Foundation Models for Biomedical NLP

Elliot Bolton

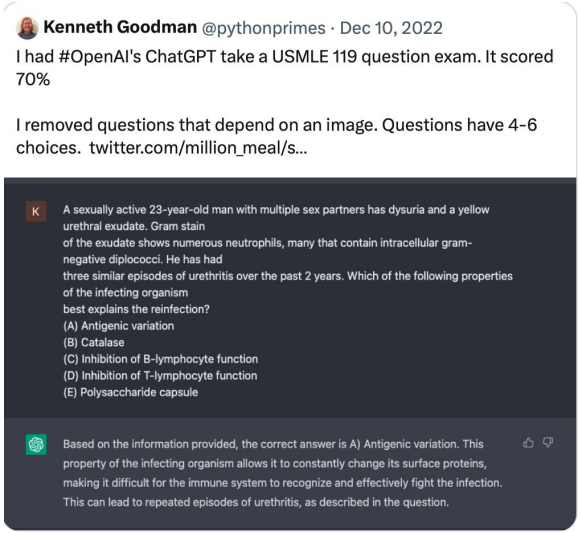
Center for
Research on
Foundation
Models

BioMedLM Team

Stanford CRFM: Elliot Bolton, Michiro Yasunaga, David Hall, Tony Lee, Chris Manning, Percy Liang

MosaicML: Abhinav Venigalla, Jonathan Frankle, Michael Carbin

ChatGPT Passes Practice USMLE

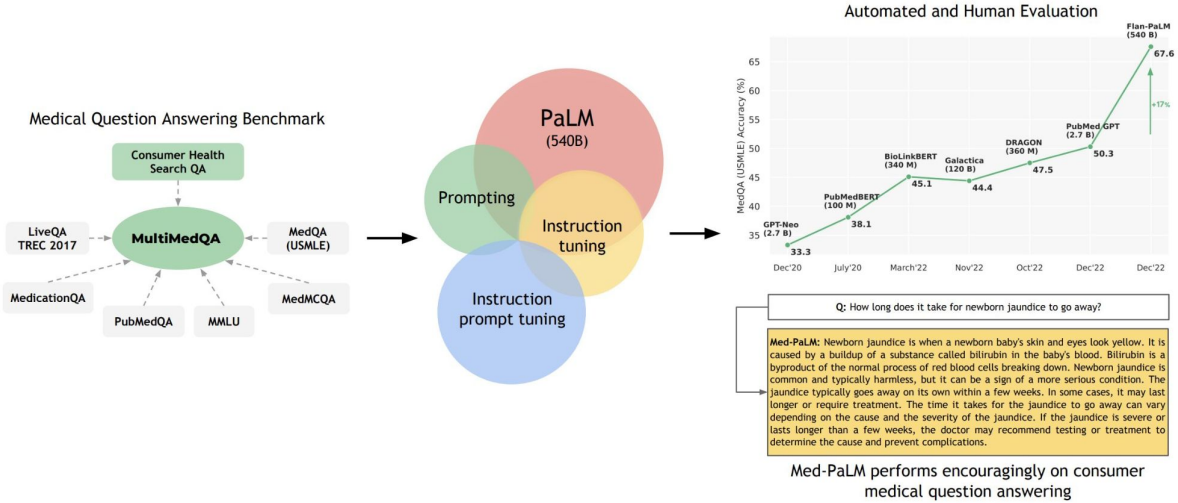


[@pythonprimes tweet on Dec 10, 2022]

BioMedLM Released, Sets New SOTA For MedQA

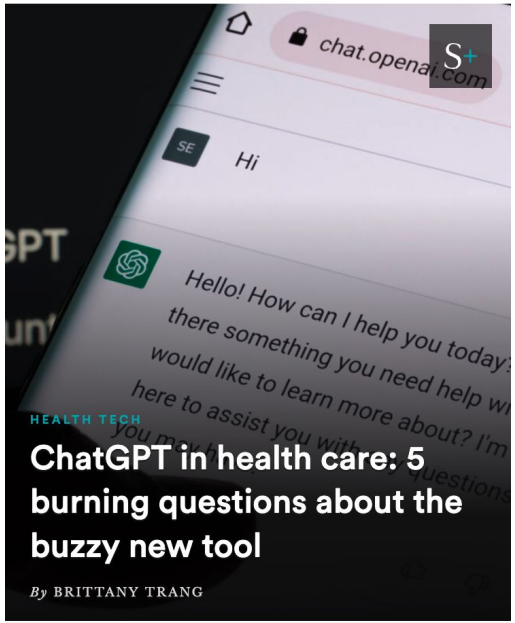
https://huggingface.co/stanford-crfm/BioMedLM

Med-PaLM Released, Sets New SOTA For MedQA



[@vivnat tweet on Dec 26, 2022]

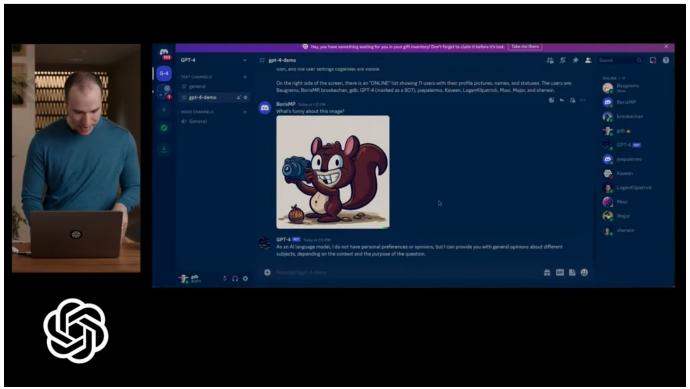
Lots Of Interest In Applying LLMs To Healthcare



STAT News - January 2023

Med-PaLM 2 And GPT-4 Released Last Week

[The Check Up With Google Health 2023]



[GPT-4 Developer Livestream]

Example MedQA Question

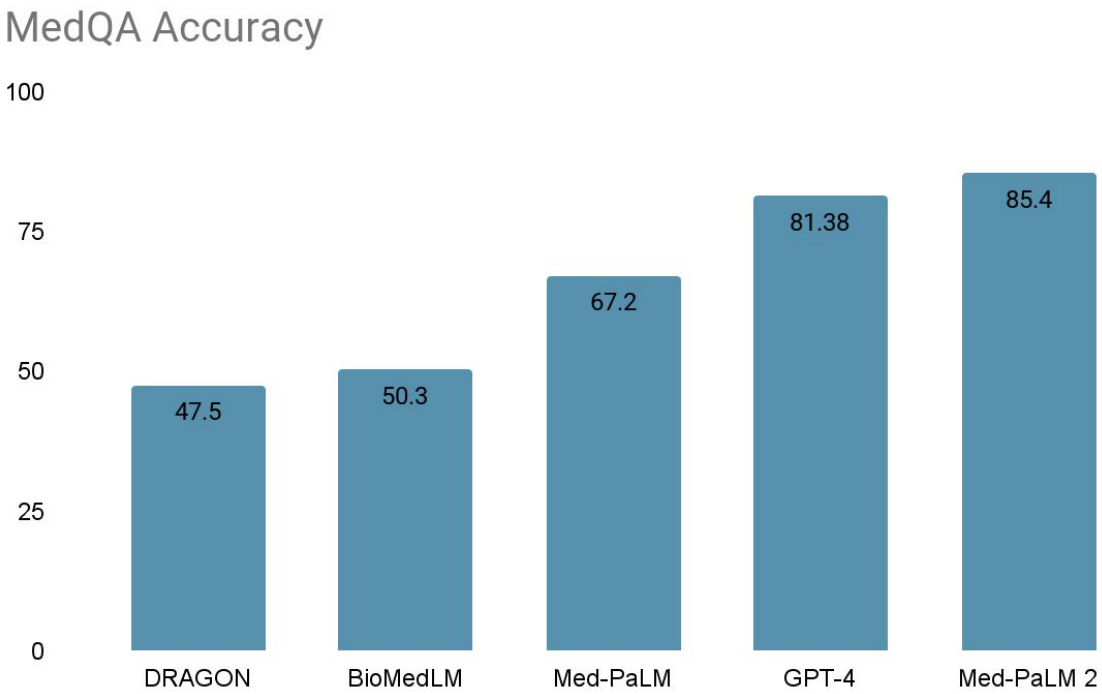
Question:

A 45-year-old woman presents to the emergency department with acute onset of severe right upper quadrant abdominal pain that radiates to the infrascapular region. Her medical history is significant for obesity, hypertension, obstructive sleep apnea, and gastric bypass surgery 2 years ago after which she lost 79 kg (150 lb). The patient complains of nausea and vomiting that accompanies the pain. Her temperature is 38.9°C (101.2°F), blood pressure is 144/88 mm Hg, heart rate is 76/min, and respiratory rate is 14/min (fever). Abdominal examination is significant for right upper quadrant tenderness along with guarding and cessation of inspired breath on deep palpation of the right upper quadrant. Which test should be ordered first for this patient?

A) Abdominal ultrasound
B) CT scan of the abdomen
C) Hepato-iminodiacetic acid scan
D) MRI of the abdomen
E) X-ray film of the abdomen

Jin+2020

Recent Performance On MedQA Task



Healthcare/Biomedical QA - Caveats

MedQA removes one option from source USMLE question, no images

The Med-PaLM 92.6% included questions such as “What do night sweats indicate?”

Training set contamination and artifacts in QA benchmarks

BioMedLM - Basics

Standard GPT-style model with 2.7B parameters, custom biomedical tokenizer

Trained from scratch on 300B tokens of PubMed data on MosaicML cluster

BioMedLM - Pretraining Insights

Improved performance on domain specific tasks when pre-training in-domain

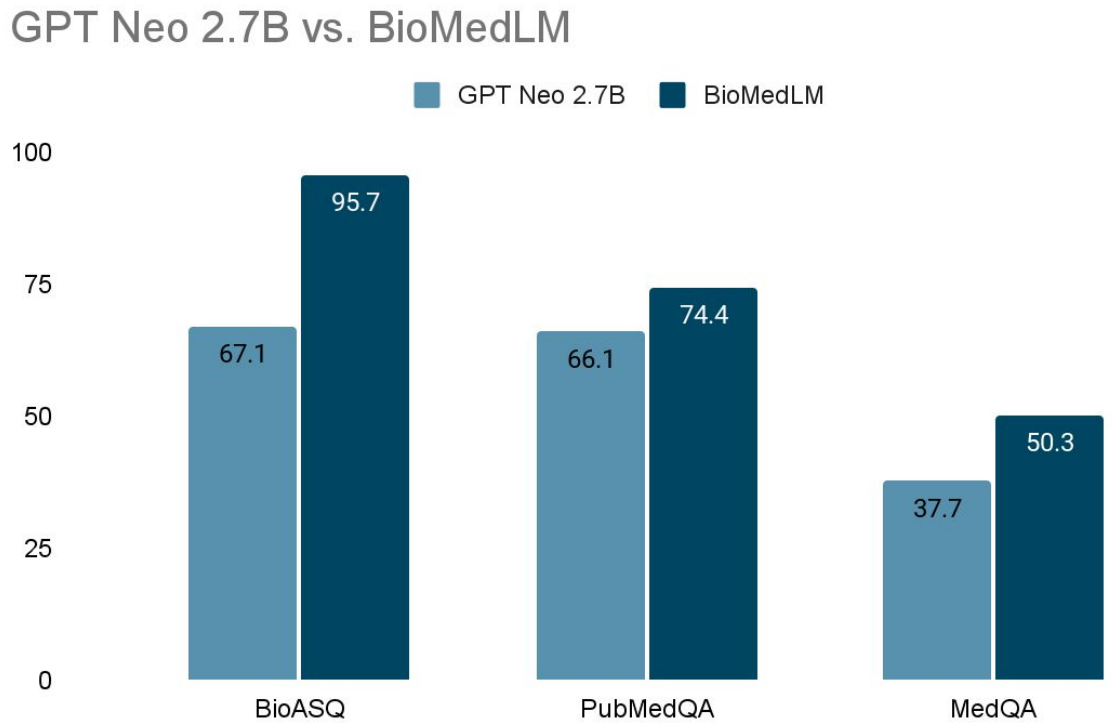
Domain specific tokenizer boosts performance

Continued downstream task performance benefits throughout 300B tokens

BioMedLM - Biomedical Tokenizer

BioMedLM Tokenizer	GPT-2 Tokenizer
chromatography	chrom/atography
GATA	G/ATA
Immunohistochemistry	Immun/oh/ist/ochemistry
photosynthesis	photos/ynthesis
probiotic	prob/iotic

GPT Neo 2.7B vs. BioMedLM



BioMedLM - Customization

Code/Model is open source, you can download your own copy of the model

Continue fine-tuning on your custom datasets (e.g. hospital EHRs) and specific tasks

Easy integration with Hugging Face

BioMedLM - Privacy/Security

Download To Your Private Cluster

Training/Inference Can Be Done In-House

No Need To Communicate With External Service

BioMedLM - Affordability

BioMedLM only has 2.7B parameters

Only cost $40k to pre-train for 300B tokens on MosaicML cluster

Can fine-tune on 1 GPU with 20GB of memory