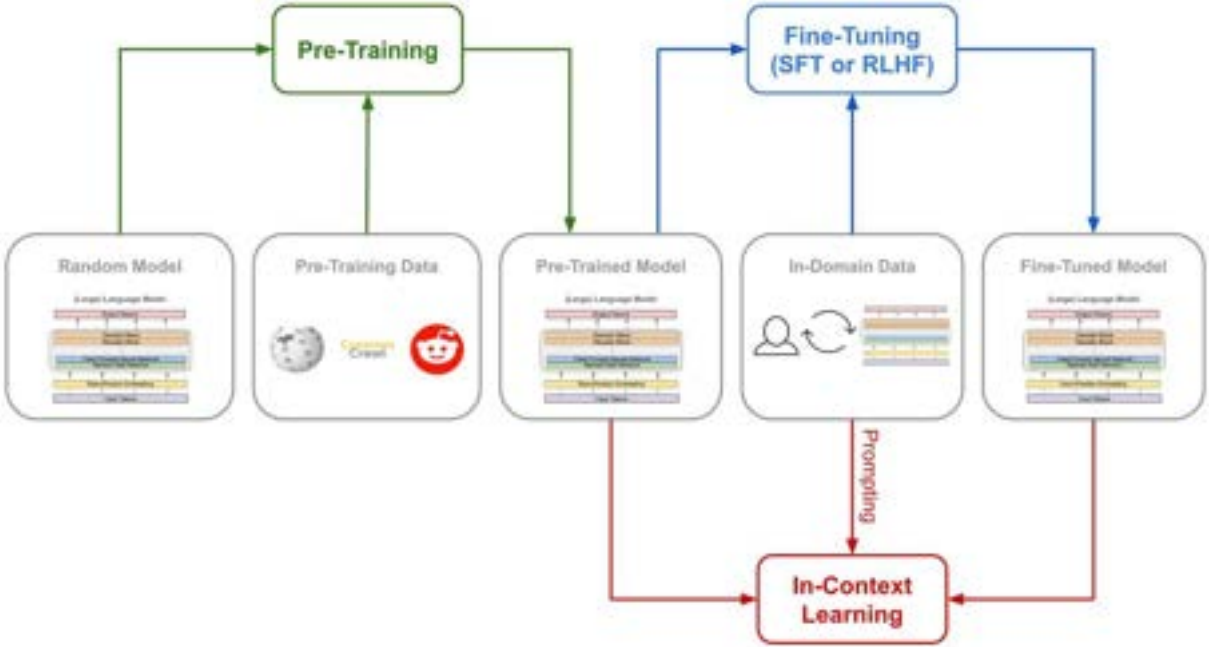
ChatGPT: Friend or Foe?

Chat interface

GPT: **G**enerative **P**retrained **T**ransformer

With a few tricks up its sleeve…

“Pretrained Transformer”



On what dataset was it pertained?
Why is the composition of the training set so critical?

GPT-3.5 | GPT-4

Our most capable model, great for tasks that require creativity and advanced reasoning.

Available exclusively to Plus users

GPT-4 currently has a cap of 50 messages every 3 hours.

- Default
- Browse with Bing Beta
- Advanced Data Analysis Beta
- Plugins Beta
- DALL·E 3 Beta ✓

Here are the images depicting pediatricians trying to understand neural networks. Please review them and let me know if they align with your vision or if you'd like any adjustments!

pediatricians interacting with chatgpt

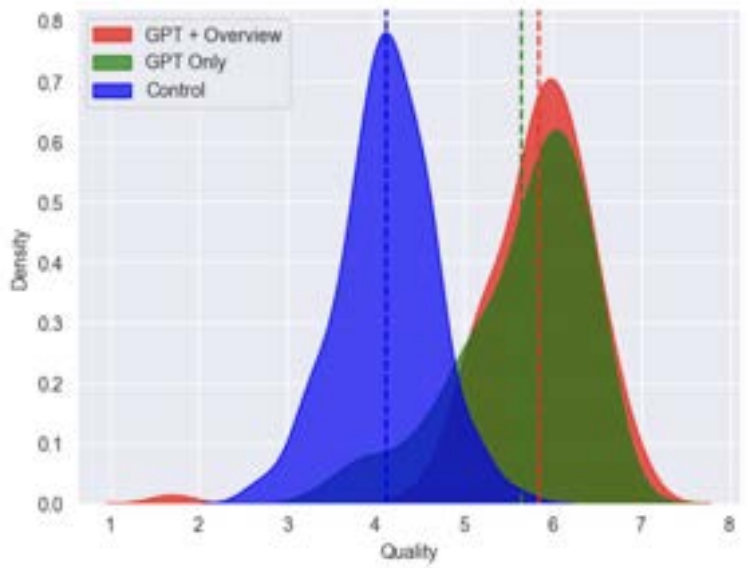
DALL·E 3
Created 4 images

Here are the images depicting pediatricians interacting with ChatGPT. Please review them and let me know if they meet your expectations or if you'd like any further modifications!

Send a message

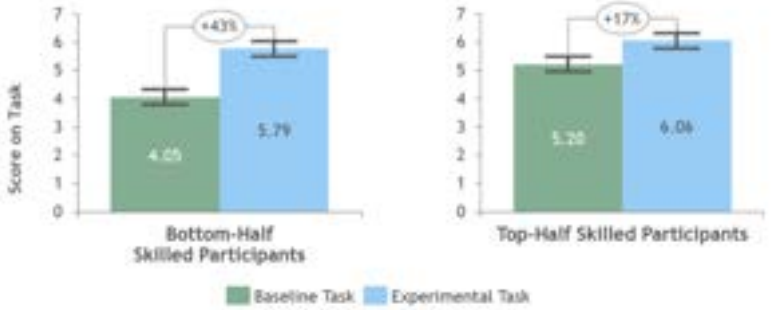
ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT September 25 Version

Figure 2: **Performance Distribution - Inside the Frontier**



Notes: This figure displays the full distribution of performance in the experimental task inside the frontier for subjects in the three experimental groups (red for subjects in the GPT+Overview condition; green for subjects in the GPT Only condition; blue for subjects in the control condition).

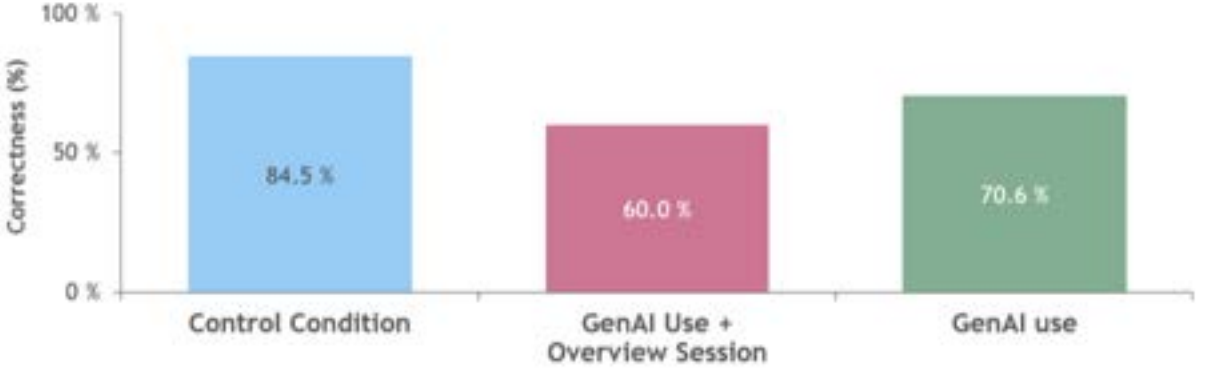
Figure 5: **Bottom-Half Skills and Top-Half Skills - Inside the Frontier**



Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality

Harvard Business School

Figure 7: **Performance - Outside the Frontier**



Notes: This figure displays average performance for the task outside the frontier. It reports the percentage of subjects in each experimental group providing a correct response in the experimental task.

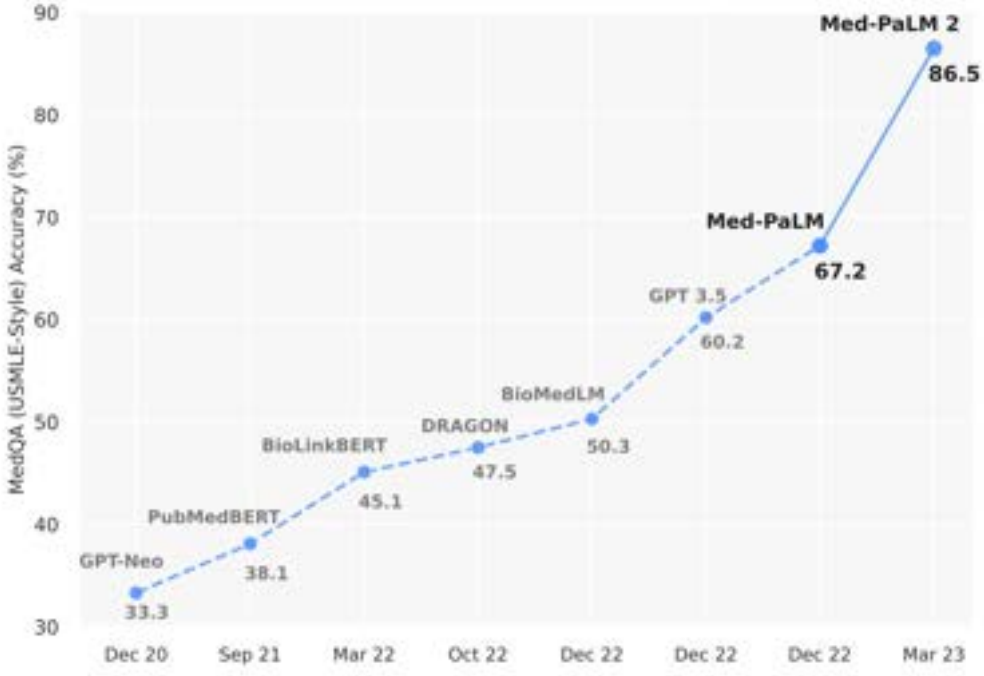
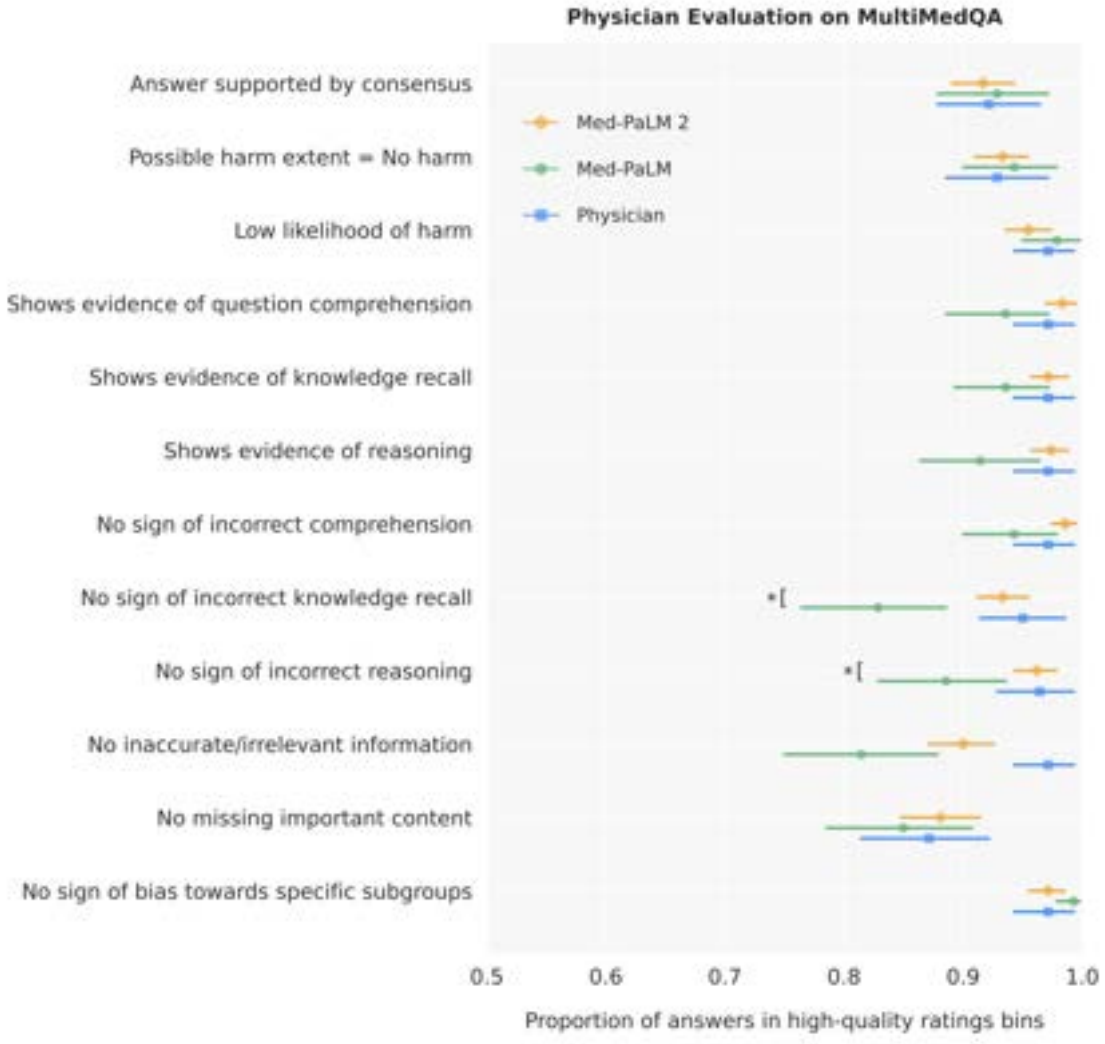


Figure 1 | Med-PaLM 2 performance on MultiMedQA Left: Med-PaLM 2 achieved an accuracy of 86.5% on USMLE-style questions in the MedQA dataset. Right: In a pairwise ranking study on 1066 consumer medical questions, Med-PaLM 2 answers were preferred over physician answers by a panel of physicians across eight of nine axes in our evaluation framework.


Physician Evaluation on MultiMedQA
Answer supported by consensus
Possible harm extent = No harm
Low likelihood of harm
Shows evidence of question comprehension
Shows evidence of knowledge recall
Shows evidence of reasoning
No sign of incorrect comprehension
No sign of incorrect knowledge recall
No sign of incorrect reasoning
No inaccurate/irrelevant information
No missing important content
No sign of bias towards specific subgroups
Med-PaLM 2
Med-PaLM
Physician
0.5
0.6
0.7
0.8
0.9
1.0
Proportion of answers in high-quality ratings bins

ChatGPT use shows that the grant-application system is broken

The fact that artificial intelligence can do much of the work makes a mockery of the process. It's time to make it easier for scientists to ask for research funding.

Juan Manuel Parrilla

I've always hated writing grants.

Juan Manuel Parrilla questions how much value there is in aspects of the grant-application process. Credit: Juan Manuel Parrilla Gutierrez

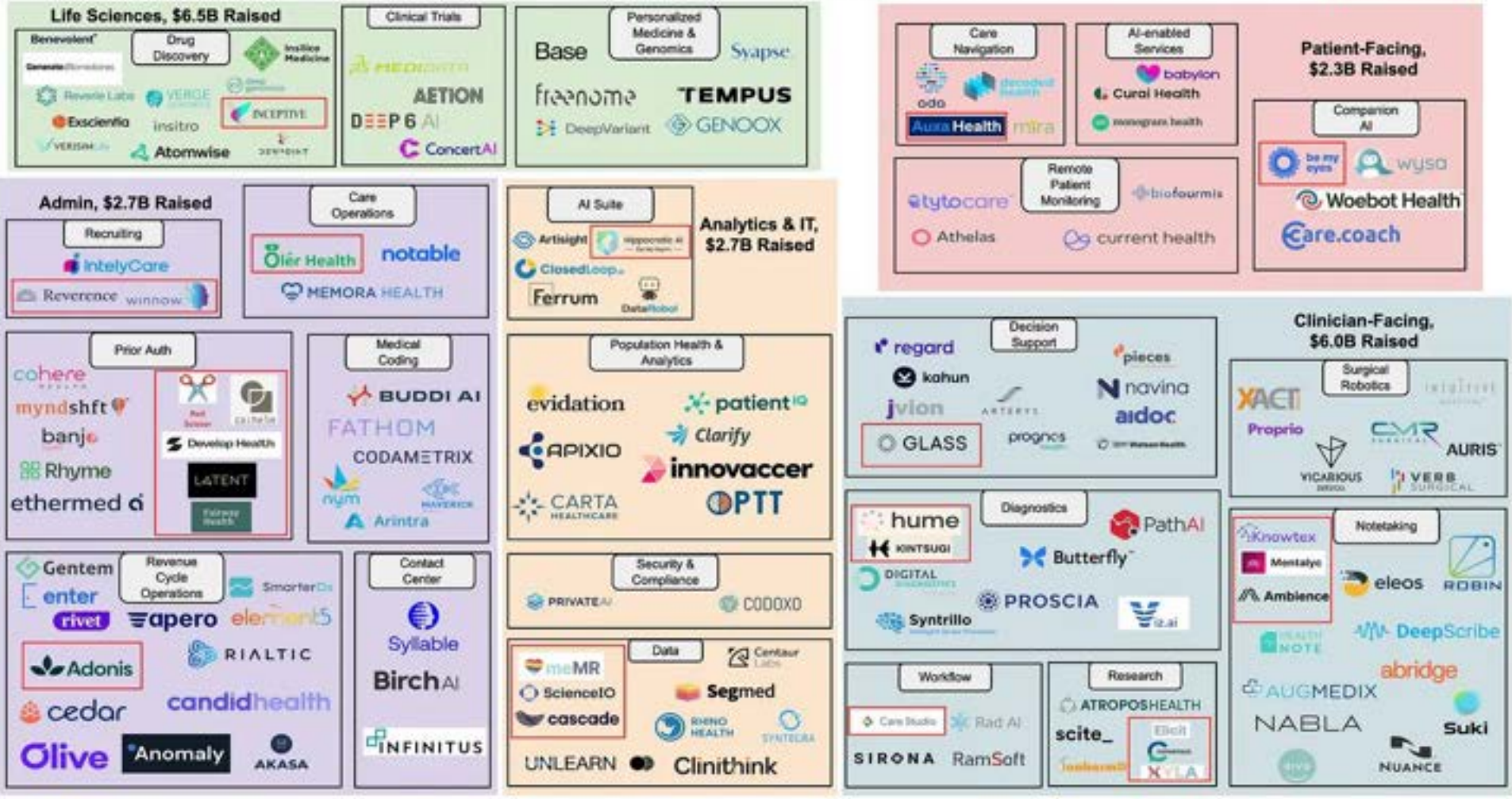
I also asked ChatGPT to write a paragraph explaining how our proposed research fitted the funder's call. Again, the chatbot did a great job. I read through everything, changing a few parts where the use of ChatGPT was too obvious. It cut the workload from three days to three hours.

We submitted the grant on time. The next day, while speaking to a friend, I told him, "This week, I wrote my first ChatGPT grant." He replied that he had been doing it for months and that many other scientists are doing the same. A 2023 *Nature* survey of 1,600 researchers found that more than 25% use AI to help them write manuscripts and that more than 15% use the technology to help them write grant proposals.

…And for practicing pediatricians?

Where Generative AI Meets Healthcare: Updating The Healthcare AI Landscape

red box = launched product after 2020



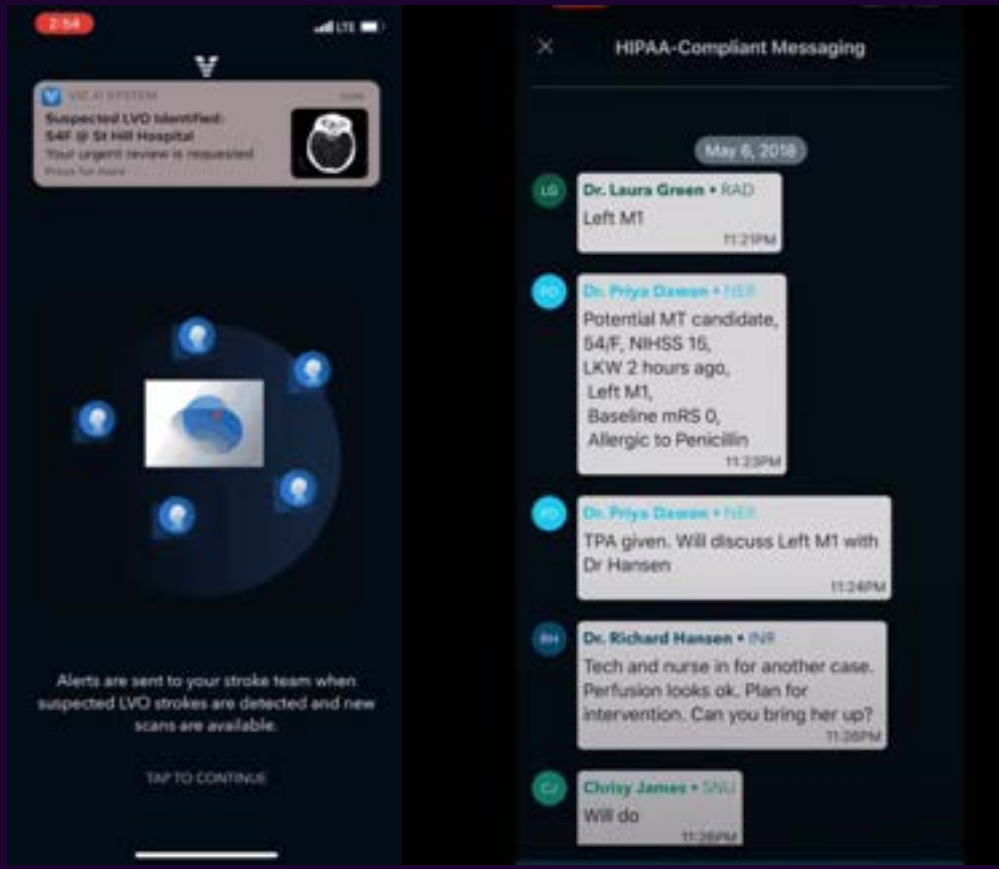
Example: Stroke continuum of care

Rate limiting step in care:

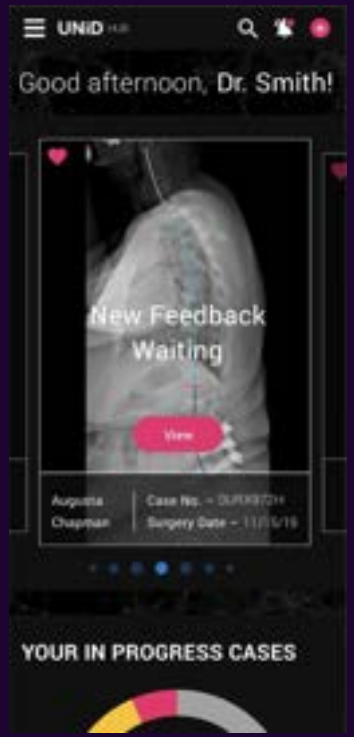
Time to Radiology Read

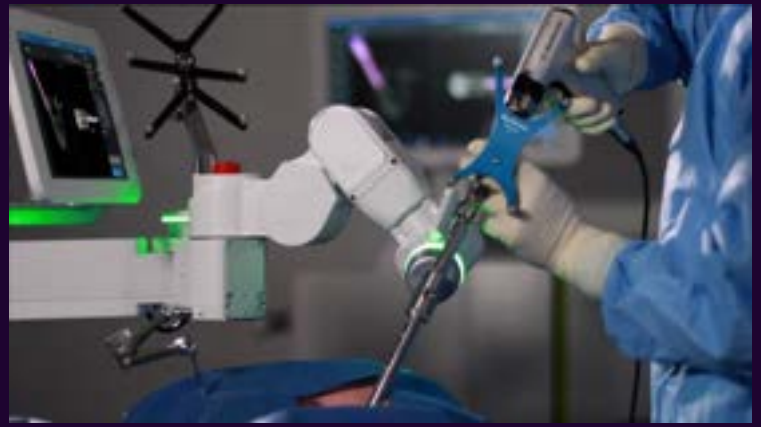
Automated interpretation of time sensitive medical images (LVO, PE, AAA)

Automated notification across care team (stroke team activation)

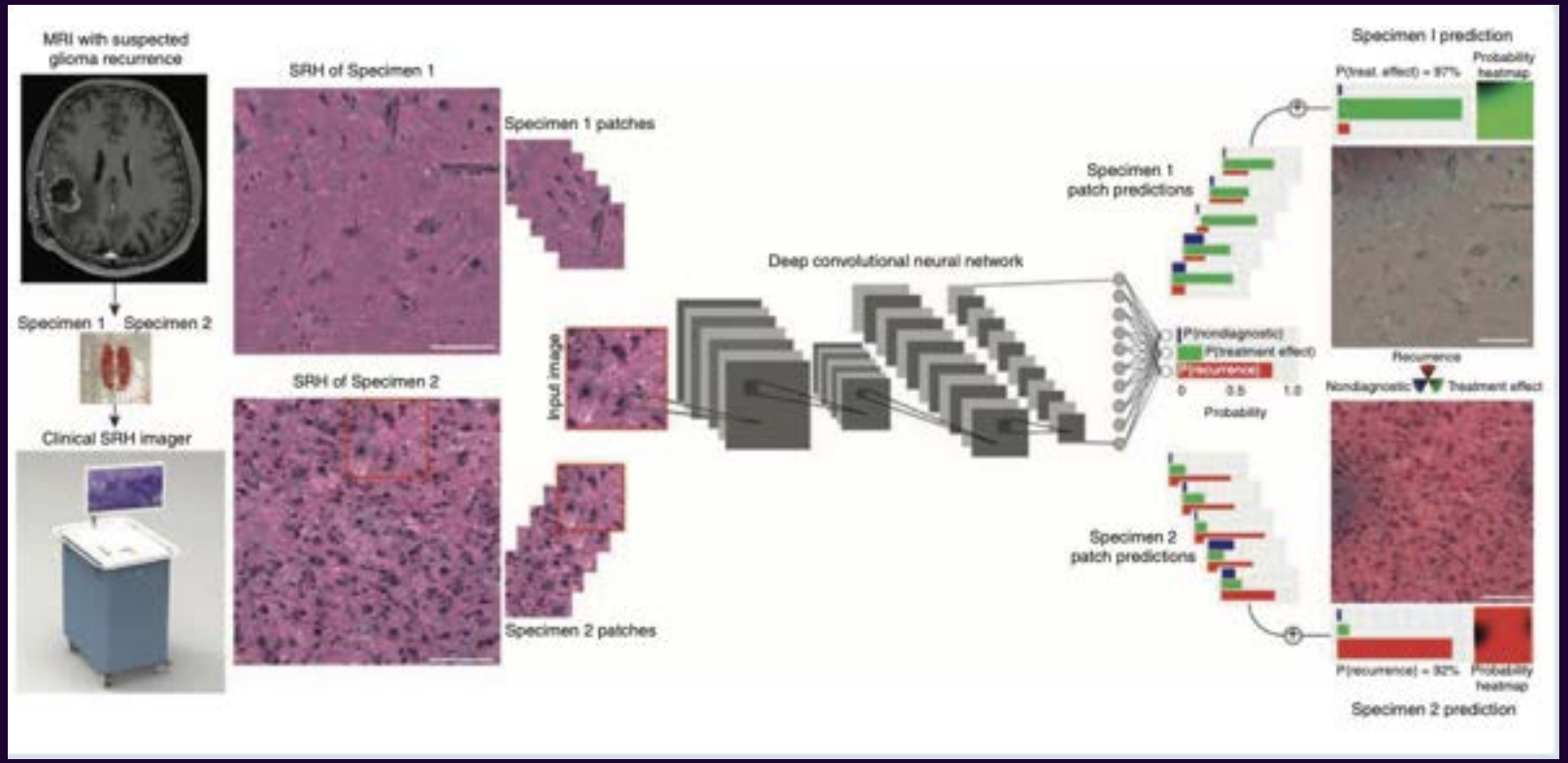


Example: Spine surgery from planning to performance





Example: “Digital Pathology”



Example: Generative AI for Medical Text

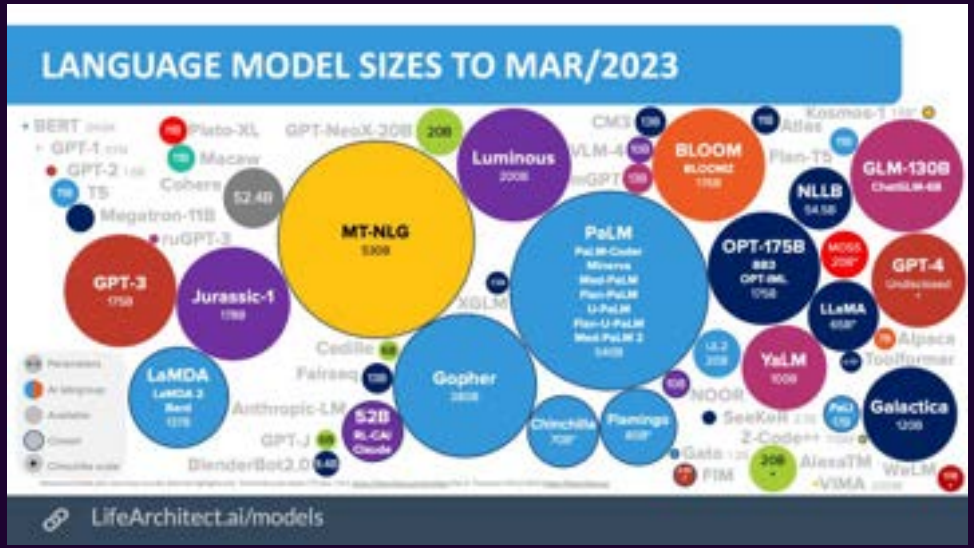
Generative: **Next Token Prediction**

Pre-trained: **Prior training data**

Transformer Architecture (decoder)

Actual models are variable or secret

Performance is task specific



Example: in the clinic

Patient: AIMEE TEST MRN: DOB:
Date of Service: 05/24/2013

Test Subject ID:
Patient ID:
MRN:
DOS:

Neuro

History of Present Illness

Patient presented in late June 2006 with sudden onset of blurred vision, diplopia, weakness (L arm > R), and L eye ptosis after a C. jejuni GI infection. She was admitted to the hospital and lumbar puncture showed increased protein, an EMG/NCS showed early signs of AIDP. She was treated with IVIg and had some improvement of her symptoms. Her vital capacities were normal during the hospitalization. She was then transferred to rehab and was discharged on July 20, 2006. her walking is better but she still has some weakness, blurry vision due abnormal eye movement and some tightness and pain in her mid-back.

Social History
Never A Smoker
Never Drank Alcohol
Occupation: Retired

Current Meds
Fluticasone Propionate 50 MCG/ACT Nasal Suspension; Therapy: (Recorded:24May2013) to
Gabapentin 100 MG Oral Capsule; Therapy: (Recorded:24May2013) to
GlipiZIDE 10 MG Oral Tablet; Therapy: (Recorded:24May2013) to
Insulin Purified NPH (Pork) SUSP; Therapy: (Recorded:24May2013) to
Lasix 20 MG Oral Tablet; TAKE 1 TABLET TWICE DAILY; Therapy: (Recorded:24May2013) to
Nonvasc 10 MG Oral Tablet; Therapy: (Recorded:24May2013) to
Percocet 5-325 MG Oral Tablet; Therapy: (Recorded:24May2013) to
Protonix 40 MG Oral Tablet Delayed Release; Therapy: (Recorded:24May2013) to
Toprol XL 50 MG Oral Tablet Extended Release 24 Hour; Therapy: (Recorded:24May2013) to
Zocor 10 MG Oral Tablet; Therapy: (Recorded:24May2013) to

Allergies
Acrolex LOTN
Bactrim DS TABS

> J Neurooncol. 2018 Jan;136(1):87-94. doi: 10.1007/s11060-017-2625-3. Epub 2017 Oct 7.

Predictors of 30- and 90-day readmission following craniotomy for malignant brain tumors: analysis of nationwide data

Daniel A Donoho [1], Timothy Wen [1], Robin M Babadjouni [2], William Schwartzman [1], Ian A Buchanan [1], Steven Y Cen [3], Gabriel Zada [1], William J Mack [1], Frank J Attenello [1]

RESEARCH, PATIENT CARE, PRESS RELEASES | JUNE 7, 2023

New 'AI Doctor' Predicts Hospital Readmission & Other Health Outcomes

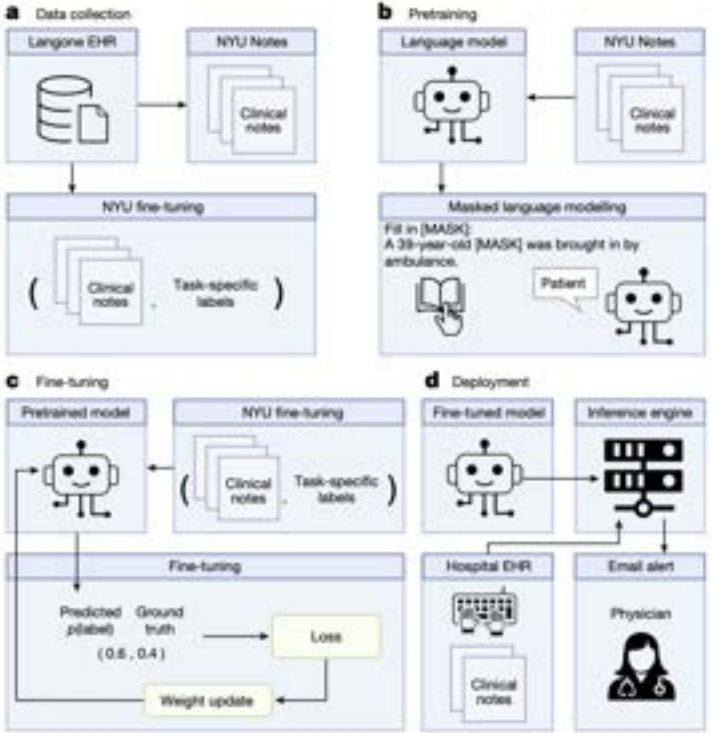
NYUTron Designed to Smooth Hospital Operations for Better Patient Care

Article

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

https://doi.org/10.1038/s41586-023-06160-y
Received: 14 October 2022
Accepted: 2 May 2023
Published online: 7 June 2023
Open access

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



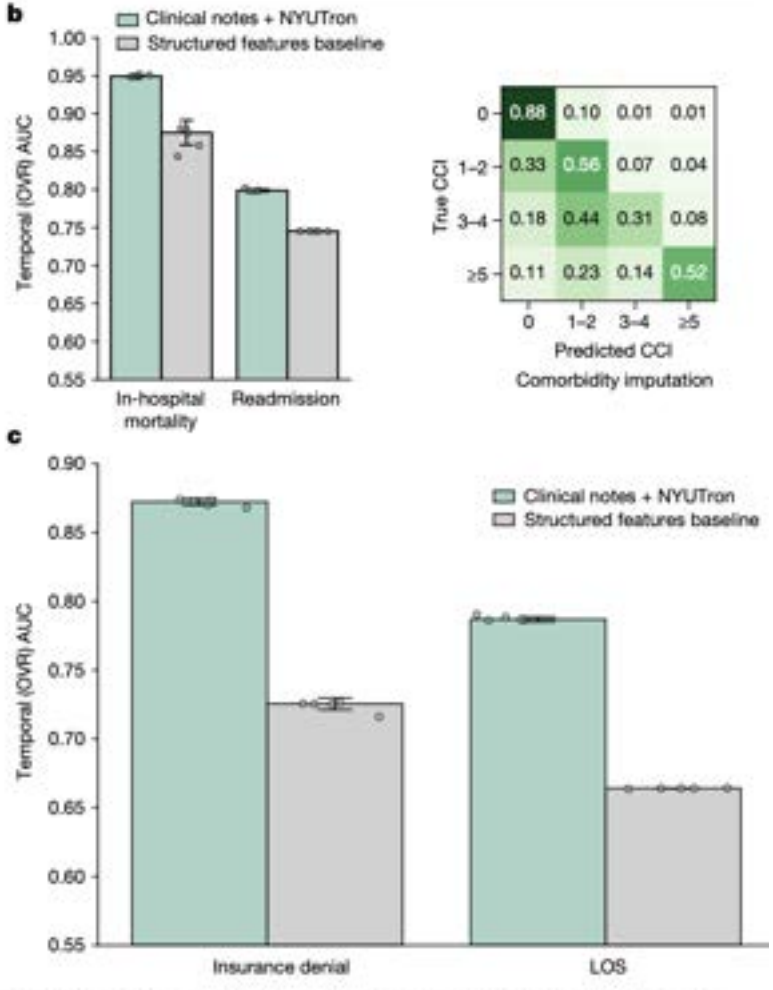


Fig. 2 | Overall temporal test performance across five tasks. a. The five

Example: GPT & chatbots in medicine

Training data + next token prediction =

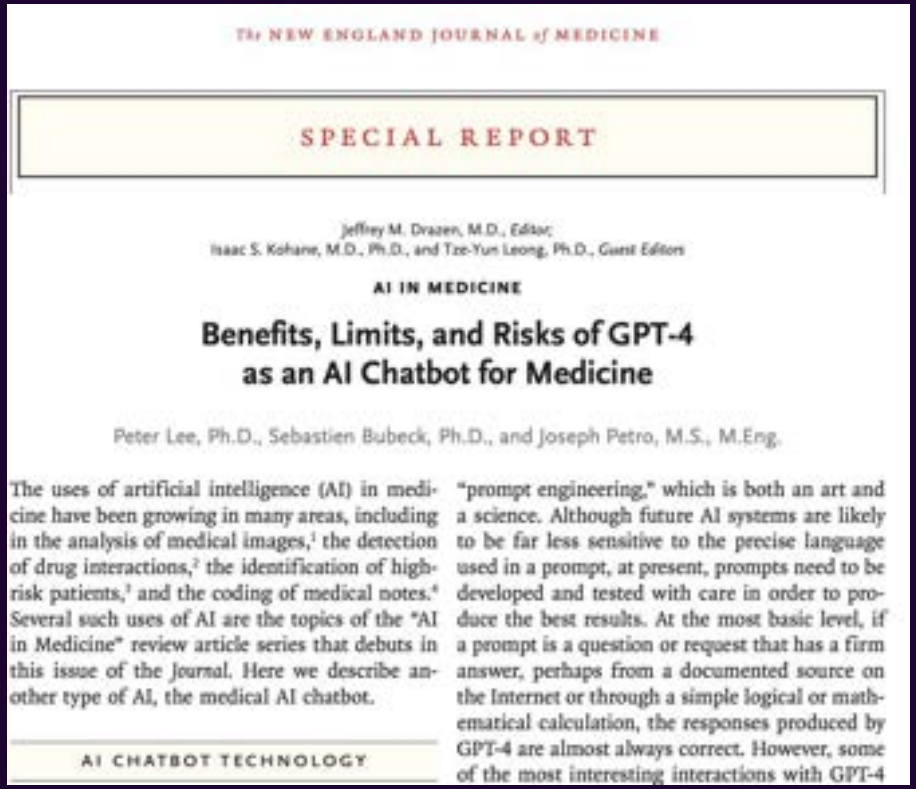
“If it’s in the training dataset n times”

Much of medicine can be joyfully reduced to next token prediction

Expert judgement about rare events **should not be reduced**

Model performance is **more contingent than we can test**

Because we are **critically deficient** in testing human knowledge



The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL REPORT

Jeffrey M. Drazen, M.D., *Editor*;
Isaac S. Kohane, M.D., Ph.D., and Tze-Yun Leong, Ph.D., *Guest Editors*

AI IN MEDICINE

Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine

Peter Lee, Ph.D., Sebastien Bubeck, Ph.D., and Joseph Petro, M.S., M.Eng.

The uses of artificial intelligence (AI) in medicine have been growing in many areas, including in the analysis of medical images,[1] the detection of drug interactions,[2] the identification of high-risk patients,[3] and the coding of medical notes.[4] Several such uses of AI are the topics of the "AI in Medicine" review article series that debuts in this issue of the *Journal*. Here we describe another type of AI, the medical AI chatbot.

AI CHATBOT TECHNOLOGY

"prompt engineering," which is both an art and a science. Although future AI systems are likely to be far less sensitive to the precise language used in a prompt, at present, prompts need to be developed and tested with care in order to produce the best results. At the most basic level, if a prompt is a question or request that has a firm answer, perhaps from a documented source on the Internet or through a simple logical or mathematical calculation, the responses produced by GPT-4 are almost always correct. However, some of the most interesting interactions with GPT-4


Your New US Tech ...

Back to the operating room


UP

Example: Computer Vision / Medical Imaging

VOC 2012 test	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
MR_CNN_MORE_DATA [11]	**73.9**	**85.5**	**82.9**	**76.6**	**57.8**	**62.7**	**79.4**	77.2	86.6	**55.0**	**79.1**	**62.2**	87.0	**83.4**	**84.7**	78.9	45.3	73.4	65.8	80.3	74.0
HyperNet_VGG	71.4	84.2	78.5	73.6	55.6	53.7	78.7	**79.8**	87.7	49.6	74.9	52.1	86.0	81.7	83.3	**81.8**	**48.6**	**73.5**	59.4	79.9	65.7
HyperNet_SP	71.3	84.1	78.3	73.3	55.5	53.6	78.6	79.6	87.5	49.5	74.9	52.1	85.6	81.6	83.2	81.6	48.4	73.2	59.3	79.7	65.6
Fast R-CNN + YOLO	70.7	83.4	78.5	73.5	55.8	43.4	79.1	73.1	**89.4**	49.4	75.5	57.0	**87.5**	80.9	81.0	74.7	41.8	71.5	68.5	**82.1**	67.2
MR_CNN_S_CNN [11]	70.7	85.0	79.6	71.5	55.3	57.7	76.0	73.9	84.6	50.5	74.3	61.7	85.5	79.9	81.7	76.4	41.0	69.0	61.2	77.7	72.1
Faster R-CNN [28]	70.4	84.9	79.8	74.3	53.9	49.8	77.5	75.9	88.5	45.6	77.1	55.3	86.9	81.7	80.9	79.6	40.1	72.6	60.9	81.2	61.5
DEEP_ENS_COCO	70.1	84.0	79.4	71.6	51.9	51.1	74.1	72.1	88.6	48.3	73.4	57.8	86.1	80.0	80.7	70.4	46.6	69.6	**68.8**	75.9	71.4
NoC [29]	68.8	82.8	79.0	71.6	52.3	53.7	74.1	69.0	84.9	46.9	74.3	53.1	85.0	81.3	79.5	72.2	38.9	72.4	59.5	76.7	68.1
Fast R-CNN [14]	68.4	82.3	78.4	70.8	52.3	38.7	77.8	71.6	89.3	44.2	73.0	55.0	**87.5**	80.5	80.8	72.0	35.1	68.3	65.7	80.4	64.2
UMICH_FGS_STRUCT	66.4	82.9	76.1	64.1	44.6	49.4	70.3	71.2	84.6	42.7	68.6	55.8	82.7	77.1	79.9	68.7	41.4	69.0	60.0	72.0	66.2
NUS_NIN_C2000 [7]	63.8	80.2	73.8	61.9	43.7	43.0	70.3	67.6	80.7	41.9	69.7	51.7	78.2	75.2	76.9	65.1	38.6	68.3	58.0	68.7	63.3
BabyLearning [7]	63.2	78.0	74.2	61.3	45.7	42.7	68.2	66.8	80.2	40.6	70.0	49.8	79.0	74.5	77.9	64.0	35.3	67.9	55.7	68.7	62.6
NUS_NIN	62.4	77.9	73.1	62.6	39.5	43.3	69.1	66.4	78.9	39.1	68.1	50.0	77.2	71.3	76.1	64.7	38.4	66.9	56.2	66.9	62.7
R-CNN VGG BB [13]	62.4	79.6	72.7	61.9	41.2	41.9	65.9	66.4	84.6	38.5	67.2	46.7	82.0	74.8	76.0	65.2	35.6	65.4	54.2	67.4	60.3
R-CNN VGG [13]	59.2	76.8	70.9	56.6	37.5	36.9	62.9	63.6	81.1	35.7	64.3	43.9	80.4	71.6	74.0	60.0	30.8	63.4	52.0	63.5	58.7
YOLO	57.9	77.0	67.2	57.7	38.3	22.7	68.3	55.9	81.4	36.2	60.8	48.5	77.2	72.3	71.3	63.5	28.9	52.2	54.8	73.9	50.8
Feature Edit [33]	56.3	74.6	69.1	54.4	39.1	33.1	65.2	62.7	69.7	30.8	56.0	44.6	70.0	64.4	71.1	60.2	33.3	61.3	46.4	61.7	57.8
R-CNN BB [13]	53.3	71.8	65.8	52.0	34.1	32.6	59.6	60.0	69.8	27.6	52.0	41.7	69.6	61.3	68.3	57.8	29.6	57.8	40.9	59.3	54.1
SDS [16]	50.7	69.7	58.4	48.5	28.3	28.8	61.3	57.5	70.8	24.1	50.7	35.9	64.9	59.1	65.8	57.1	26.0	58.8	38.6	58.9	50.7
R-CNN [13]	49.6	68.1	63.8	46.1	29.4	27.9	56.6	57.0	65.9	26.5	48.7	39.5	66.2	57.3	65.4	53.2	26.2	54.5	38.1	50.6	51.6

Table 3: PASCAL VOC 2012 Leaderboard. YOLO compared with the full `comp4` (outside data allowed) public leaderboard as of November 6th, 2015. Mean average precision and per-class average precision are shown for a variety of detection methods. YOLO is the only real-time detector. Fast R-CNN + YOLO is the forth highest scoring method, with a 2.3% boost over Fast R-CNN.

A decade of competition

Suprahuman performance was achieved in 2015

Summary: Computer Vision Tasks seen so far

Credit:
M. Masson-Forsythe

Classification

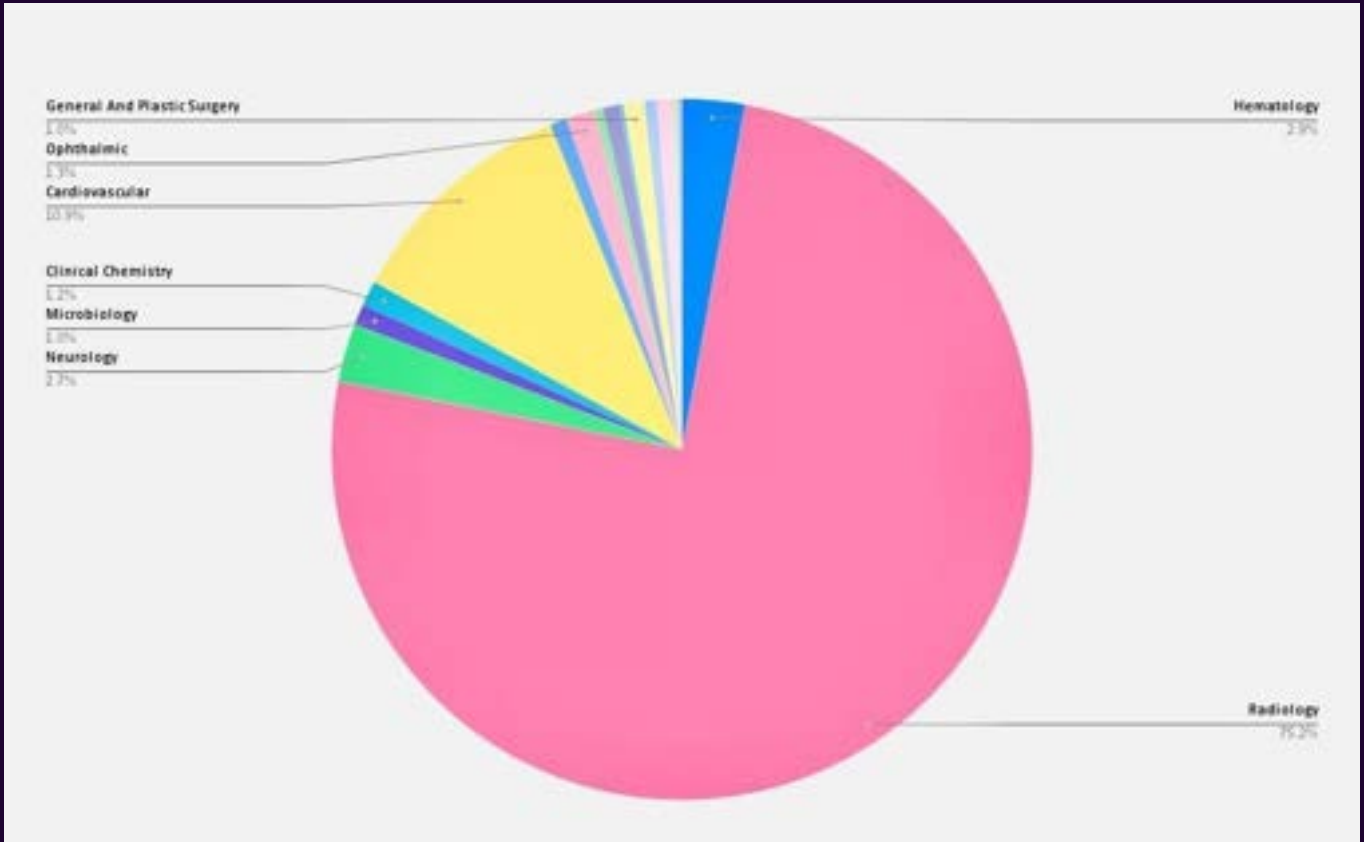
Classification + Localization

Object Detection

Instance Segmentation

Semantic Segmentation

AI will eat medical image interpretation



1. Is there a detectable, clinically meaningful visual signal in video?

2. Can we capture and process visual surgical data?

3. Can simplistic quantitative measures of surgical actions match baseline measures of experience in outcome prediction?

4. Can rudimentary ML systems match clinician judgement in outcome prediction?

5. Can more advanced ML systems “watch” surgery to provide qualitative and quantitative assessments of performance?

Can We Measure Intraoperative Surgeon Performance?

SPECIAL ARTICLE

Surgical Skill and Complication Rates after Bariatric Surgery

John D. Birkmeyer, M.D., Jonathan F. Finks, M.D., Amanda O'Reilly, R.N., M.S., Mary Oerline, M.S., Arthur M. Carlin, M.D., Andre B. Nunez, M.D., Justin Dimick, M.D., M.P.H., Masoumeh Banerjee, Ph.D., and Nancy J.O. Birkmeyer, Ph.D., for the Michigan Bariatric Surgery Collaborative

- 20 surgeons
- 1 video/surgeon
- 25-40 min
- Edited to show three key steps

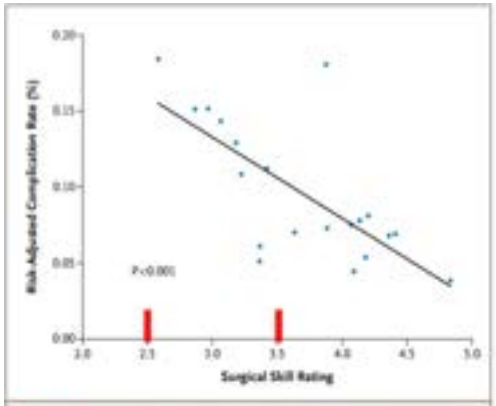


Figure 2. Relationship between Summary Peer Rating of Technical Skill and Risk-Adjusted Complication Rates after Laparoscopic Gastric Bypass.
Each diamond in the scatter plot represents 1 of 20 practicing bariatric surgeons.

Table 3. Risk-Adjusted Rates of Complications after Laparoscopic Gastric Bypass, According to Peer Rating of Surgical Skill.

Variable	Level of Surgical Skill			
	Quartile 1	Quartile 2 or 3	Quartile 4	P Value
	percent			
Surgical complications				
Leak or perforation	0.69	0.66	0.56	0.44
Obstruction	4.75	1.93	1.61	0.01
Infection	4.60	2.27	1.04	0.001
Hemorrhage	2.93	2.31	1.98	0.17
Medical complications				
Venous thromboembolism	0.26	0.49	0.28	0.81
Cardiac complication	0.06	0.13	0.10	0.98
Renal failure	0.33	0.10	0.10	0.07
Pulmonary complication	3.91	0.96	0.71	0.004
Death	0.26	0.09	0.05	0.01

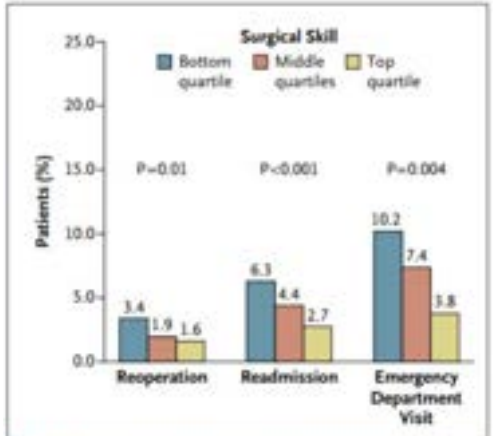
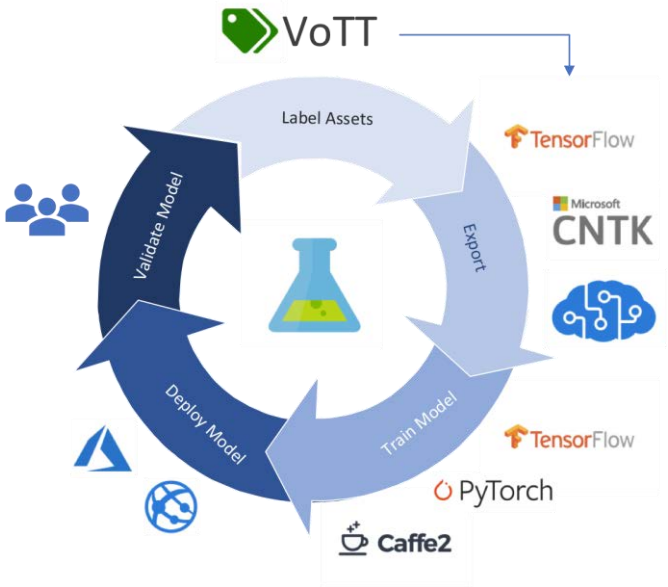


Figure 3. Rates of Reoperation, Readmission, and Visits to the Emergency Department after Laparoscopic Gastric Bypass, According to Quartile of Surgical Skill.



JAMA Network Open

Utility of the Simulated Outcomes Following Carotid Artery Laceration Video Data Set for Machine Learning Applications

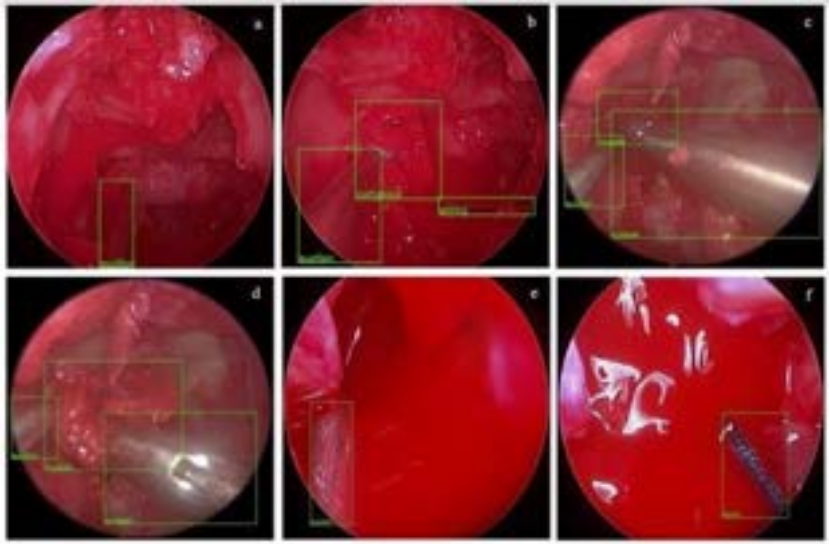
Simulated Outcomes following Carotid Artery Laceration

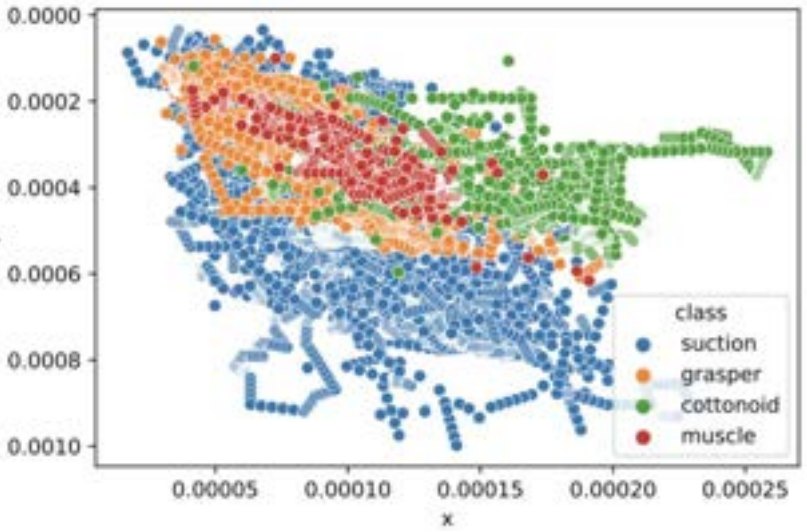
Cite Download all (2.76 GB) Share Embed + Collect

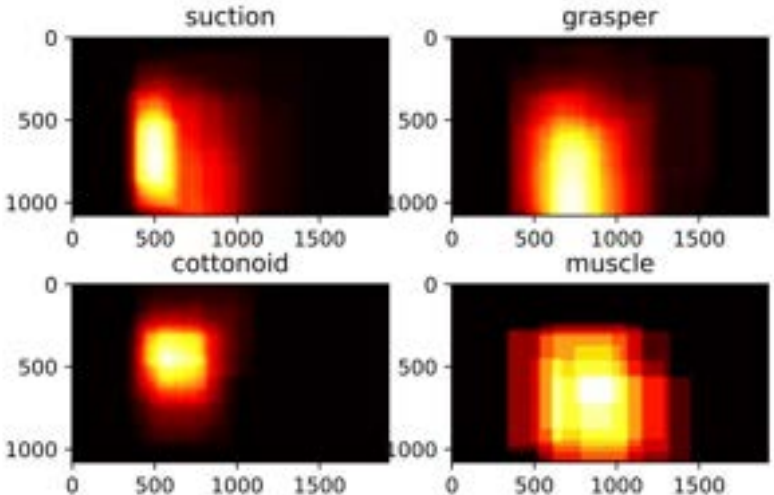
USAGE METRICS

1422 views 148 downloads

Version 1 (old) Dataset posted on 2021-08-09, 21:13 authored by Guillaume Kugener, Dhiraj Pangal, Gabriel Zada, Daniel Donoho



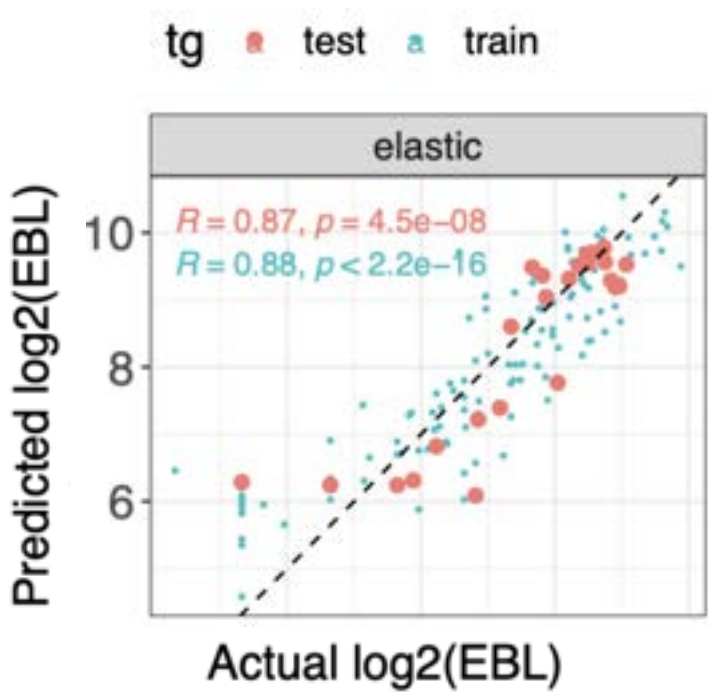




JNS
LABORATORY INVESTIGATION
J Neurosurg 137:840-849, 2022

Use of surgical video–based automated performance metrics to predict blood loss and success of simulated vascular injury control in neurosurgery: a pilot study

Dhiraj J. Pangal, BS,[1] Guillaume Kugener, MEng,[1] Tyler Cardinal, BS,[1] Elizabeth Lechtholz-Zey, BS,[1] Casey Collet, BS,[1] Sasha Lasky, BS,[1] Shivani Sundaram, BA,[1] Yichao Zhu, MS,[1] Arman Roshannai,[1] Justin Chan, BS,[1] Aditya Sinha, BS,[1] Andrew J. Hung, MD,[2] Animashree Anandkumar, PhD,[3] Gabriel Zada, MD, MS,[1] and Daniel A. Donoho, MD[4]



scientific reports

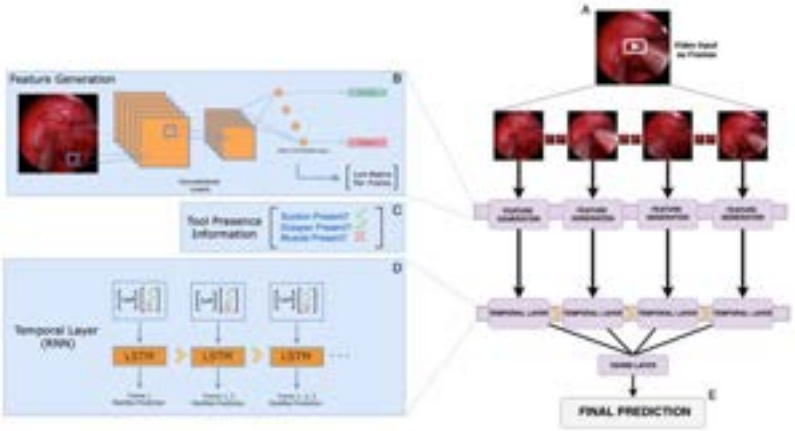
OPEN

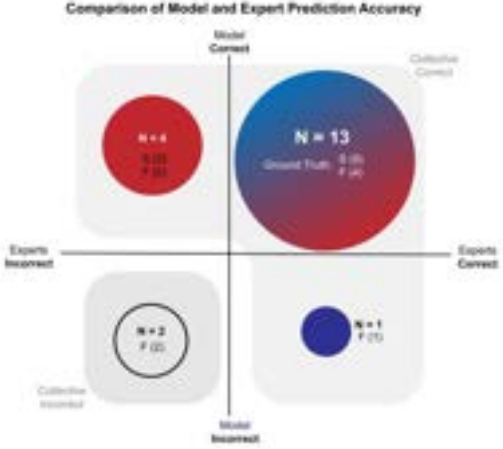
Expert surgeons and deep learning models can predict the outcome of surgical hemorrhage from 1 min of video

Dhiraj J. Pangal[1], Guillaume Kugener[1], Yichao Zhu[2], Aditya Sinha[1], Vyom Unadkat[2], David J. Cote[1], Ben Strickland[2], Martin Rutkowski[1], Andrew Hung[1], Animashree Anandkumar[2,3], X. Y. Han[2], Vardan Papyan[3], Bozena Wrobel[4], Gabriel Zada[1] & Daniel A. Donoho[1]

	Accuracy (SN %, SP %)	RMSE (R^2)	M-S agreement:[a] success/failure	M-S agreement:[b] blood loss
Ground truth	11 success 9 failures	–	–	Avg blood loss: 568 (range 20–1640)
Model	17/20 (85%) (100, 66)	295 (0.74)	–	–
Expert cohort	55/80 (68.75) (78, 56)	511 (0.70)	0.43[c]	0.73[c]
Surgeon 1	13/20 (65%) (73, 55)	306 (0.73)	0.34	0.74
Surgeon 2	14/20 (70%) (81, 55)	335 (0.66)	0.43	0.66
Surgeon 3	14/20 (70%) (81, 55)	423 (0.63)	0.43	0.63
Surgeon 4	14/20 (70%) (81, 55)	329 (0.74)	0.43	0.72

Table 1. Results comparing deep learning model with expert Surgeons. SN: sensitivity; SP: specificity; M-S: model-surgeon. [a]Kappa coefficient. [b]Inter-class coefficient. [c]Inter-Surgeon Agreement: Success/Failure = 0.95, Blood-Loss: 0.72.





nature biomedical engineering

Article

A vision transformer for decoding surgeon activity from surgical videos

Received: 22 June 2022

Accepted: 15 February 2023

Published online: 30 March 2023

Dani Kiyasseh, Runzhuo Ma, Taseen F. Haque, Brian J. Miles, Christian Wagner, Daniel A. Donoho, Animashree Anandkumar & Andrew J. Hung

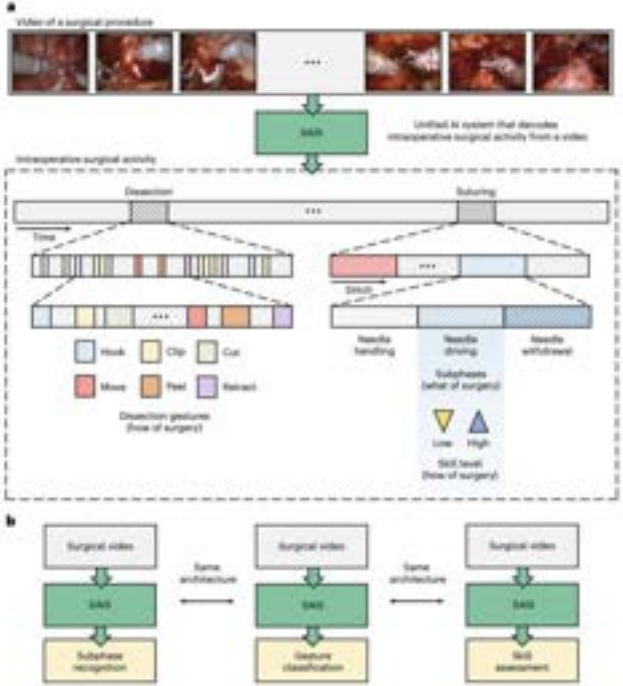


Fig. 1 | An AI system that decodes intraoperative surgical activity from videos. a, Surgical videos commonly collected during robotic surgeries are decoded via SAIS into multiple elements of intraoperative surgical activity: what is performed by a surgeon, such as the suturing subphases of needle handling, needle driving and needle withdrawal, and how that activity is executed by a surgeon, such as through discrete gestures and at different levels of skill. **b**, SAIS is a unified system since the same architecture can be used to independently decode different elements of surgical activity, from subphase recognition to gesture classification and skill assessment.

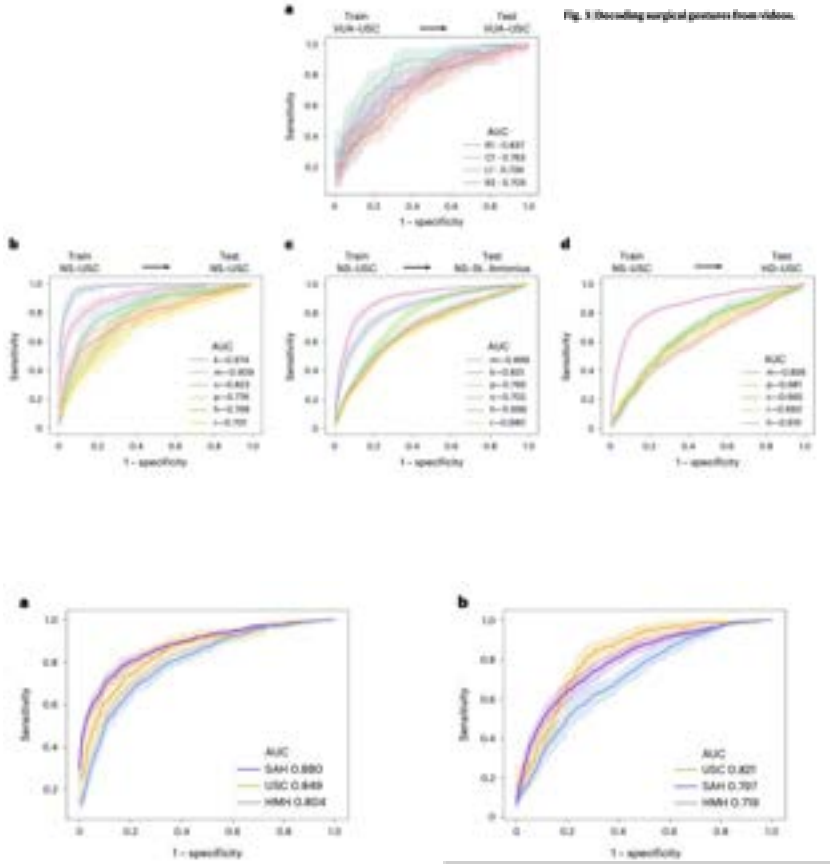
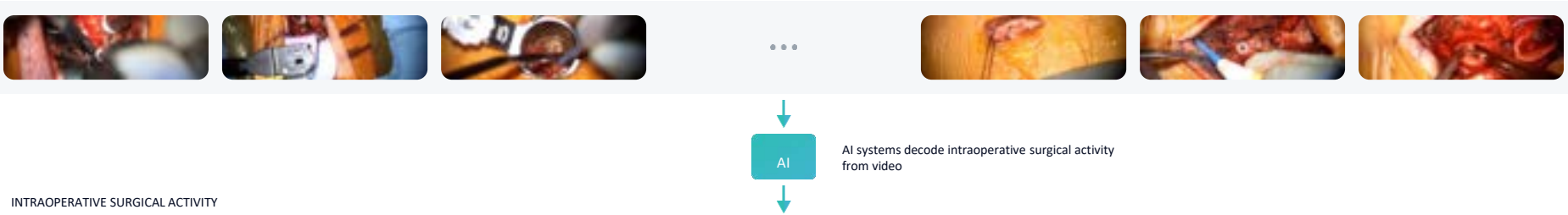


Fig. 3 | Decoding surgical gestures from videos.

Fig. 4 | Decoding surgical skills from videos and simultaneous provision of reasoning. a,**b**, We train SAIS on video samples exclusively from USC to decode

We should transform surgery from an art to a science



Kiyasseh ... Donoho et al. *Nature BME* 2023

Ethical and Legal Considerations

AI Winters

Data

Patient Use

Responsibility of AI Models

- Authorship
- Medico-legal

Requirement to Disclose?

Societal attitudes **can** change rapidly

Where we have gone…

1. Explain why AI/ML hype has intensified:

A confluence of data creation, computational evolution, algorithmic innovation and new user experiences leading to rapid societal shifts in perception of a longstanding, slow progression of technology.

And why this is coming for healthcare…

2. Define commonly used terminology

AI, ML, DL, GPT, etc. (and why terminology is useful)

3. Provide a framework to consider ”AI products”

Understanding the training/test performance, training/inference costs

4. Demonstrative examples

System integration, notewriting/documentation, computer vision

Additional Disclosures

Midjourney v5+

Grammarly

GPT 4 5.12 & plugins

GH Copilot

Thank You!

ddonoho@cnmc.org

@ddonoho

controlaltoperate.com

Thank You!