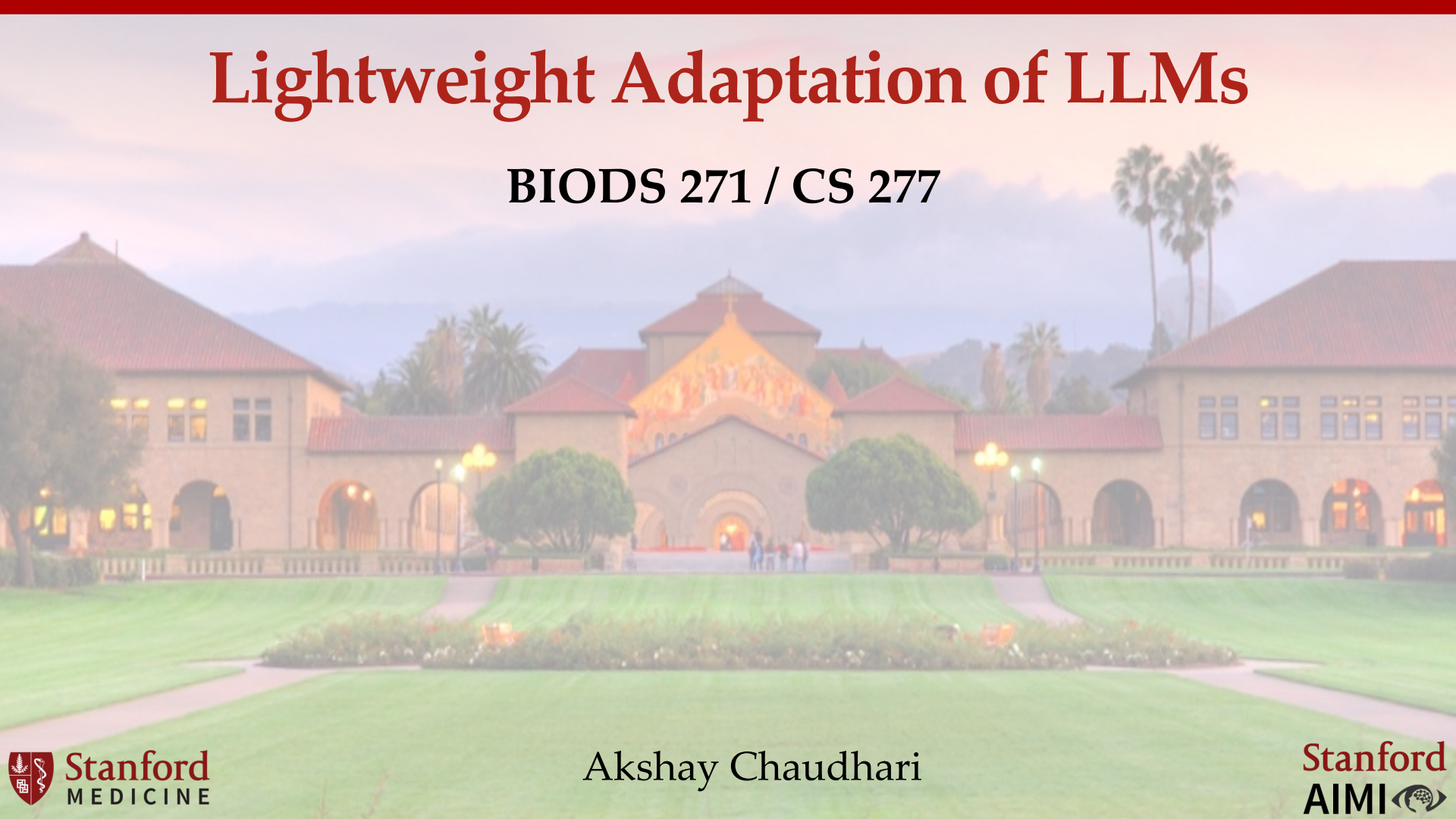
Lightweight Adaptation of LLMs

BIODS 271 / CS 277

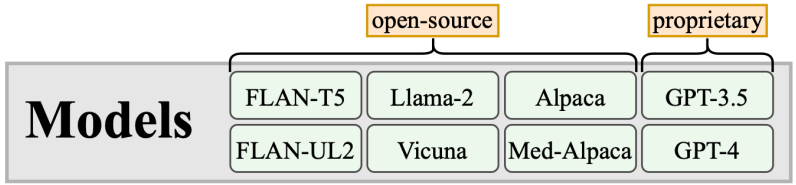


Stanford MEDICINE

Akshay Chaudhari

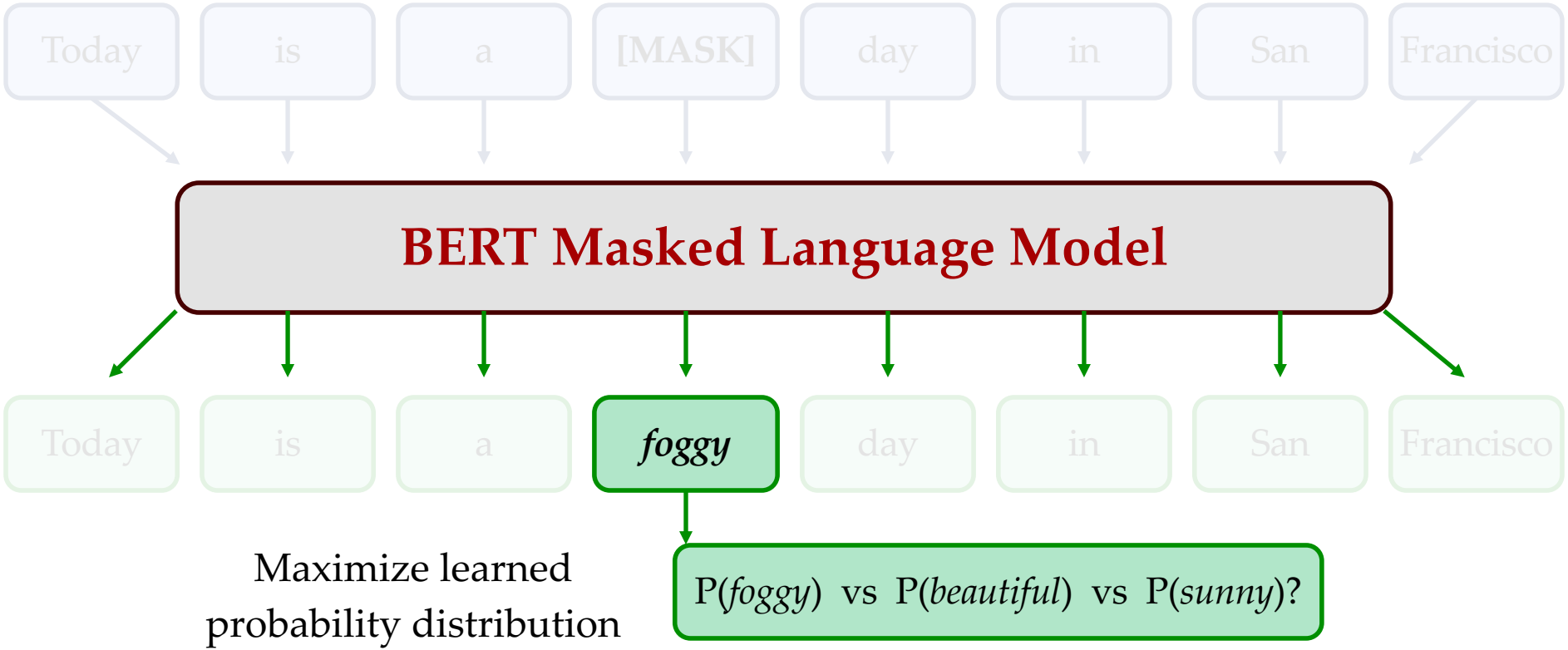
Stanford AIMI

LLM to Summarize Medical Text

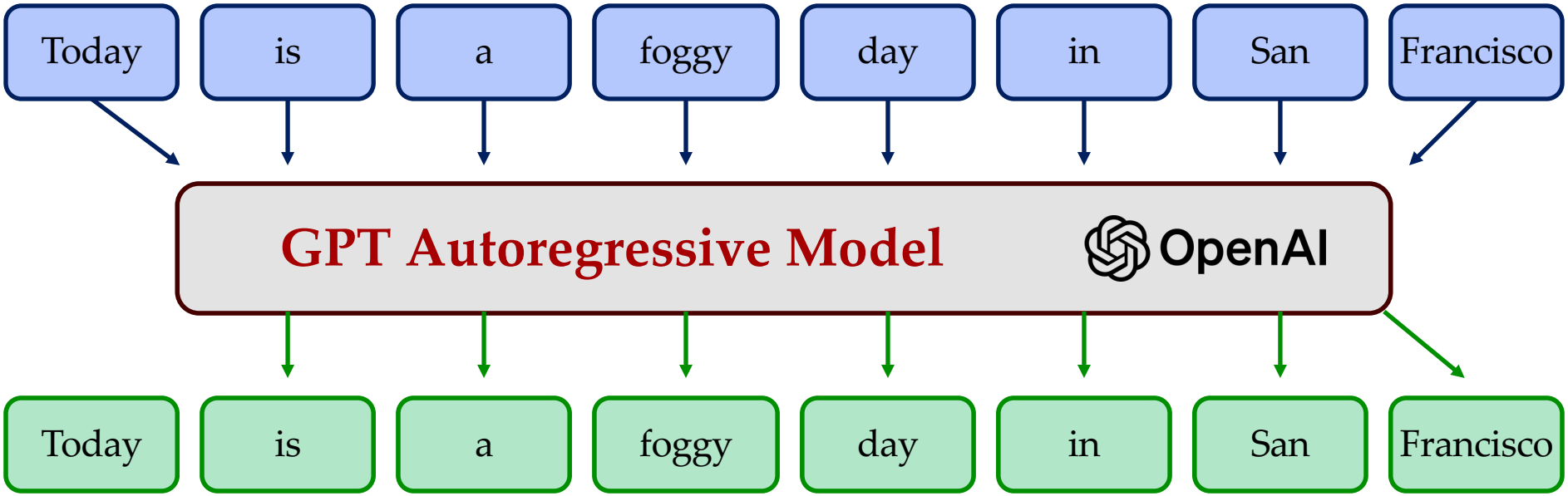


Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

BERT Pretraining



GPT Pretraining



Maximize conditional probability distribution

Example Datasets

Radiology Report Findings → Report Impressions

The patient is s/p left frontal craniotomy. A small amount of intracranial gas is seen posterior to the surgical intervention, which could represent postoperative changes. Extensive edema is seen in the left frontal lobe at the site of presumed surgery. Additionally multiple foci of hemorrhage are seen in the region of the left frontal lobe. Midline shift to the right is seen in the frontal region. The ventricles, cisterns, and sulci are unremarkable, without effacement. Comparison with prior studies from outside institution would be helpful in further evaluation of these findings.

1. Left frontal craniotomy.
2. Frontal midline shift to the right.
3. Extensive left frontal lobe edema.
4. Multiple foci of hemorrhage in the right frontal lobe.

Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Example Datasets

Patient Questions

Hello, I have been dealing with trimethylaminuria since I was a child. I have done some of my own research and it looks like not much can be done for this condition. I do not have it all over my body it's only in my armpits. In the past I've gone to doctors and dermatologist they gave me no answers until I looked online today and finally found out what I have. I don't know maybe I'm wrong. But this disease isn't even consider common because no one has done anything about it. I'm sure they're thousands of women with it... Can I be tested for it and help in some kind of way to finding a cure or something? What testing is done for this? And where? Thank you

Summary

How can I get tested and treated for trimethylaminuria?

Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Progress Notes

Summary

<ASSESSMENT>
*Ms. [**Known lastname 12031**] is a [**Age over 90 **] yo female with HTN, CAD s/p CABG, osteoporosis, COPD, here with painless lower GI bleeding and active extravasation from branch of middle colic artery on CTA now s/p angiographic coiling of middle colic artery branch.*

<SUBJECTIVE>
*UOP low, gave 500cc NS bolus doing very well clinically track serial hcts still having bloody bowel movements as expected if hct stable likely plan for scope 2am hct dropped to 29 from 35 [**Doctor First Name 91**] - give 2 units and recheck 1 hr after 2nd unit, 3-4 hours*
Lactose Intolerance (Oral) (Lactase) Unknown; Codeine
Nausea/Vomiting Bactrim Ds (Oral)
(Sulfamethoxazole/Trimethoprim) Unknown; Changes to and f
Review of systems is unchanged from admission except as noted below

Review of systems:
<OBJECTIVE>
*Last dose of Antibiotics: Ciprofloxacin - [**2196-3-31**] 12:29 AM*
*Infusions: Other ICU medications: Pantoprazole (Protonix) - [**2196-3-30**] 08:20 PM*
*Other medications: Flowsheet Data as of [**2196-3-31**] 06:40 AM*
*Vital signs Hemodynamic monitoring Fluid balance 24 hours Since [**98**] AM*
Tmax: 36.3 C (97.3 Tcurrent: 36.3 C (97.3
HR: 79 (79 - 92) bpm
BP: 115/45(62) {93/32(48) - 126/85(96)} mmHg
RR: 19 (18 - 29) insp/min
SpO2: 95%
Heart rhythm: SR (Sinus Rhythm)
Height: 62 Inch
Total In: 3,554 mL 2,328 mL
PO: TF: IVF: 179 mL 1,698 mL
Blood products: 375 mL 630 mL
Total out: 230 mL 191 mL
Urine: 230 mL 191 mL
NG: Stool: Drains:
Balance: 3,324 mL 2,137 mL
Respiratory support O2 Delivery Device: None
SpO2: 95%
ABG: ///27/

General: Alert, oriented, no acute distress
HEENT: Sclera anicteric, dry MM, oropharynx clear, dentures on upper teeth
Neck: supple, JVP not elevated, no LAD
Lungs: Clear to auscultation bilaterally, no wheezes, rales, rhonchi
CV: Regular rate and rhythm, normal S1 + S2, II/VI SEM LUSB, well-healed thoracotomy scar
Abdomen: soft, non-tender, very mildly distended, hyperactive bowel sounds, no rebound tenderness or guarding, no organomegaly appreciated
Ext: upper extremities WWP, 2+ pulses; LE cool with weak but palpable distal pulses
*107 K/uL 12.6 g/dL 139 mg/dL 0.5 mg/dL 27 mEq/L 4.4 mEq/L 13 mg/dL 107 mEq/L 139 mEq/L 29.7 % 10.7 K/uL image002.jpg] [**2196-3-30**] 03:10 PM [**2196-3-30**] 09:25 PM [**2196-3-31**] 01:54 AM*

WBC 10.7
Hct 30 35.9 29.7
Plt 107
Cr 0.5
Glucose 139

Other labs: PT / PTT / INR:13.5/28.2/1.2, ALT / AST:14/23, Alk Phos / T Bili:43/2.0, Lactic Acid:1.1 mmol/L, Albumin:3.0 g/dL, LDH:223 IU/L, Ca++:7.8 mg/dL, Mg++:1.7 mg/dL, PO4:3.9 mg/dL

GI bleed;
CAD;
UTI;
HTN;
Osteoporosis

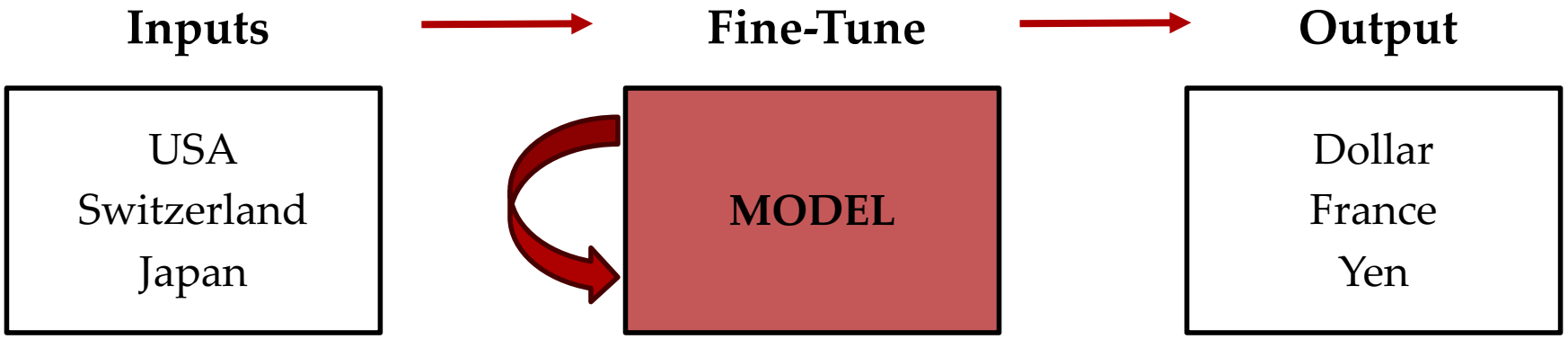
Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Prompt Anatomy

Expertise	`You are an expert medical professional.`
Instruction (task-specific)	`Summarize the [radiology report findings] into an [impression with minimal text].`
Examples $i = 1, \ldots, m$ #: delimiters *Note*: examples for ICL only, else $m = 0$	`Use the examples to guide word choice.` ⋮ `input` i`: {example input}` `summary` i`: {example summary}` `##` ⋮
Input	`input` $m+1$`: {input text}` `summary` $m+1$`:`

Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Supervised Finetuning of LLMs

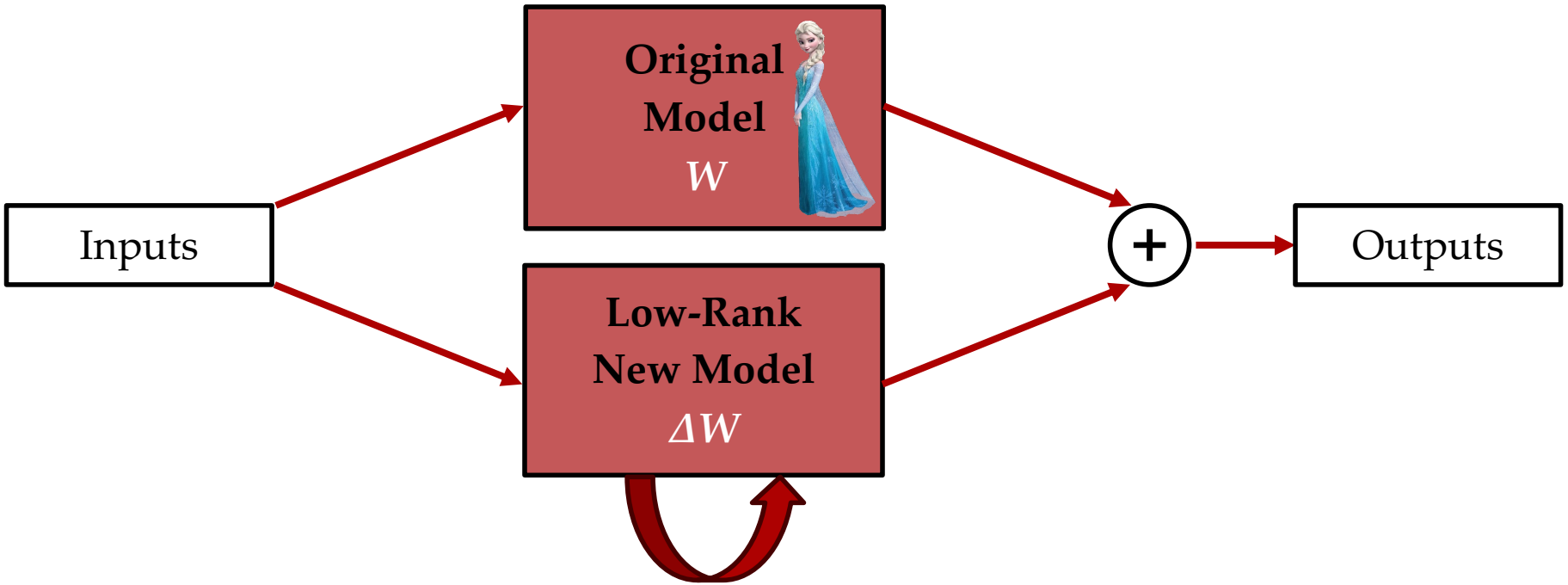


Model Weights $= W$

Fine-Tuning $= W + \Delta W$

It is challenging to fine-tune billion+ parameter LLMs!

Low Rank Adaptation



Hu et al. LoRA: Low-Rank Adaptation of Large Language Models. ICLR 2022

In Context Learning

- Simply pass training examples as inputs in prompts

Complete this sentence

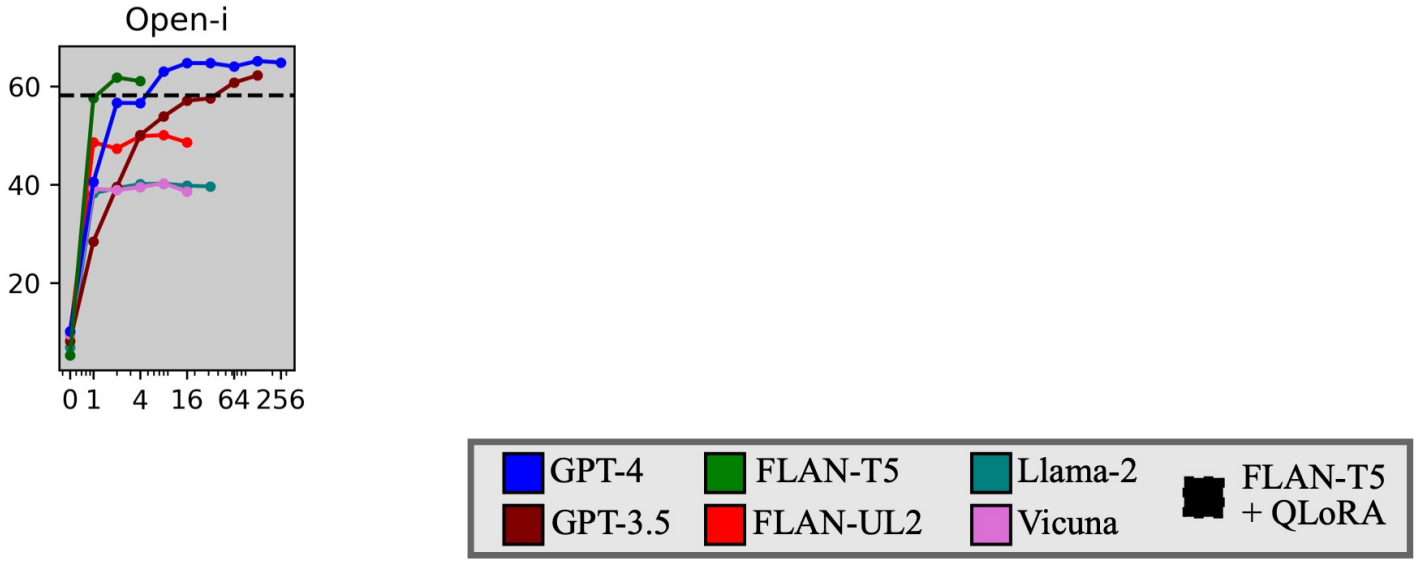
USA: Dollar,

Switzerland: Franc,

Japan: Yen,

Denmark: _____

In Context Learning



Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Win-Rates


BLEU
... against this model
Vicuna
Llama-2
FLAN-UL2
FLAN-T5
GPT-3.5
Win rate of this model...
GPT-4
GPT-3.5
FLAN-T5
FLAN-UL2
Llama-2
0.64 0.63 0.46 0.35 0.50
0.66 0.64 0.49 0.35
0.78 0.74 0.52
0.69 0.66
0.61
0.8
0.7
0.6
0.5
0.4
0.3
0.2

Reader Study Results

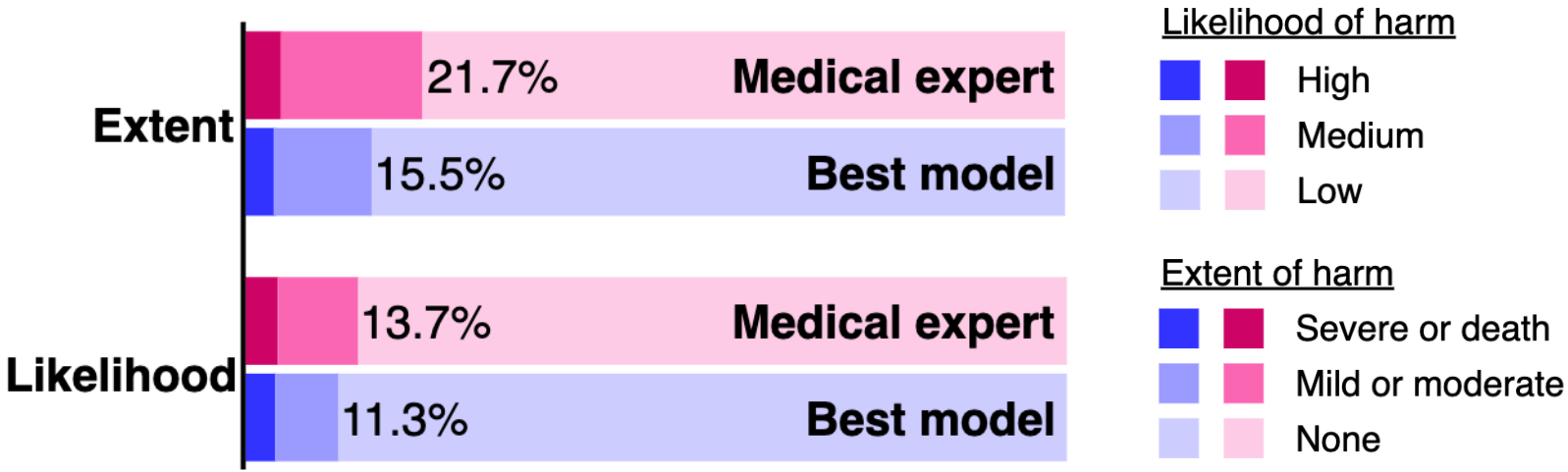
Which summary...

[Completeness] ... more completely captures important information?
[Correctness] ... includes less false information?
[Conciseness] ... contains less non-important information?

Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

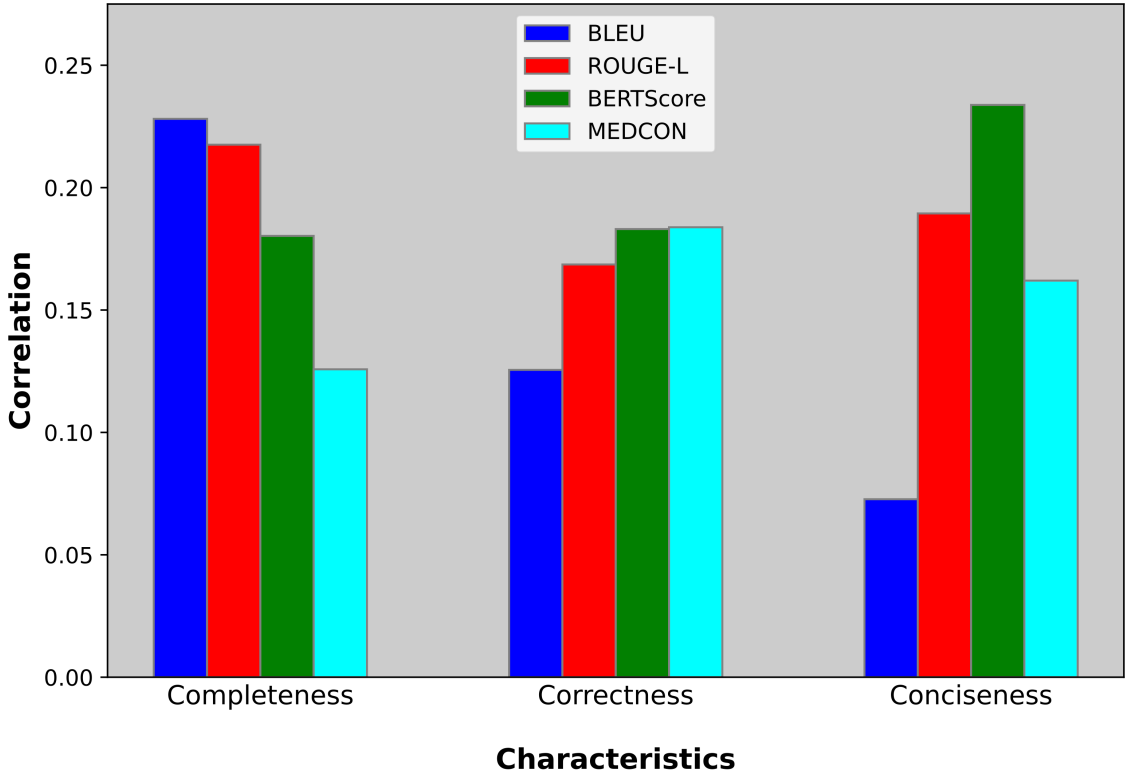
Reader Study Results

- Where either the human/LLM output was preferred, if the other inferior summary were to be used…



Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

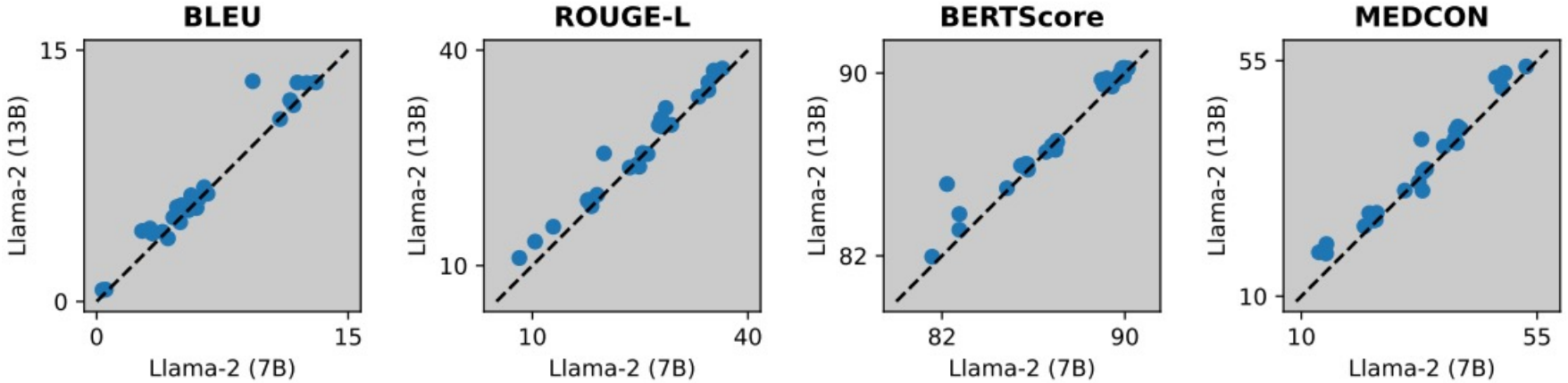
Metrics Correlation



Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Interesting Comparisons

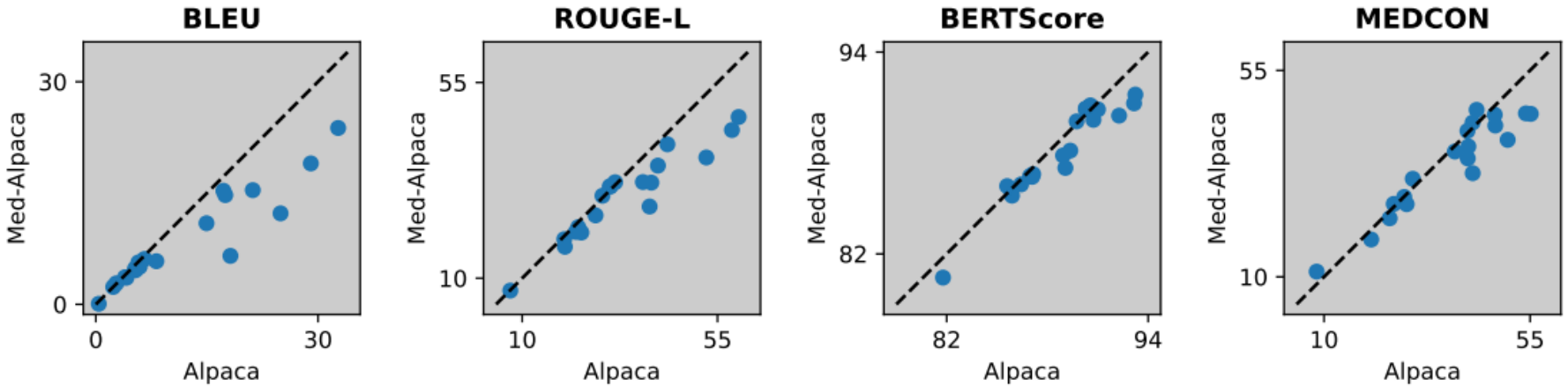
- Impact on 7B vs 13B models



Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Interesting Comparisons

- Impact of “medical” fine-tuning



Van Veen et al. Clinical Text Summarization: Adapting Large Language Models Can Outperform Human Experts. Nature Medicine (accepted), 2024.

Questions?

akshaysc@stanford.edu