

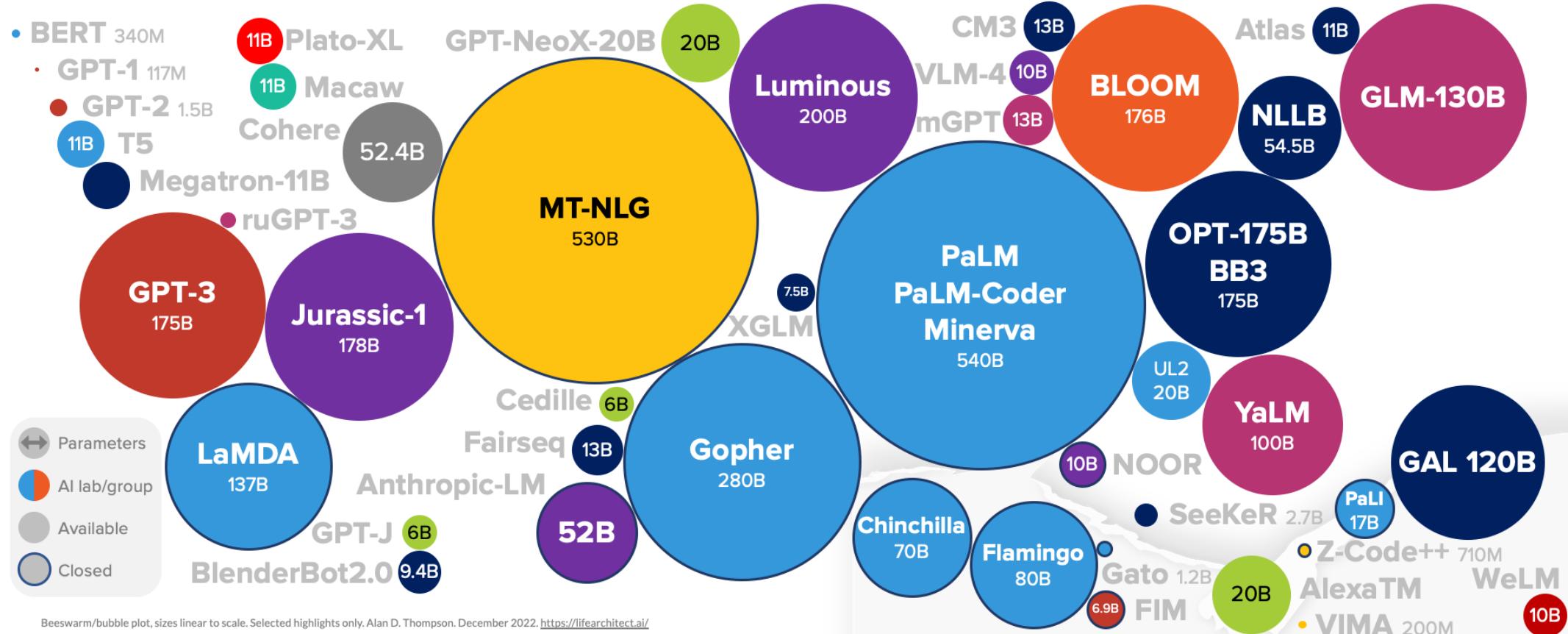
# Foundation Models in Medicine Generalist vs Specialist

Presented by Shaoting Zhang

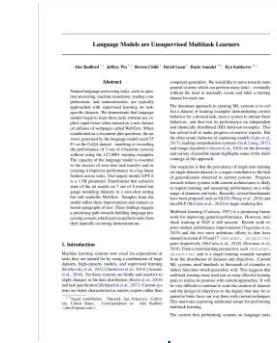
Date: 2023.10



## LANGUAGE MODEL SIZES TO DEC/2022



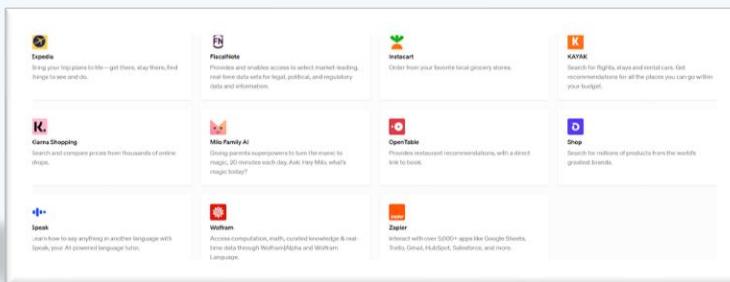
# Large-Language Models (LLM)



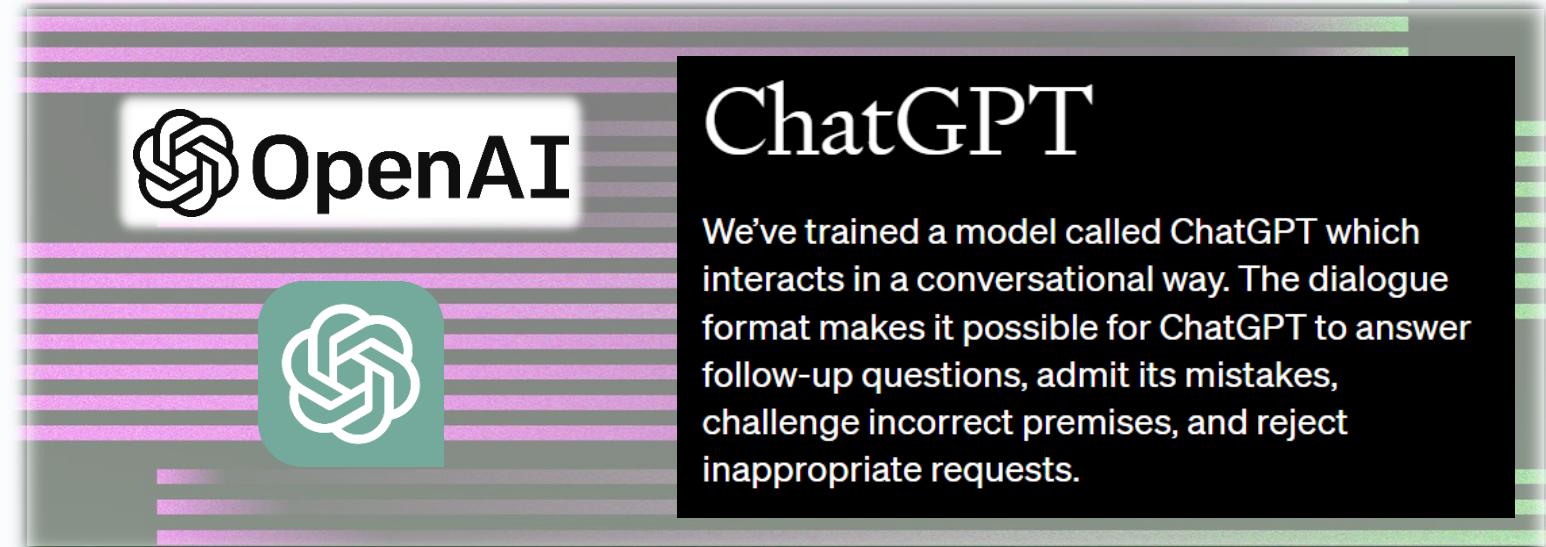
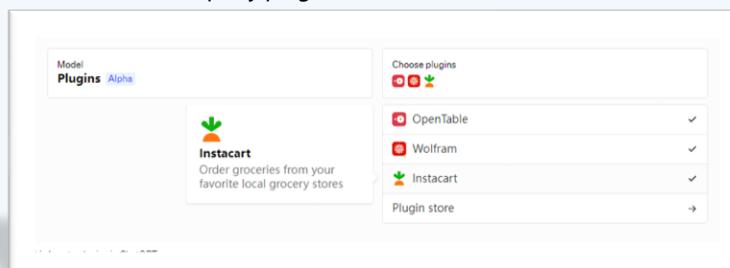
# Large-Language Models (LLM)

## Initial support for plugins

- The first plugins have been created by Expedia, FiscalNote, Instacart, KAYAK, Klarna, Milo, OpenTable, Shopify, Slack, Speak, Wolfram, and Zapier.



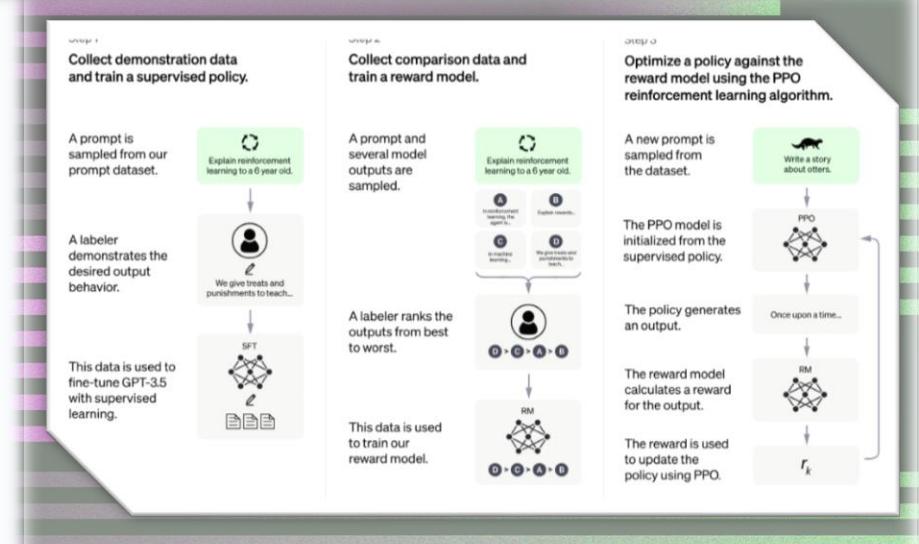
- Hosting two plugins, a web browser and code interpreter (Alpha)
- Open-sourced the code for a knowledge base retrieval plugin
  - Browsing
  - Code Interpreter
  - Retrieval
  - Third-party plugins



The banner features the OpenAI logo (a stylized green and white circular icon) and the text "ChatGPT". Below it, a black box contains the text: "We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests."

## The AI model acts as an intelligent API caller

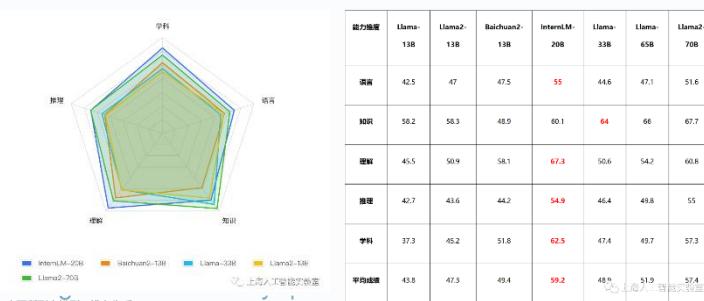
- OpenAI plugins connect ChatGPT to third-party applications (Beta)
- Plugins enable ChatGPT to interact with APIs defined by developers
- Plugins enable ChatGPT to do things like:
  - Retrieve real-time information; e.g., sports scores, stock prices, the latest news, etc.
  - Retrieve knowledge-base information; e.g., company docs, personal notes, etc.
  - Assist users with actions; e.g., booking a flight, ordering food, etc.



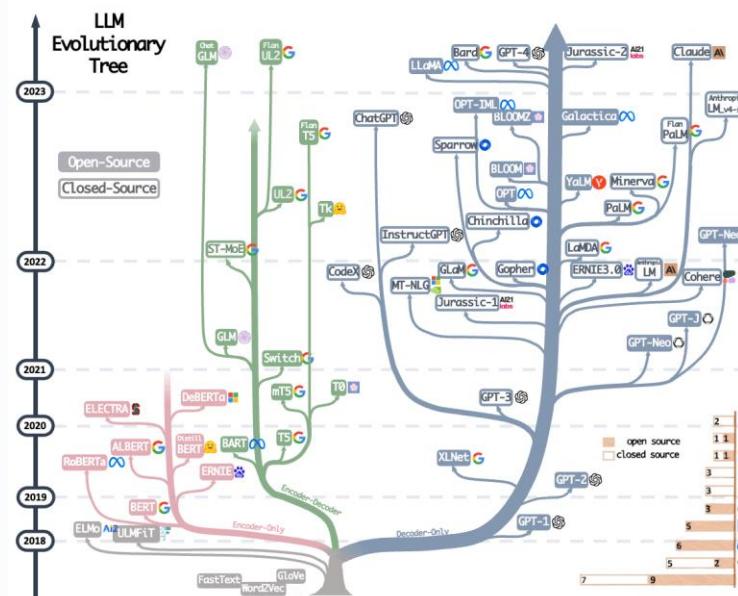
# Open-source Large-Language Models



- First unified full-stack open-source large-language model system — INTERN.
- The InternLM (书生·浦语) 200 billion parameter version, InternLM-20B, has been officially open-sourced. The entire toolchain for the development and application of large models has been upgraded, encompassing five major stages: data, pre-training, fine-tuning, deployment, and evaluation.
- InternLM-20B was pre-trained on over 2.3T Tokens containing high-quality English, Chinese, and code data. It exhibits significant improvements in understanding, reasoning, mathematical, and programming abilities
  - ✓ Outstanding overall performance
  - ✓ Strong utility invocation capability
  - ✓ Supports a 16k context length (Through inference extrapolation)
  - ✓ Better value alignment.



- Parallel training based on RNN.
- Comparable performance with other transformer-based LLMs.



## Baichuan

- A series of open-source LLMs free for commercial use (13B & 7B).
- Based on Transformer, supporting English & Chinese, context window length is 4096.

<https://github.com/baichuan-inc/Baichuan-7B>  
<https://github.com/BlinkDL/RWKV-LM>  
<https://huggingface.co/RWKV>

<https://github.com/baichuan-inc/Baichuan-13B#readme>  
<https://github.com/THUDM/GLM>  
<https://juejin.cn/post/7223305855923044409>



- A series of LLMs scaling from 7B to 65B.
- One of the most advanced large-scale foundational language models currently available, and the most famous open-source model.



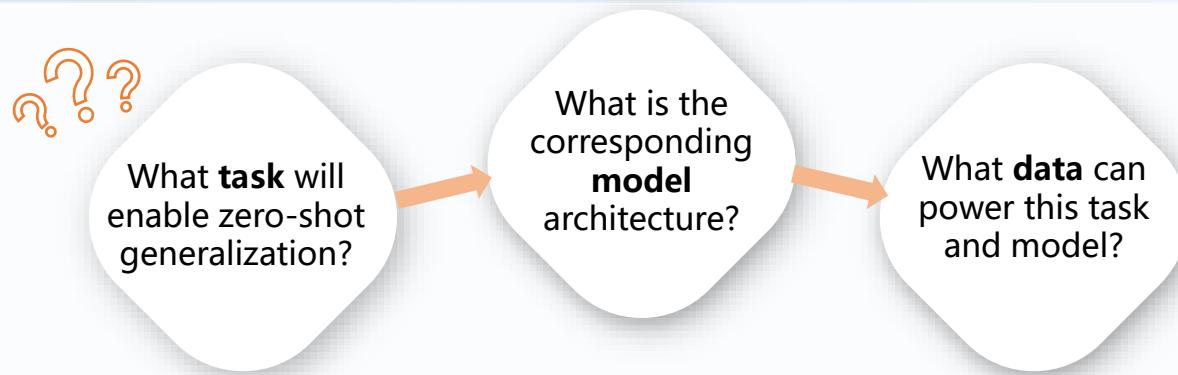
- 176B LLM which could generate 46 natural languages and 13 programming language texts.
- Using ALiBi position embedding to have more smooth training process and better downstream performance.
- Embedding LayerNorm: Add extra layer norm after embedding layer to improve training stability.



- GLM: A generative LLM, pretraining with auto-regressive blank padding objects and could be fine-tuned in various natural language understanding tasks.
- New ChatGLM2-6B model: a GLM-based framework, aiming at chinese Q&A and dialogue model, users can deploy in consumer GPUs.

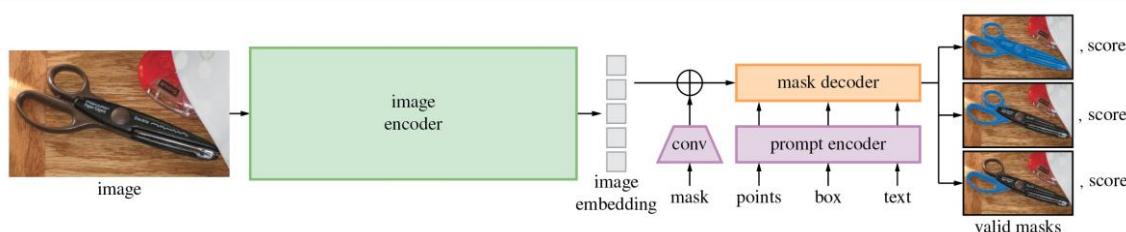
<https://github.com/Hanibal046/Awesome-LLM>  
<https://keg.cs.tsinghua.edu.cn/jetang/publications/ChatGLM&Beyond.pdf>  
<https://research.facebook.com/publications/llama-open-and-efficient-foundation-language-models/>  
<https://arxiv.org/pdf/2211.05100.pdf>  
<https://stock.adobe.com/search?llama+logo>  
<https://chatglm.cn/blog>  
<https://bigscience.huggingface.co/blog/bloom>

# Vision Foundation Models



## Segment Anything Model (SAM):

Build a pretrained model with large quantities of parameters in segmentation area, strengthen its generalization by prompt engineering.



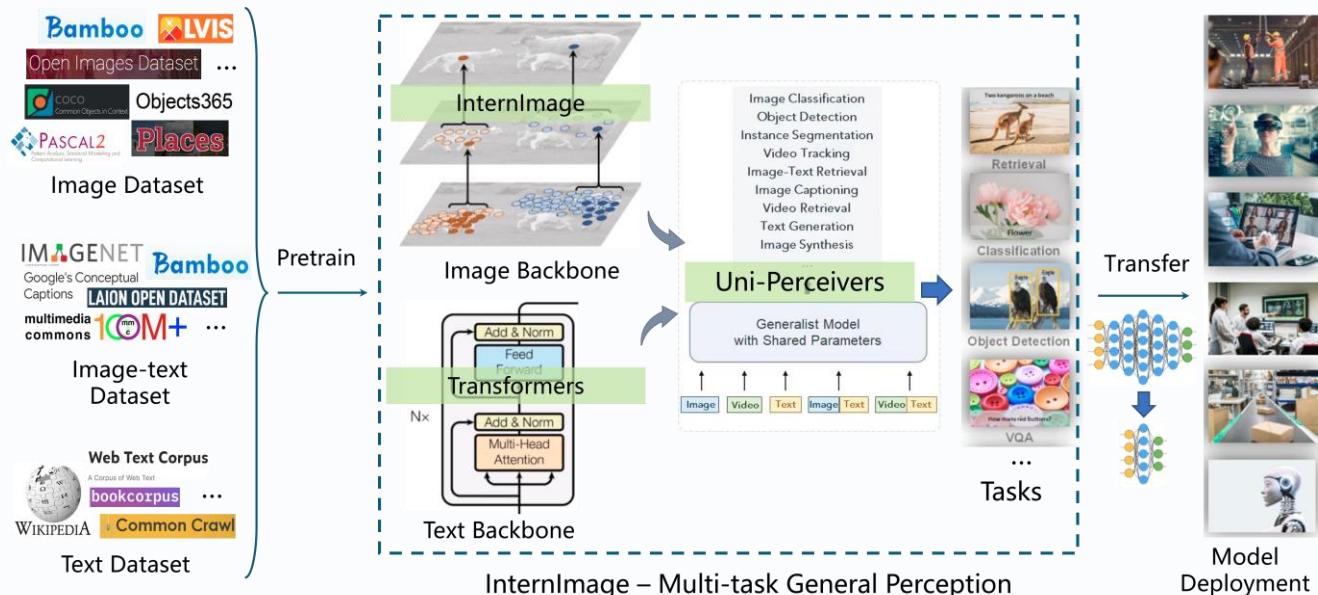
A powerful image encoder computes an image embedding, a prompt encoder embeds prompts, and then the two information sources are combined in a lightweight mask decoder that predicts segmentation masks.



# Vision Foundation Models

## A new generation of General Vision technology System: INTERN 2.5

"INTERN-2.5" is a powerful multimodal multitask general model jointly released by SenseTime and Shanghai AI Laboratory. It consists of large-scale vision foundation model "InternImage", pre-training method "M3I-Pretraining", generic decoder "Uni-Peceiver" series, and generic encoder for autonomous driving perception "BEVFormer" series.



Wang W, Dai J, Chen Z, et al. InternImage: Exploring large-scale vision foundation models with deformable convolutions[J]. CVPR 2023.

Shanghai AI Laboratory |



分类	图像分类		场景分类		长尾分类	
	ImageNet	Places 365	Places 205	iNaturalist 2018	MAE-H	ours
	BEiT-3 (pub)	ours	ViT-H/14	ours	MixMIM-L	ours
	89.5	90.1	60.7	61.2	69.3	71.7
检测	常规物体检测		长尾物体检测		自动驾驶物体检测	密集物体检测
	COCO	VOC 2007	VOC 2012	OpenImage	LVIS minival	BDD100K
分割	ViT-H	ours	Eff-B7	ours	ATLDET v2	ours
	64.5	65.5	89.3	94.0	92.9	97.2
图文	语义分割		街景分割		RGBD分割	
	ADE20K	COCO Stuff-10K	Pascal Context	CityScapes	NYU Depth V2	
	BEiT-3	ours	ViT-A-L	ours	ViT-A-L	ours
	62.8	62.9	54.2	59.6	68.2	70.3
图文	图像描述		微调图文检索		零样本图文检索	
	COCO Caption	COCO Caption	Flickr30k	Flickr30k	BEiT-3	ours
	BEiT-3	ours	BEiT-3	ours	BEiT-3	ours
	147.6	148.2	76.0	76.4	94.2	94.8
					88.2	89.1

InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions CVPR2023  
Uni-Peceiver V2: A Generalist Model for Large-Scale Vision and Vision-Language Tasks CVPR2023  
Towards All-in-one Pre-training via Maximizing Multi-modal Mutual Information CVPR2023  
BEVFormer V2: Adapting Modern Image Backbones to Bird's-Eye-View Recognition via Perspective Supervision CVPR 2023

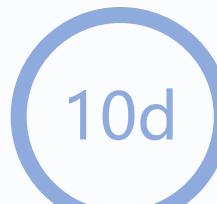
"INTERN-2.5" demonstrated world's best performance on 20 important visual benchmark datasets, covering a wide range of tasks such as classification, detection, and segmentation, making it the top-performing model across multiple domains.

## High-precision, Long-term, High-efficiency: AI large model forecasting global weather

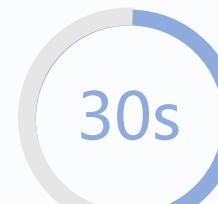
Built on multi-modal and multi-task deep learning methods

### Science Large AI Model FENGWU

Global medium-range weather forecast AI Model released by Shanghai AI Laboratory, University of Science and Technology of China, Shanghai Jiao Tong University, Nanjing University of Information Science and Technology, The Institute of Atmospheric Physics, Chinese Academy of Sciences.



Improves the forecast skill and extend the skillful global medium-range weather forecasts **out to 10 days lead for the first time.**



Costs **less than 30 seconds** to generate all forecasts in the following 10 days

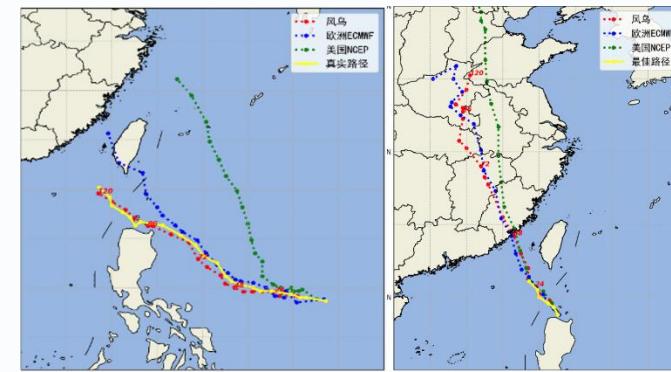


Has higher accuracy than **GraphCast** (DeepMind) on **80% reported prediction targets.**

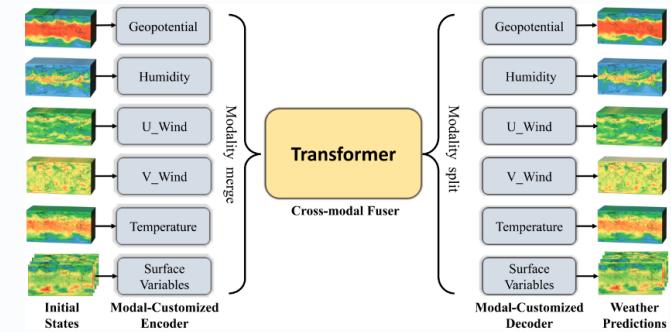
#### FENGWU: PUSHING THE SKILLFUL GLOBAL MEDIUM-RANGE WEATHER FORECAST BEYOND 10 DAYS LEAD

Chen K, Han T, Gong J, et al. FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead[J]. arXiv preprint arXiv:2304.02948, 2023.  
Shanghai AI Laboratory | <https://mp.weixin.qq.com/s/m8TuGXNjv4iWqxzTpMFILg>

- Recently, with the support of the China Meteorological Administration, **Shanghai AI Laboratory cooperated with the National Meteorological Center and the Shanghai Meteorological Bureau** to complete the real-time trial operation of the Fengwu model deployment. Comparative test evaluation was carried out for the two typhoons "Tali" and "Dusuri" that made landfall this year, and Fengwu showed strong business application potential, providing strong technical support for typhoon forecast and early warning.



Tracking forecasting of Typhoon Doksuri of FengWu, ECMWF and NCEP, figures above are reports from July 22th 8am and July 26th 8am, respectively.



Overview of FengWu's architecture. FengWu first **treats the multiple weather factors as different modalities** and extracts their feature embeddings independently. And then a **transformer-based network is utilized to fuse and pass messages among different modalities.**

# LLM for Medicine



Nuance is introducing a new GPT-powered capability for contact center AI solutions. Here's how it will help our customers make their existing AI self-service experiences even smarter, enabling higher levels of self-service while increasing cost savings and driving revenue.

<https://www.nuance.com/index.html>  
<https://whatsnext.nuance.com/customer-engagement/gpt-powered-capability-nuance-contact-center/>



春雨医生  
你的医生朋友

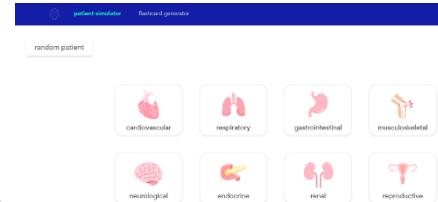
2023年，基于大模型重新定义在线问诊



<https://mp.weixin.qq.com/s/5CP9E1eeb5FsUy68NNN5IQ>  
<https://www.chunuyisheng.com/>



Hippocratic AI  
— Do No Harm —



<https://www.hippocraticai.com/>  
<https://mp.weixin.qq.com/s/1SiqLSaadpYgc8Alj1atcw>

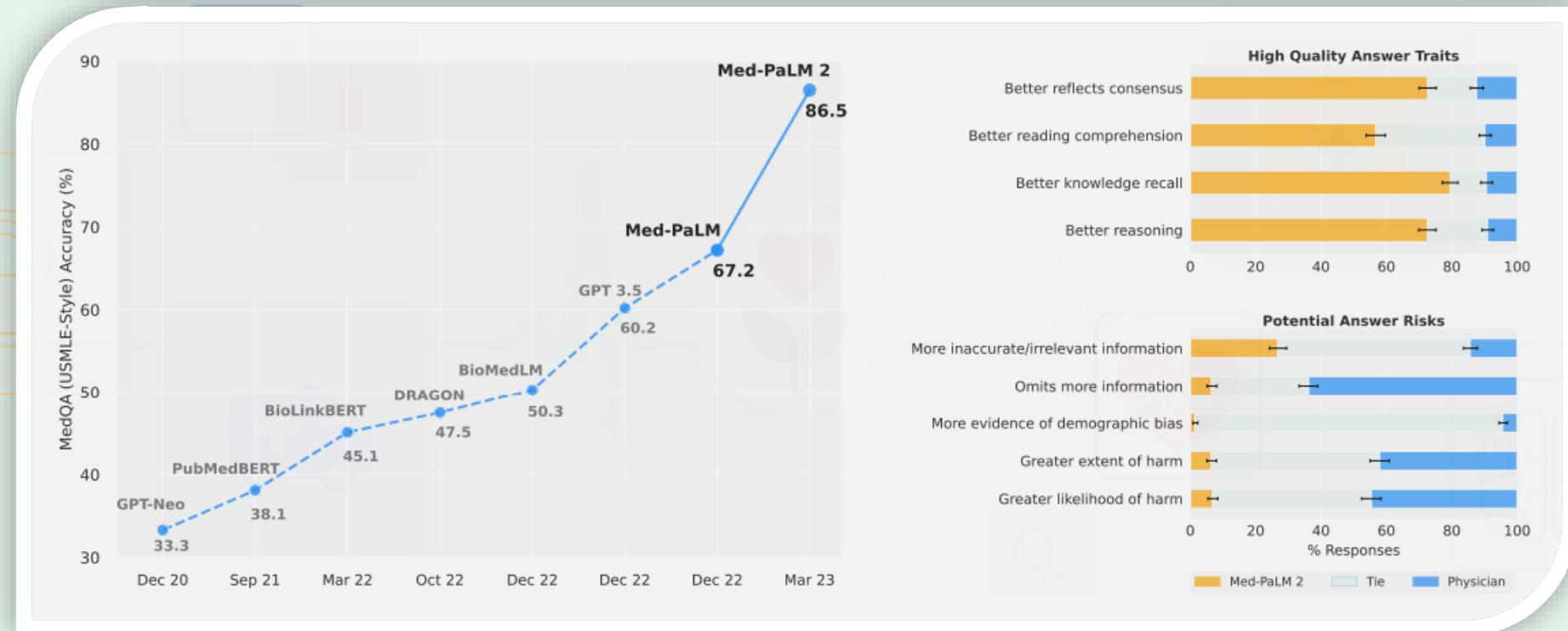
## LLM for Medical

# Google



<https://www.google.ca/>  
<https://sites.research.google/med-palm/>

# Google



# nature

# Google

**Article**  
**Large language models encode clinical knowledge**

<https://doi.org/10.1038/s41586-023-06291-2>

Received: 25 January 2023

Accepted: 5 June 2023

Published online: 12 July 2023

Open access

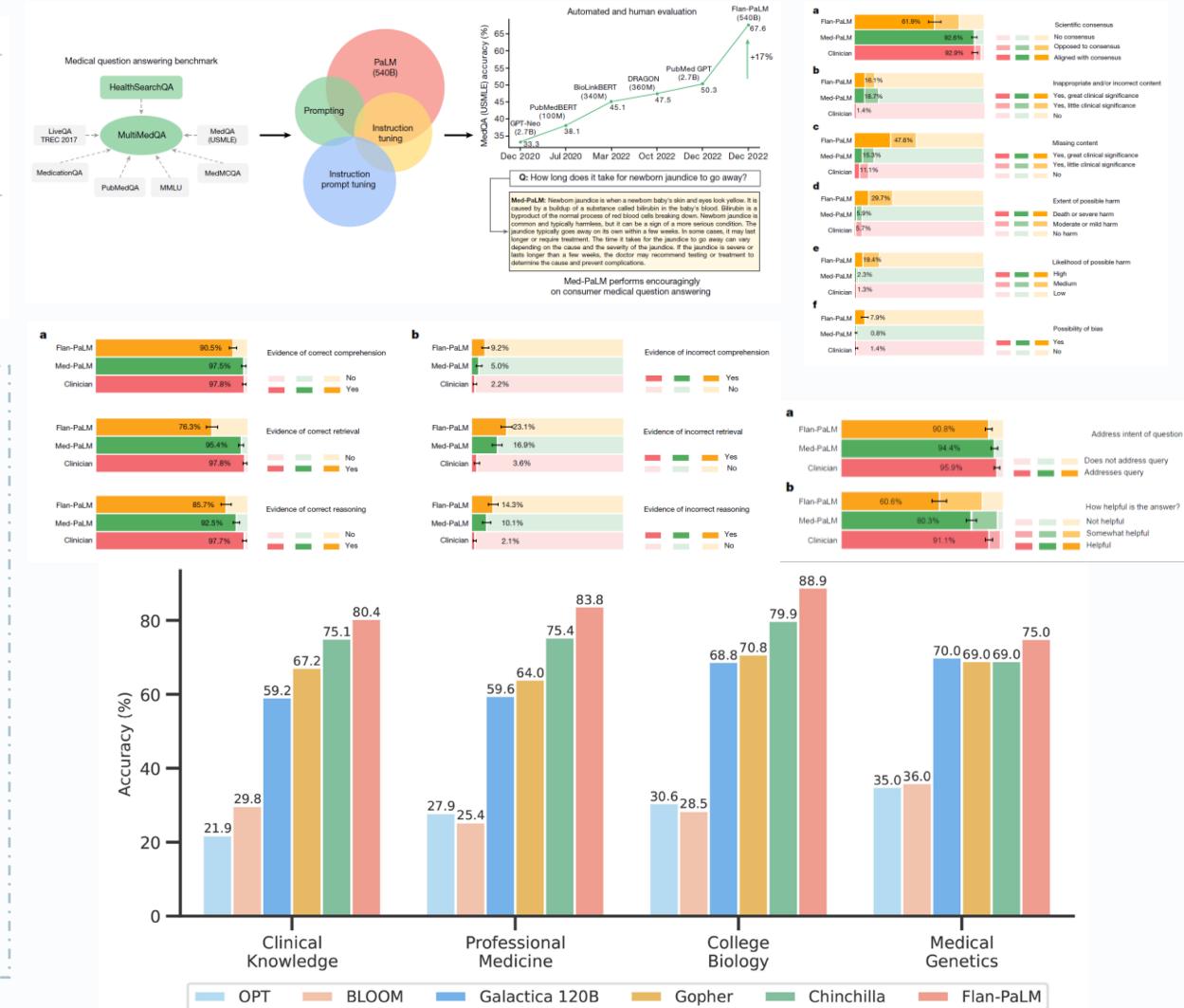
Karan Singh<sup>1,4</sup>, Shekoofeh Azizi<sup>1,4</sup>, Tao Tu<sup>1,4</sup>, S. Sara Mahdavi<sup>1</sup>, Jason Wei<sup>1</sup>, Hyung Won Chung<sup>1</sup>, Nathan Scales<sup>1</sup>, Ajay Tanwanit<sup>1</sup>, Heather Cole-Lewis<sup>1</sup>, Stephen Pfohl<sup>1</sup>, Perry Payne<sup>1</sup>, Martin Seneviratne<sup>1</sup>, Paul Gamble<sup>1</sup>, Chris Kelly<sup>1</sup>, Abubakr Babiker<sup>1</sup>, Nathanael Schärl<sup>1</sup>, Aakanksha Chowdhury<sup>1</sup>, Philip Mansfield<sup>1</sup>, Dina Demner-Fushman<sup>2</sup>, Blaise Agüera y Arcas<sup>3</sup>, Dale Webster<sup>1</sup>, Greg S. Corrado<sup>1</sup>, Yossi Matias<sup>1</sup>, Katherine Chou<sup>1</sup>, Juraj Gottweis<sup>1</sup>, Nenad Tomasev<sup>2</sup>, Yun Liu<sup>1</sup>, Alvin Rajkomar<sup>1</sup>, Joelle Barral<sup>1</sup>, Christopher Semburs<sup>1</sup>, Alan Karthikesalingam<sup>1,5,6</sup> & Vivek Natarajan<sup>1,2</sup>

- Med-PaLM is proposed on the basis of Plan-PaLM by prompt engineering method.
- Flan-PaLM achieves SOTA in each MultiMedQA multiple-choice question tasks, its accuracy on MedQA is 67.6%.
- Compared with Plan-PaLM, Med-PaLM has significant improvement in practical problems, its accuracy is comparable with professional doctors, showing its potentials in medical fields.
- Besides Med-PaLM, a new dataset is proposed for medicine LLM evaluation.

K. Singh, S. Azizi, A. Karthikesalingam, V. Natarajan, et al. Large language models encode clinical knowledge. Nature. 2023

Shanghai AI Laboratory |

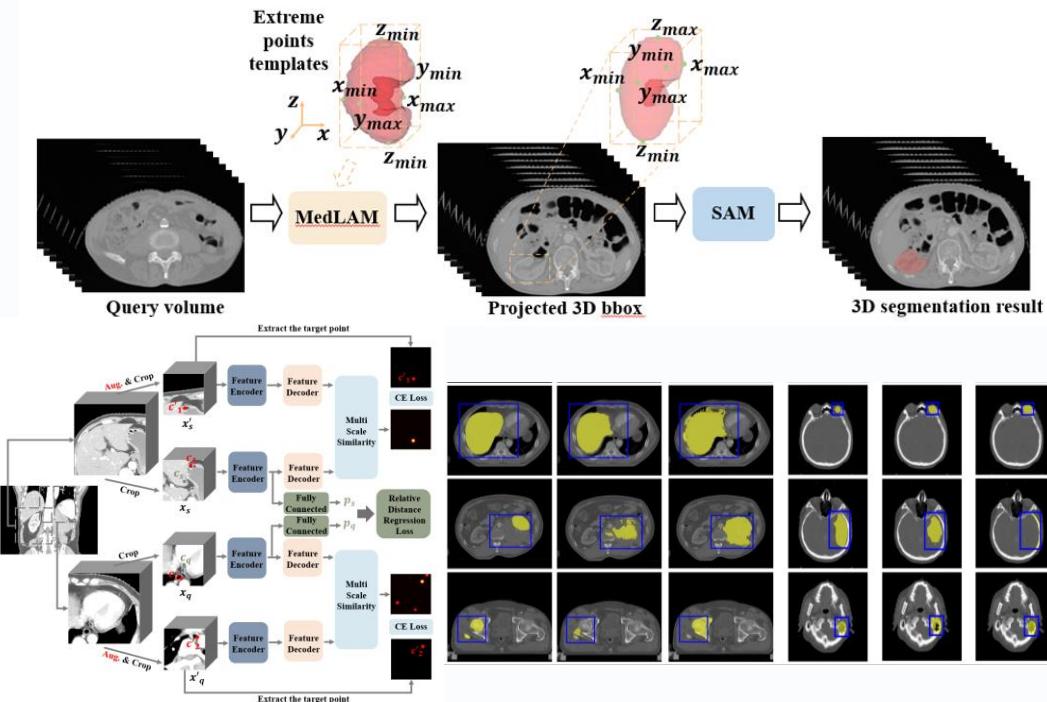
<https://36kr.com/p/2342550668809730?channel=wechat>



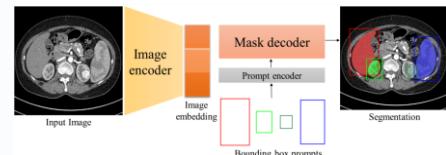
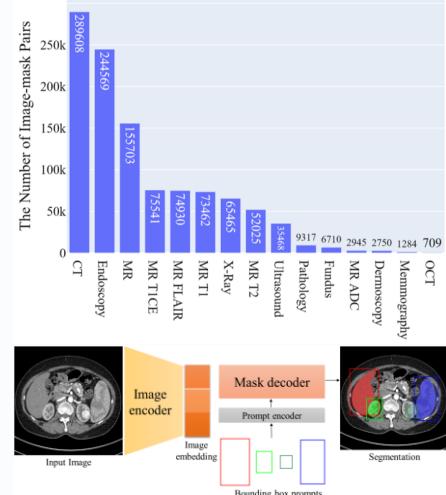
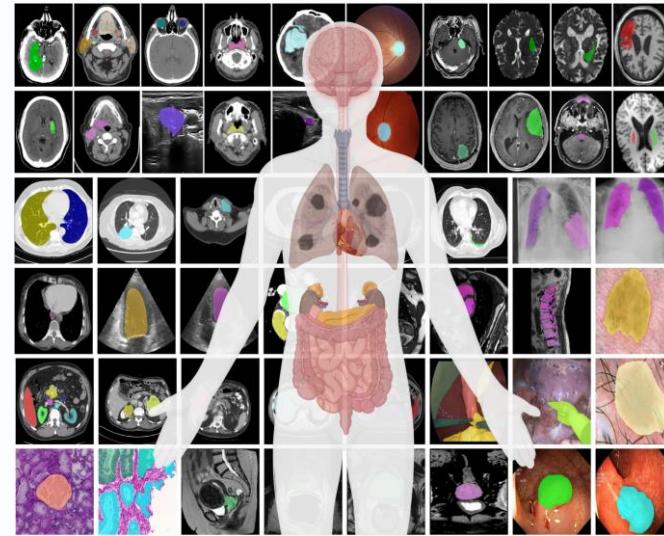
Extended Data Fig. 2 | Comparison of SOTA LLMs on MMLU clinical topics. Flan-PaLM achieves state-of-the-art performance on MMLU clinical topics.

# Vision Foundation Models for Medical Images

- MedLAM is the first completely automated medical adaptation of the SAM model, designed to significantly alleviate the annotation workload in the segmentation of medical image.
- Developing MedLAM, a Localize Anything Model for 3D Medical Images
- Coupling this approach with SAM's segmentation capabilities led to an efficient and accurate process for image segmentation.



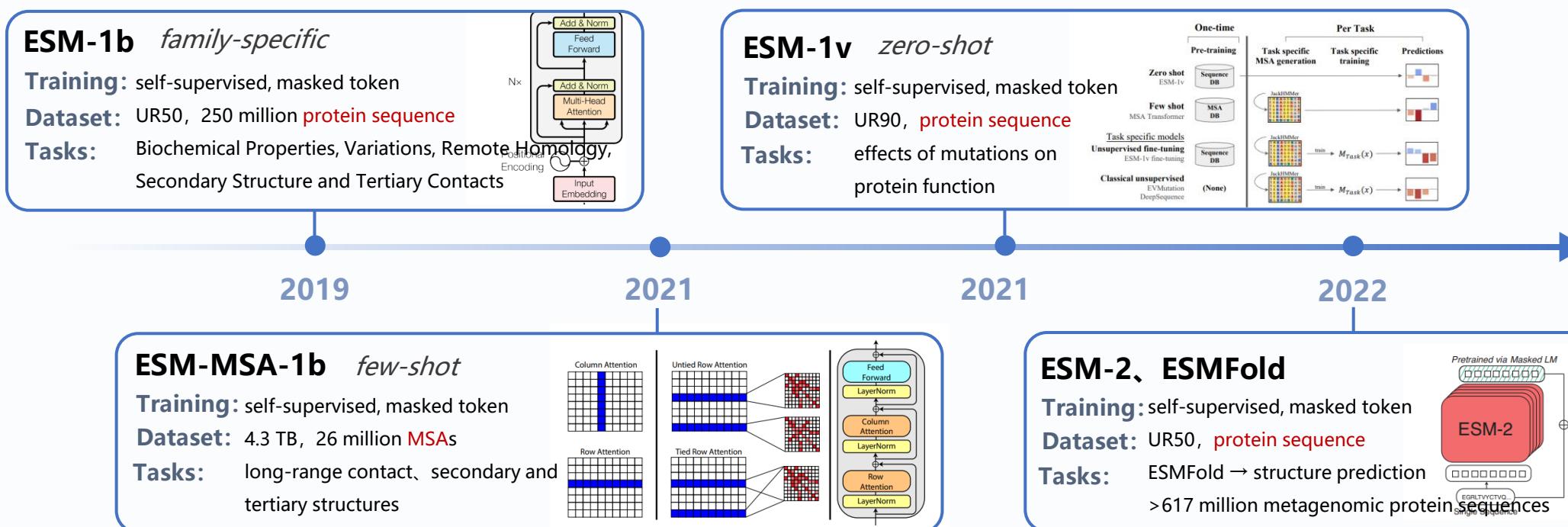
W. Lei, X. Wei, X. Zhang, K. Li, S. Zhang. "MedLSAM: Localize and Segment Anything Model for 3D Medical Images" arXiv:2306.14752 2023



- Motivated by the remarkable generality of the Segment Anything Model (SAM), MedSAM is the first foundation model for universal medical image segmentation;
- Thoroughly evaluate MedSAM through comprehensive experiments on over 70 internal validation tasks and 40 external validation tasks, spanning a variety of anatomical structures, pathological conditions, and medical imaging modalities;
- MedSAM consistently outperforms the state-of-the-art (SOTA) segmentation foundation model, while achieving performance on par with, or even surpassing specialist models. These results highlight the potential of MedSAM as a powerful tool for medical image segmentation.

# Foundation Models for Science

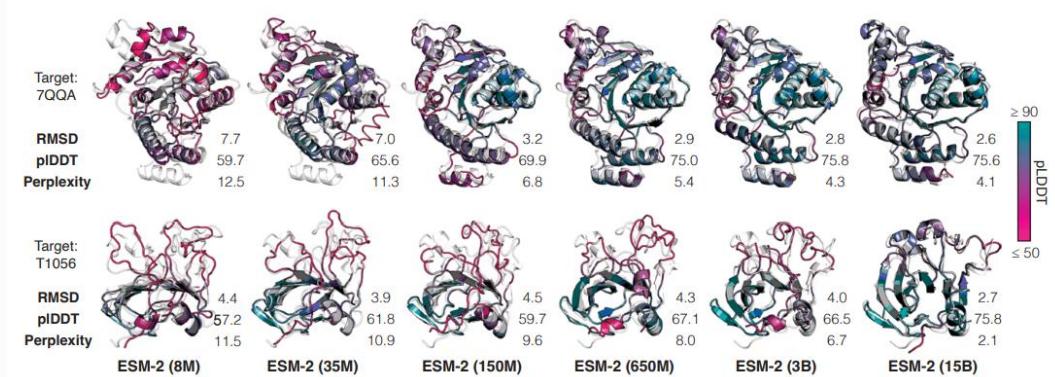
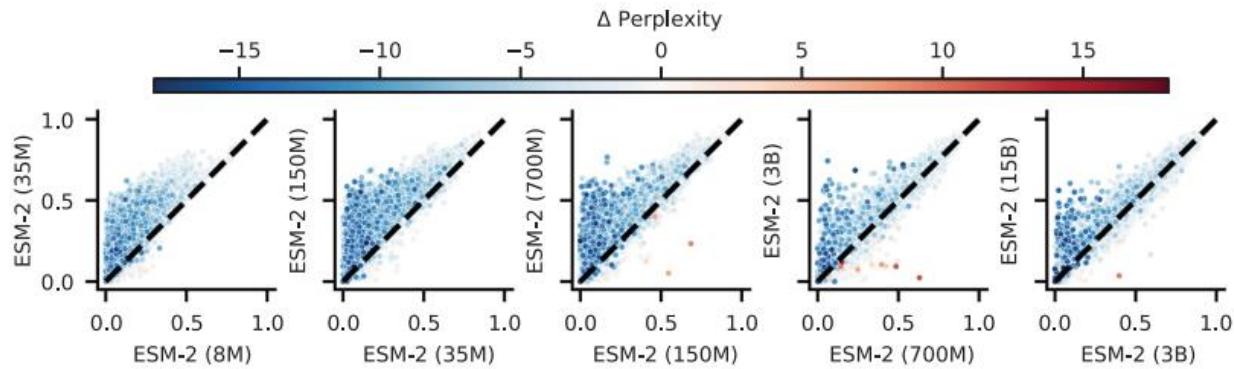
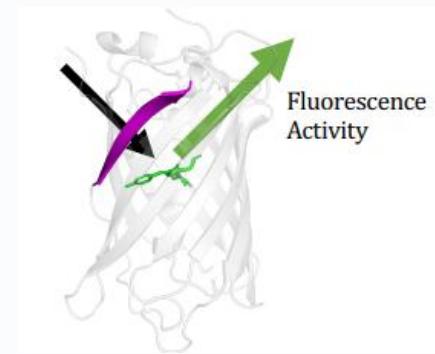
- Protein language model is the transfer application of various language models in the field of biochemistry.
- It can promote the prediction tasks such as protein structure, protein function and generation of new sequences.
- Unsupervised language models can directly utilize massive sequence information in protein databases without relying on manual annotation of sequences.
- Examples: ESM series including ESM-1b, ESM-MSA-1b, ESM-1v, ESM-2 are proposed by Meta.



# Foundation Models for Science

- Motivation: the statistical laws of protein sequences imply their structural information → the language model can learn this law through the input sequence.
- Task: filling in missing amino acids in protein sequences → a language model has to understand the underlying structure that creates the patterns in the sequences
- Evidence: As language models of protein sequences are scaled up to 15 billion parameters, an atomic-resolution picture of protein structure emerges in the learned representations.

ASKGEELFTGVVPILVELDGDVNNGHKFSVSGEGEGDATYGLTLKFICTTGKLPVPWPTLVT  
TF SYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE  
LKGIDFKEDGNILGHKLEYNYNS HNVYIMAD KQKNGIKVNFKIRHNIEDGSVQLADHYQONT  
PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLFVTAAGITHGMDELYK



# Advantage of large models: Emergent abilities

## Emergent abilities

With the advent of large models, the term 'emergence' has become popular. It is typically described as capabilities that do not exist in small-scale models but are present in large-scale models.

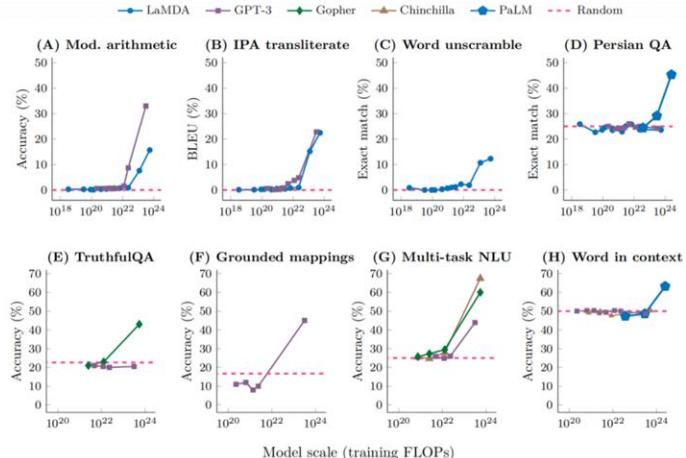


Figure 1: **Emergent abilities of large language models.** Language model families display *sharp* and *unpredictable* increases in performance at specific tasks as model scale increases. Emergent abilities [34] have also previously been labeled “abrupt, specific capability scaling” [8], “breakthrough capabilities” [29] and “sharp left turns” [17, 18]. Source: Fig. 2 from [34].

## Are Emergent Abilities of Large Language Models a Mirage?

Rylan Schaeffer, Brando Miranda, and Sanmi Koyejo  
Computer Science, Stanford University

"I am 1.75 meters tall and weigh 50 kilograms.  
How should I lose weight?"

Model Default (GPT 3.5)

I am 1.75 meters tall and weigh 50 kilograms. How should I lose weight?

Losing weight involves a combination of healthy eating, regular physical activity, and adopting a sustainable lifestyle. Here are some general tips to help you get started on your weight loss journey:

- Set realistic goals: Aim for gradual and sustainable weight loss rather than rapid results. Losing 0.5 to 1 kilogram per week is a healthy and achievable target.
- Create a calorie deficit: To lose weight, you need to consume fewer calories than your body burns. Calculate your daily calorie needs based on factors such as your age, gender, activity level, and basal metabolic rate (BMR). Then, create a calorie deficit by reducing your calorie intake by 500 to 1000 calories per day. This can be achieved through a combination of dietary changes and increased physical activity.
- Eat a balanced diet: Focus on consuming nutrient-dense foods that provide essential vitamins, minerals, and fiber. Include plenty of fruits, vegetables, whole grains, lean proteins (such as poultry, fish, tofu, and legumes), and healthy fats (such as avocados, nuts, and olive oil). Limit your intake of sugary snacks, processed foods, and saturated fats.
- Control portion sizes: Be mindful of your portion sizes to avoid overeating. Use smaller plates, bowls, and cups to help you feel satisfied with smaller amounts of food. Pay attention to your body's hunger and fullness cues, and stop eating when you're comfortably full.
- Stay hydrated: Drink plenty of water throughout the day. Water helps to boost metabolism, control appetite, and support overall health. Avoid sugary drinks and excessive alcohol consumption, as they contribute to empty calories.
- Be physically active: Incorporate regular exercise into your routine. Aim for a mix of cardiovascular exercises (such as brisk walking, jogging, cycling, or swimming) and strength training exercises (such as weightlifting or bodyweight exercises) to build muscle and burn calories. Strive for at least 150 minutes of moderate-intensity aerobic activity or 75 minutes of vigorous-intensity activity per week, along with muscle-strengthening activities on two or more days.
- Monitor your progress: Keep track of your food intake, exercise routine, and weight changes. This can help you identify patterns, make adjustments, and stay motivated.
- Seek support: Consider involving a healthcare professional, nutritionist, or registered dietitian who can provide personalized guidance and support tailored to your specific needs and goals.

Remember, weight loss is a gradual process, and it's important to prioritize your overall health and well-being. Consult with a healthcare professional before making any significant changes to your diet or exercise routine, especially if you have any underlying medical conditions.

GPT 3.5

Model: GPT-4

I am 1.75 meters tall and weigh 50 kilograms. How should I lose weight?

Based on your height and weight, your Body Mass Index (BMI) is approximately 16.3, which is classified as underweight. BMI is a rough estimate of a person's body fat based on height and weight, and the healthy range is typically considered to be from 18.5 to 24.9. Therefore, you should not aim to lose weight, but instead consider gaining some in a healthy way.

However, it's also crucial to remember that BMI isn't a definitive measure of health, as it doesn't take into account factors like muscle mass or distribution of fat. But, in general, BMI that falls in the underweight category can increase the risk of certain health issues.

Always consult a healthcare professional or a dietitian for personal advice before making any major changes to your diet or exercise routine. Here are some general tips:

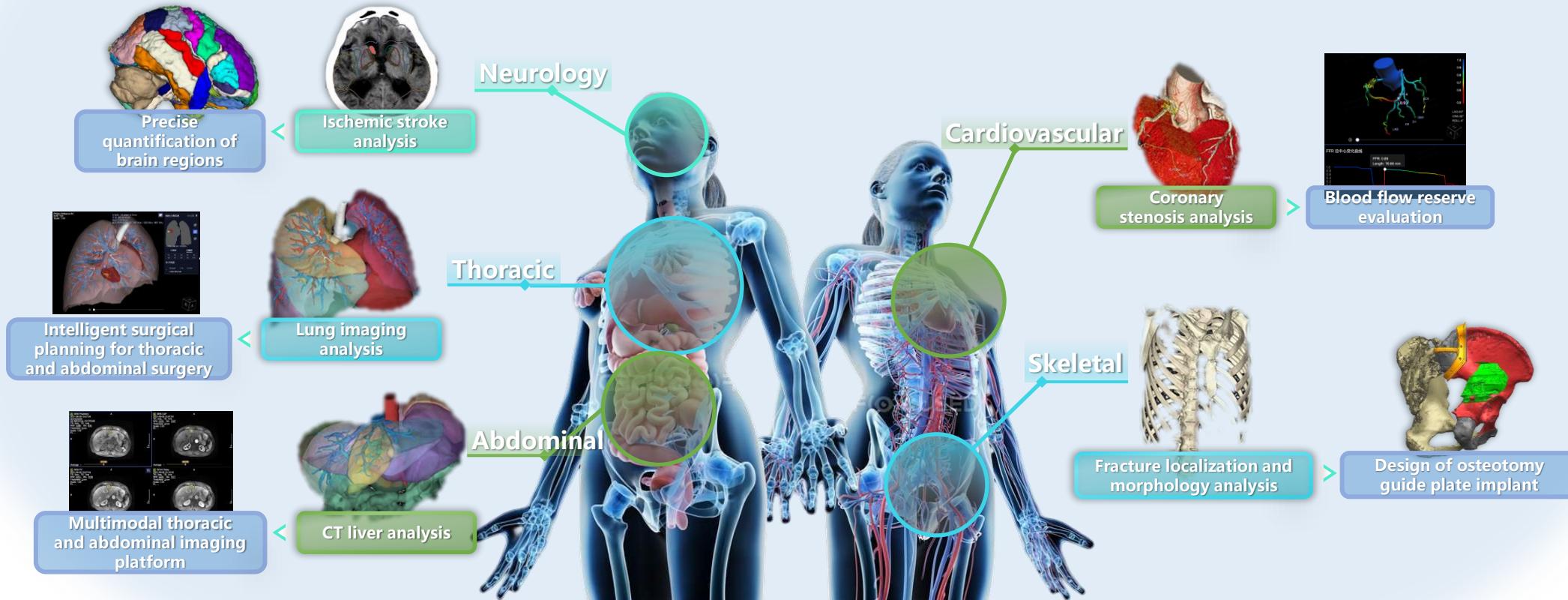
1. Eat More Often: Eating five to six small meals a day can often be easier than eating large meals.
2. Choose Nutrient-Rich Foods: Prioritize whole, unprocessed foods that are high in protein, fiber, and healthy fats. These foods will be more satisfying and nutritious than junk food.
3. Include Protein: Proteins are the building blocks of your body. Consuming enough protein can help you maintain your weight and muscle mass.
4. Stay Active: Regular physical activity, including strength training, can help you build muscle and promote a healthy appetite.

Remember, it's important to gain weight in a healthy way. Avoid bingeing on sugary or fatty foods just for the sake of adding calories. Try to focus on nutrient-rich foods, as these will also provide you with the vitamins and minerals your body needs.

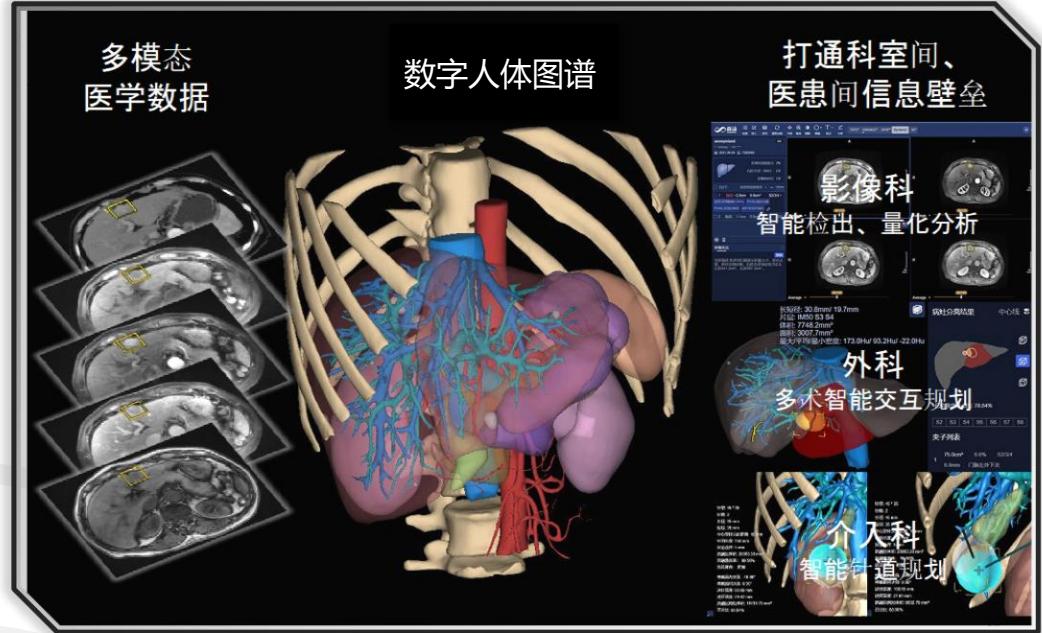
GPT 4

# Advantage of large models: Long-tail problems

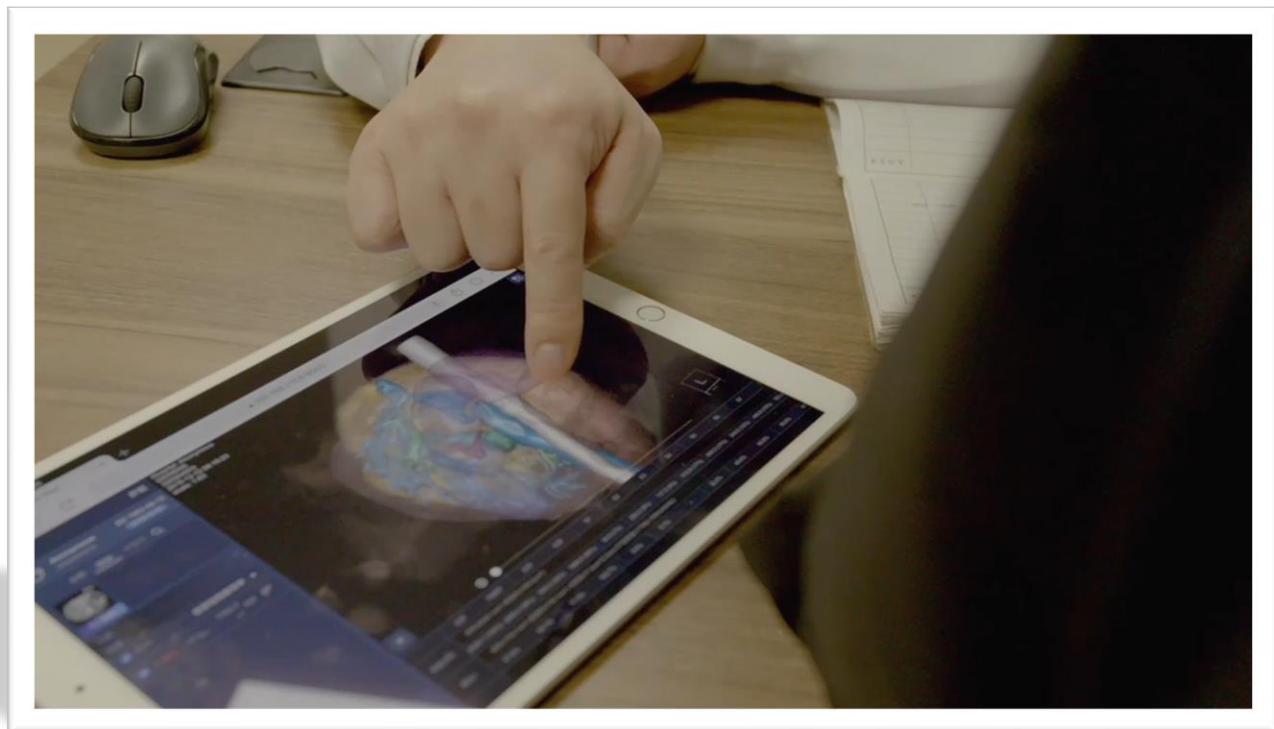
Published achievements in digital human medicine, which include 200 anatomical targets of human organs. The data modalities involved include: MRI, Ultrasound, DSA, CT, DR. Over 200+ sections of vascular networks between human organs have been restored, the precision of which surpasses the domain of what human doctors can visually distinguish between arteries and veins. This has reduced the amount of annotated data required for downstream tasks to less than a hundred instances. The model and some finely annotated data are open-source.



# Advantage of large models: Long-tail problems



Based on a full-chain AI technology stack, we construct a digital human body atlas to achieve efficient and adaptive diagnosis and treatment.



# Advantage of large models: Data security, MaaS

## Model-as-a-Service

### Automated data annotation

Intelligent annotation efficiency increased by a hundredfold

### Deployment of large model inference

Large model inference efficiency improved by 100%

### Parallel training of large models

Single cluster with 4000 cards - 500 billion dense parameter model

### Incremental training of large models

Incremental fine-tuning cost reduced by 90%

### Developer efficiency

Open model and AI development toolchain

Intelligent annotation efficiency increased by a hundredfold

Large model inference efficiency improved by 100%

Single cluster with 4000 cards - 500 billion dense parameter model

Incremental fine-tuning cost reduced by 90%

Open model and AI development toolchain

## Training

- Hardware: 384 pieces of A100, with a single card memory of 80GB, accommodating the batch training of 175 billion parameters
- Cost: \$9.6 million (48 servers, \$200,000 per server)
- Rental: Single training round ~\$3 million (AWS p4d × 2 months)

## GPT-3.5

Model	Input	Output
4K context	\$0.0015 / 1K tokens	\$0.002 / 1K tokens
16K context	\$0.003 / 1K tokens	\$0.004 / 1K tokens

## Inference

- Hardware: 8 pieces of A100 for single operation, capable of generating about 15-20 words per second (before deployment optimization)

## GPT-4

Model	Input	Output
8K context	\$0.03 / 1K tokens	\$0.06 / 1K tokens
32K context	\$0.06 / 1K tokens	\$0.12 / 1K tokens

## Estimated Resources

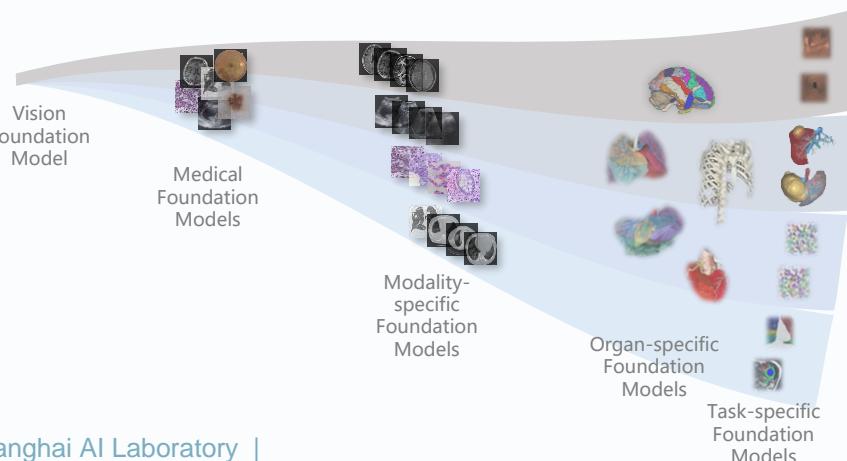
- Training from scratch for a trillion-parameter model: 500-1000 pieces of A100
- Fine-tuning for a trillion-parameter model: 50-100 pieces of A100

# Our Related Work: Medical Foundational Models



The first multi-level, multi-scenario, high-precision and implementable medical basic model group "OpenMEDLab" has recently been open sourced, integrating the world's top AI research and development capabilities, massive medical data and medical expert knowledge.

The first group of basic models released includes basic models trained on **more than 10 data modalities** such as **medical images**, **medical texts**, **biological information**, and **protein engineering**. Aiming at promoting cross-field, cross-disease, and cross-modal scientific research breakthroughs based on basic medical models, while helping to solve long-tail problems in the medical field and promote the industrial implementation of large medical models.



## World's first open-source platform for medical foundation models "OpenMEDLab"

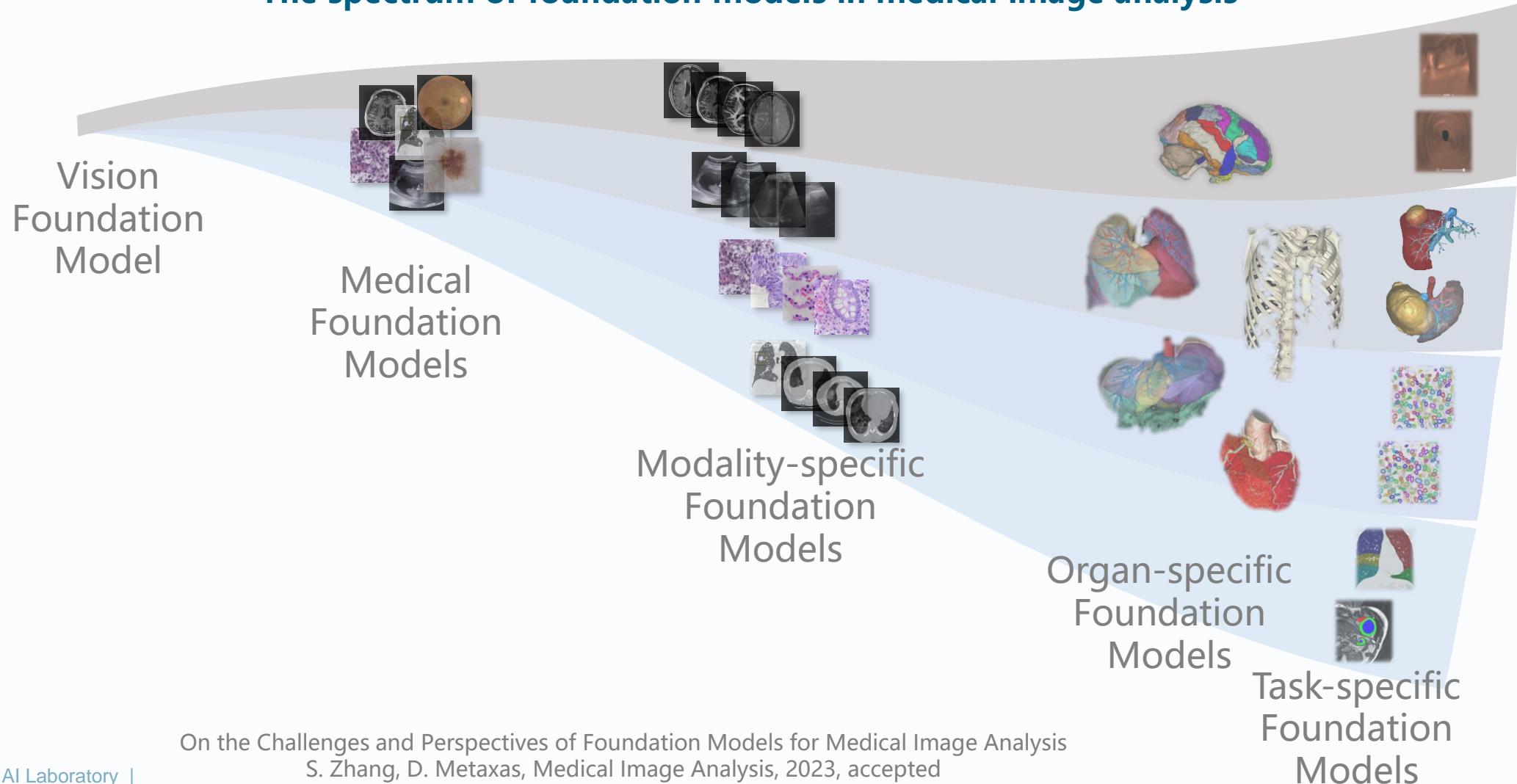
<https://github.com/openmedlab>



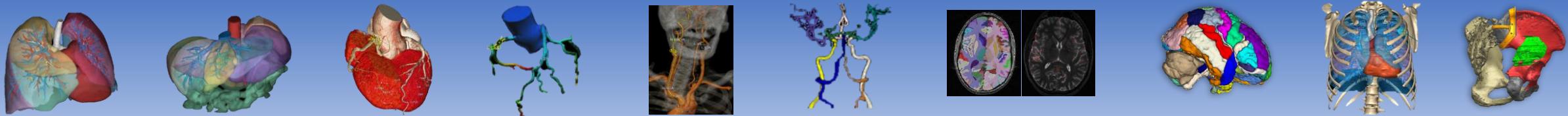
- The large Chinese medical language model, based on massive medical knowledge and **real doctor-patient interaction dialogue data**, can provide multi-scenario and multi-round conversation capabilities such as **guidance**, **consultation**, **health consultation**, and **assisted decision-making**.
- A series of image model groups can achieve cutting-edge research and clinical tasks such as **high-precision detection**, **segmentation**, and **classification** for different modalities such as **radiological images**, **pathological images**, **endoscopy**, and **ultrasound images**, and efficiently enable solutions to clinical medical imaging-based problems. The model is lightweight and can meet the clinical deployment and convenient application of innovative research.
- In terms of biomedicine, the large-scale protein sequence model is developed. It can help design and synthesize highly stable and bioactive proteins by integrating biomedical expertise, solving the common problems of long time and high cost in the development of new protein drugs.

# Our Related Work: Spectrum of Foundation Models

## The spectrum of foundation models in medical image analysis



# Our Related Work: Vision Foundational Model



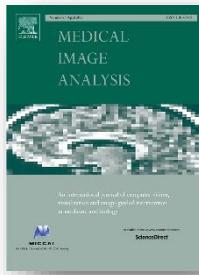
Pathology

Ultrasound

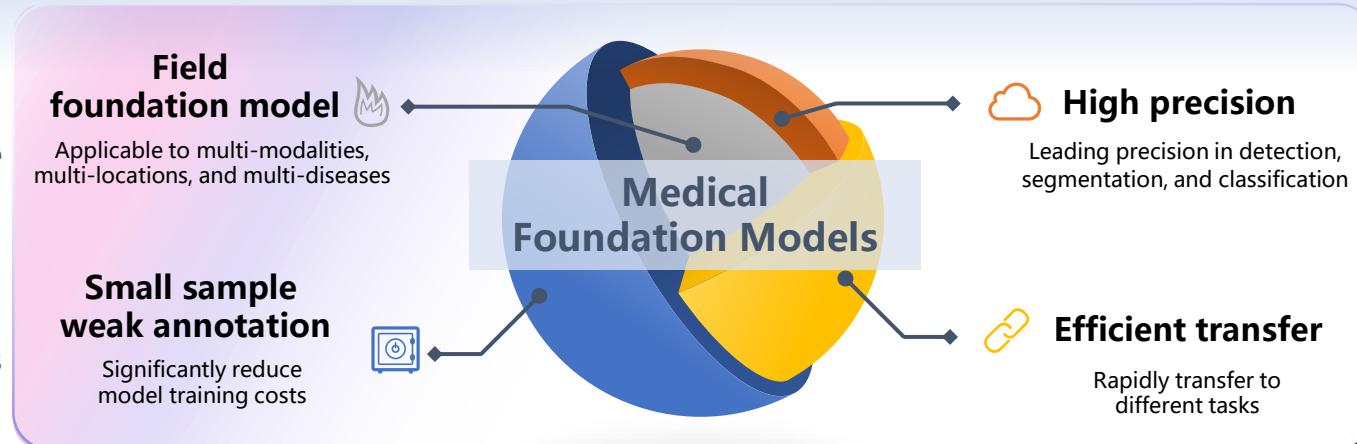
CT

MR

Electrocardiogram



Building global academic influence  
Medical Image Analysis  
Special issue on foundational models IF 13.8



WAIC 2023 officially released

Grand Challenge  
International challenge

# Our Related Work: Vision Foundational Model

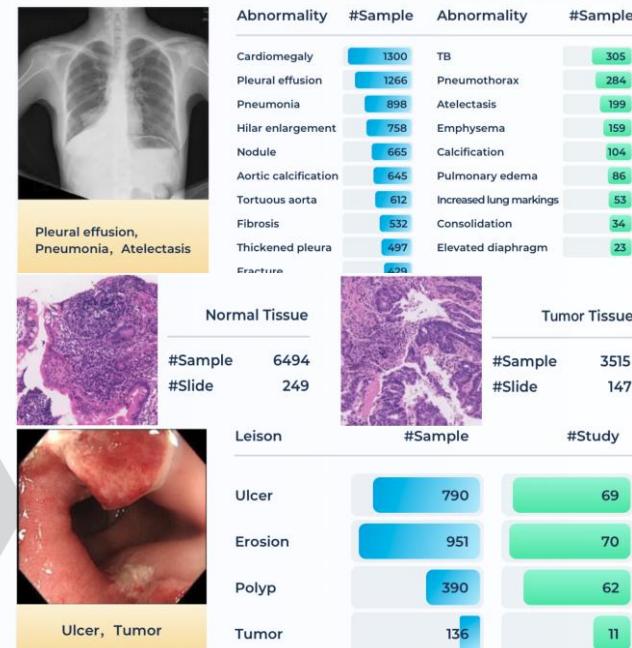
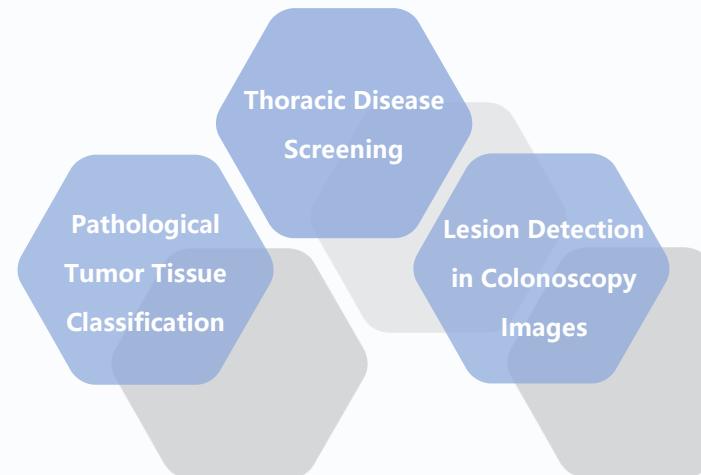
## Grand Challenge

A platform for end-to-end development of machine learning solutions in biomedical imaging.

82,000+ users | 356 challenges | 2,750 algorithms

powered by AWS

### Three Medical Image Classification Tasks



<https://medfm2023.grand-challenge.org/medfm2023/>

Shanghai AI Laboratory |

## MedFMC: Foundation Models for Medical Image Analysis Challenge

This challenge aims to promote technological innovation and explore new learning paradigms in the field of medical image analysis.

Using a small number of data (1-shot, 5-shot, 10-shot) provided by the challenge, transferring the large-scale pre-trained general foundation model to three medical image classification tasks.



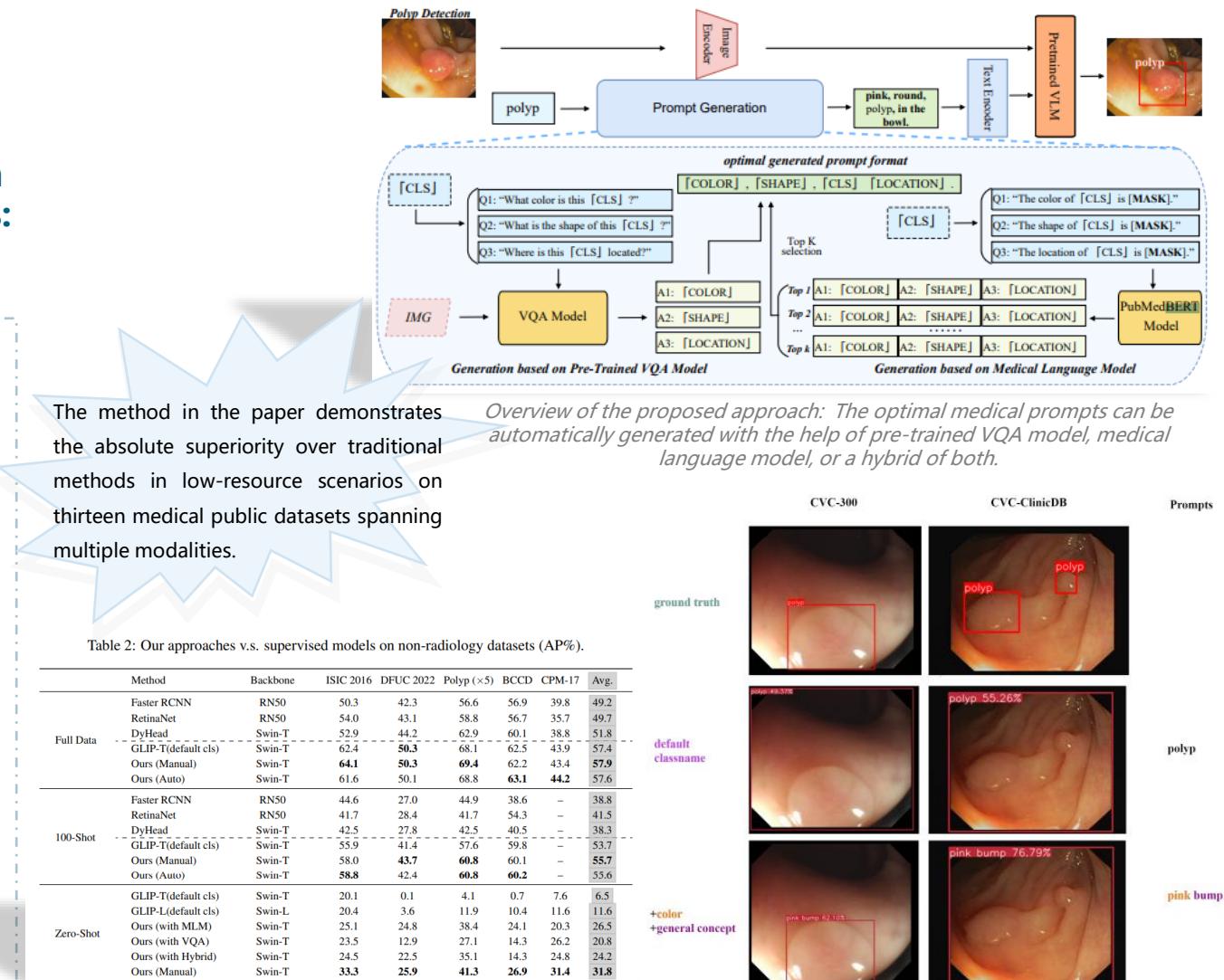
"A Real-world Dataset and Benchmark For Foundation Model Adaptation in Medical Image Classification"  
by X. Wang, D. Wang, L. Wang, M. Li, S. Zhang, et al. Scientific Data. 2023 Accepted

# Our Related Work: Vision Foundational Model



## Medical Image Understanding with Pretrained Vision Language Models: A Comprehensive Study

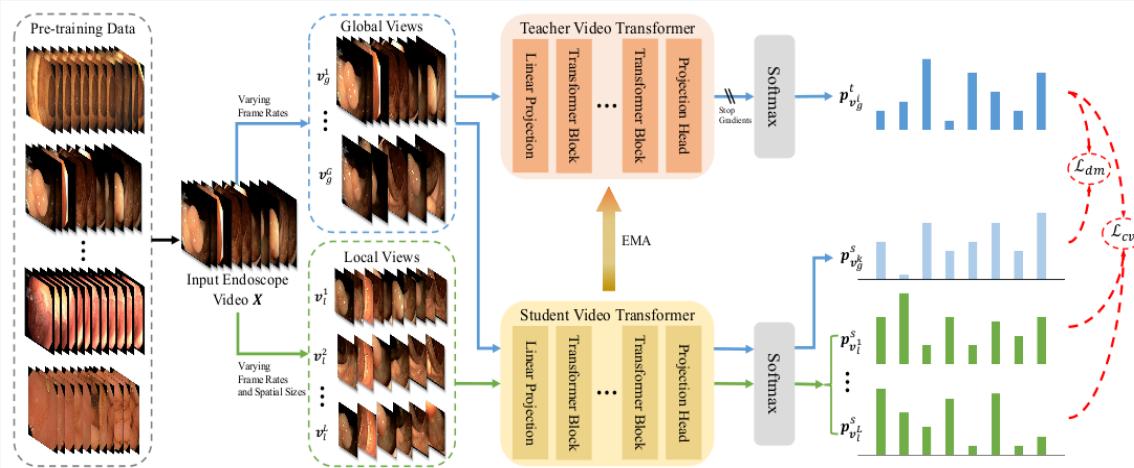
- As one of the pioneering work in applying the visual-language pre-trained model to medical image analysis area, this work first explored how to automatically generate appropriate **language prompts** to guide the pre-trained model identifying the target object in few-shot or even zero-shot situations.
- Using the invariance of descriptive languages such as color and shape in different fields, we can recognize some unseen concepts even in the case of zero-shot tasks. And by inducing a large language model (LLM) to complete adjectives, we successfully automates the generation of these descriptive hints.



# Our Related Work: Vision Foundational Model

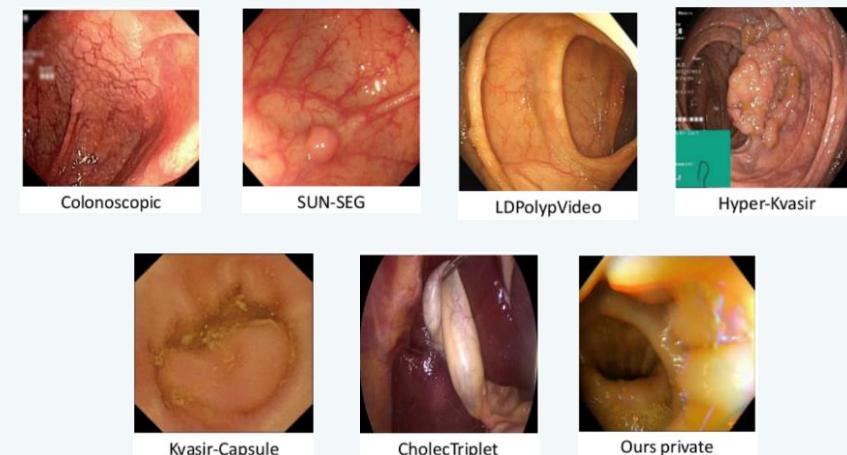
## Endo-FM

- Constructing a large-scale endoscopic video dataset consisting of over 32K video clips (5M frames), encompassing varying modalities, target organs, and disease types.
- Developing a foundation video transformer model capable of capturing both local and global long-range dependencies across spatial and temporal dimensions.
- Endo-FM achieves promising performance when applied to various downstream tasks, surpassing state-of-the-art methods by a significant margin.



**Table 2.** Comparison with other latest SOTA methods on 3 downstream tasks. We report F1 score for PolypDiag, dice for CVC-12k, and F1 score for KUMC.

Method	Venue	PolypDiag	CVC-12k	KUMC
Scratch (Rand. init.)		83.5±1.3	53.2±3.2	73.5±4.3
TimeSformer [3]	ICML'21	84.2±0.8	56.3±1.5	75.8±2.1
CORP [13]	ICCV'21	87.1±0.6	68.4±1.1	78.2±1.4
FAME [10]	CVPR'22	85.4±0.8	67.2±1.3	76.9±1.2
ProViCo [26]	CVPR'22	86.9±0.5	69.0±1.5	78.6±1.7
Qian <i>et al.</i> [28]	ECCV'22	87.6±0.6	69.1±1.2	78.1±1.9
ST-Adapter [25]	NeurIPS'22	84.8±0.7	64.3±1.9	74.9±2.9
<b>Endo-FM (Ours)</b>		<b>90.7±0.4</b>	<b>73.9±1.2</b>	<b>84.1±1.3</b>



Example frames of 6 public and ours private endoscope video datasets.

Z. Wang, C. Liu, S. Zhang, Q. Dou. "Foundation Model for Endoscopy Video Analysis via Large-scale Self-supervised Pre-train" , MICCAI, 2023.

# Our Related Work: Vision Foundational Model

# nature

## Article

### A foundation model for generalizable disease detection from retinal images

<https://doi.org/10.1038/s41586-023-06555-x>

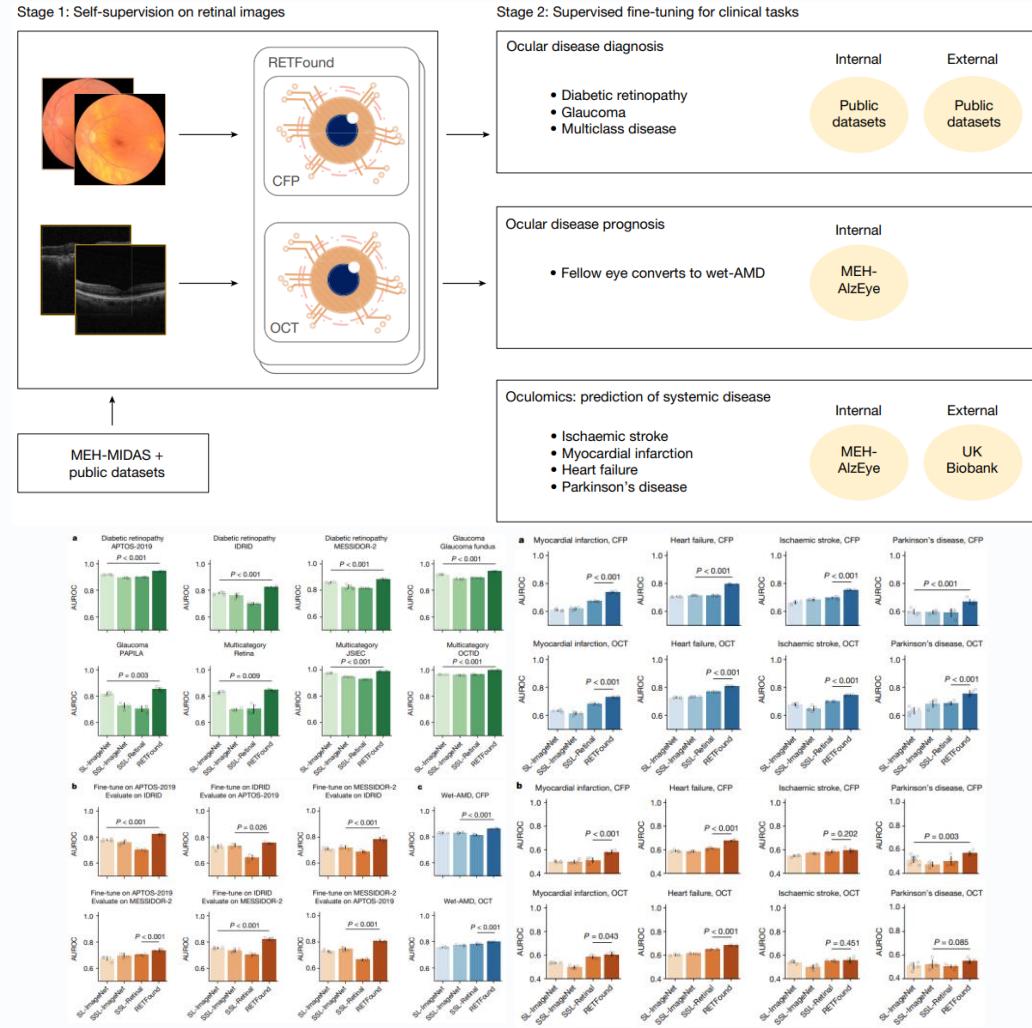
Received: 5 December 2022

Accepted: 18 August 2023

Published online: 13 September 2023

Yukun Zhou<sup>1,2,3,✉</sup>, Mark A. Chia<sup>2,4</sup>, Siegfried K. Wagner<sup>2,4</sup>, Murat S. Ayhan<sup>1,2,4</sup>, Dominic J. Williamson<sup>1,2,4</sup>, Robbert R. Struyven<sup>1,2,4</sup>, Timing Liu<sup>2</sup>, Moucheng Xu<sup>1,3</sup>, Mateo G. Lozano<sup>2,5</sup>, Peter Woodward-Court<sup>1,2,6</sup>, Yuka Kihara<sup>7,8</sup>, UK Biobank Eye & Vision Consortium<sup>\*</sup>, Andre Altmann<sup>1,3</sup>, Aaron Y. Lee<sup>7,8</sup>, Eric J. Topol<sup>9</sup>, Alastair K. Denniston<sup>10,11</sup>, Daniel C. Alexander<sup>1,2</sup> & Pearse A. Keane<sup>2,4,✉</sup>

- Investigated the development and validation of a foundational model for retinal images, RETFound, with the objective of mitigating the AI model's heavy reliance on massive labeled data and bolstering its generalization capability in disease detection tasks.
- RETFound stands as the inaugural foundational model in the field of ophthalmology, adopting one of the most extensive training datasets and being validating on the most comprehensive set of validation tasks. This open-source model can be tailored to custom downstream tasks, promoting diverse ophthalmological research.
- Trained on 1.6 million unlabeled retinal images by a manner of Self-Supervised Learning and later adapted to disease detection tasks based on explicit labels for the detection of various diseases.
- RETFound outperformed the other pretraining strategies using only 10% of labeled training data and showed consistently high adaptation efficiency, demonstrating the potential of this foundation model in alleviating data shortages and adapting to downstream tasks.
- RETFound provides a generalizable solution to improve model performance and alleviate the annotation workload of experts to enable broad clinical AI applications from retinal imaging.



# Our Related Work: Large Language Models



- **Large-scale Training:** A model optimized on approximately 4,000,000 instruction datasets from both the Chinese medical domain and general domain.
- **Comprehensive Support for Chinese Medical NLP Tasks:** Including health education, physician examinations, report interpretation, medical record structuring, and simulated diagnostics and treatments.



# Our Related Work: Large Language Models



### Self-supervised Pre-training

**Medical Domain Corpus:** Textbook, Guideline, EHR, papers, etc. ~12B tokens

**General Domain Corpus :** Wiki, Webpage, News, etc. ~230B tokens



### Instruction Tuning

**Medical Domain Instruction Tuning Dataset:**  
1 language, 17 tasks, 19 types of prompt, 2.7B

**Generic Domain Instruction Tuning Dataset:**  
46 languages, 16 tasks, 16 types of prompt, 5.3M

**Medical Domain QA Dataset:** ~4B tokens

**Generic Domain QA Dataset:** ~1.2B tokens



### Plugins

**General Plugins:** Search Engine, Sci Calculator, ~600M tokens

**Medical Plugins:** Knowledge Retrieval, BMI Calculation, ~109M tokens

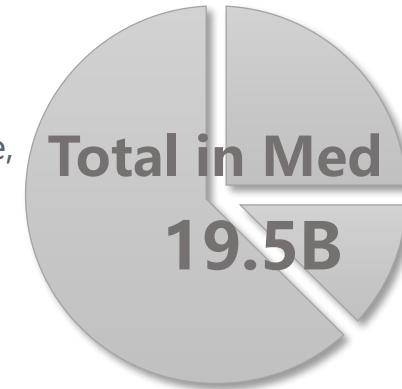


### RLHF

**Medical Domain Dialog Dataset:**  
7 scenarios, ~49M tokens

**Generic Domain Dialog Dataset:**  
16 languages, ~72M tokens

**Generic Domain Task Dataset:**  
Chinese & English, 16 tasks, ~35M tokens



# Our Related Work: Large Language Models

## Medical Domain Supervised Fine-tuning

Collect demonstration data and train a supervised policy

- Instructions: NER, plugins, etc.
- Multi-round dialog: Consultation, triage, etc.
- QA: physical examination, drug consultation, etc.
- Rejection: refuse to respond non-medical questions

## Reward Model Training

Collect comparison data and train a reward model

Design a prompt for self-evaluation

## RL Model Training

Optimize a policy against the reward model with RL

Improve the basic PPO framework

Balance the exploration & exploitation

### SFT Prompt

#### [USER]

检查途径: 经体表 图像等级: 丙 检查项目: 二维 M型 彩色 多普勒(脉冲式 连续式) 组织多普勒 一、M型主要测值(单位mm): 名称 测量值 正常参考值 主动脉根部内径 33 20-37 左房内径 35 19-40 左室舒张末期内径 48 35-56 左室收缩末期内径 28 20-37 室间隔厚度 11 6-11 左室后壁厚度 10 6-11二、二维超声心动图描述: 1.各房室无明显扩大。2.室间隔基底部偏厚, 约12mm。静息状态下左室壁各节段收缩活动未见明显异常。3.各心瓣膜未见明显增厚, 开放不受限。三、彩色多普勒超声描述: 1.房、室间隔水平未见明显分流。2.二尖瓣轻微反流。舒张期经二尖瓣口血流:E=63cm/s, A=69cm/s, E/A=0.9。3.主动脉瓣未见明显反流。4.三尖瓣轻微反流, 最大反流速度约2.0m/s, 估测肺动脉收缩压约27mmHg。5.左室流出道收缩期血流速度未见明显异常。四、左心功能测定: 名称 测量值 名称 测量值 左室舒张末期容量(ml) 106 左室收缩末期容量(ml) 30 左室射血分数(%) 71 左室短轴缩短率(%) 40 每搏输出量(ml) 76 五、组织多普勒检查: 二尖瓣瓣环水平: 室间隔侧 E'=6.9cm/s, E/E'=9.2。左室侧壁 E'=11.3cm/s, E/E'=5.6。

根据上文中信息, 判断主动脉根部内径是什么? 请提取文中对应的值

#### [AI]

根据上文中信息, 主动脉根部内径为33毫米, 正常参考值为20-37毫米。

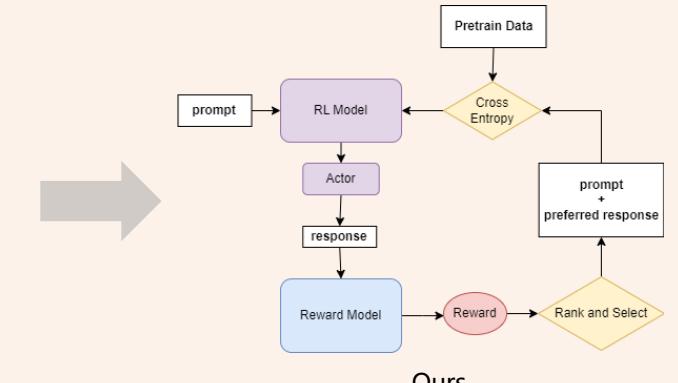
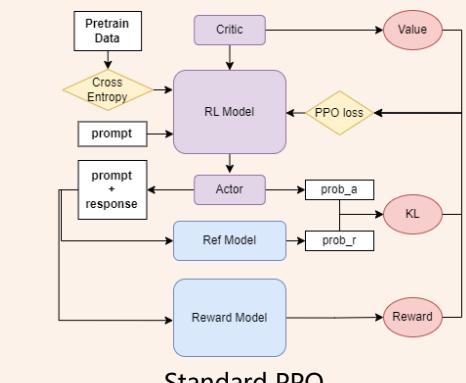
### RLHF Labeling Tool

#### [USER]

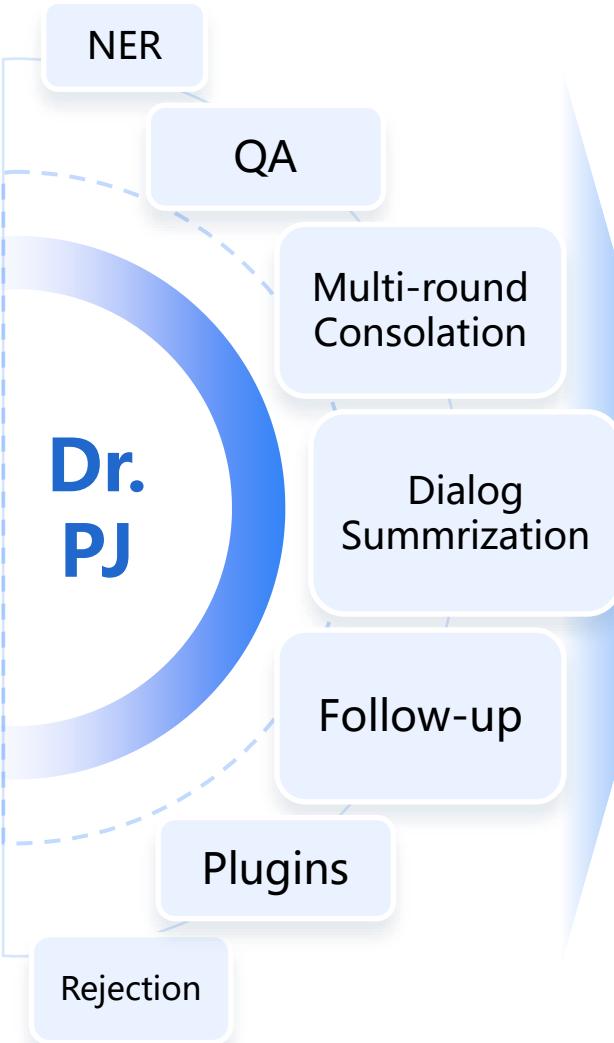
血常规里面白细胞比较低是怎么回事

1/12	白细胞低可能是由多种原因引起的, 包括病毒感染、药物副作用、疾病等。在您的情况下, 白细胞的幅度较小 (从12.6万到13.3万) 且未见其他异常指标 (如白细胞计数、血小板数量等), 因此需要进一步的检查和评估来确定具体的原因。在您的情况下, 白细胞比正常值并不一定意味着一定是疾病的征兆。但是, [AI] 仅凭一项血液检测结果无法确定具体的病因, 需要结合更多的检查和评估才能确定具体的原	1/12	白细胞低可能是由多种原因引起的, 包括病毒感染、过度使用抗菌药物、某些疾病导致的免疫系统成功抑制等。在您的情况下, 如果出现白细胞比较低的症状并不一定意味着一定是疾病的征兆。但是, [AI] 仅凭一项血液检测结果无法确定具体的病因, 需要结合更多的检查和评估才能确定具体的原	3/12	白细胞低可能是由多种原因引起的, 包括感染、疾病或药物副作用等等。在血常规检查中显示白细胞低于正常范围通常被视为异常体征之一。因此, 仅凭一项血液检测结果无法确定具体的病因, 需要结合更多的检查和评估才能确定具体的原	4/12	白细胞低可能是由多种原因引起的, 包括感染、疾病或药物副作用等等。在血常规检查中显示白细胞的数值低于正常范围通常被视为异常体征之一。因此, 仅凭一项血液检测结果无法确定具体的病因, 需要结合更多的检查和评估才能确定具体的原		
5/12	白细胞 (也称为白血球) 是一种血液中的免疫系统细胞, 它的作用是抵御细菌和病毒的入侵。当身体受到细菌的感染时, 白细胞会释放出一种叫做氯气的物质来攻击病菌或病原体。因此, 如果血常规中白细胞 [AI] 用作评估体内免疫系统的功能是否正常, 如果值偏低, 可能有以下原因:	5/12	白细胞 (也称为白血球) 是一种血液中的免疫系统细胞, 它的作用是抵御细菌和病毒的入侵。当身体受到细菌的感染时, 白细胞会释放出一种叫做氯气的物质来攻击病菌或病原体。因此, 如果血常规中白细胞 [AI] 用作评估体内免疫系统的功能是否正常, 如果值偏低, 则表示免疫力低下, 易发生感染或其他疾病	7/12	白细胞 (也称为白血球) 是一种血液中的免疫系统细胞, 它的作用是抵御细菌和病毒的入侵。当身体受到细菌的感染时, 白细胞会释放出一种叫做氯气的物质来攻击病菌或病原体。因此, 如果血常规中白细胞 [AI] 用作评估体内免疫系统的功能是否正常, 如果值偏低, 则表示免疫力低下, 易发生感染或其他疾病	8/12	白细胞 (也称为白血球) 是一种血液中的免疫系统细胞, 它的作用是抵御细菌和病毒的入侵。当身体受到细菌的感染时, 白细胞会释放出一种叫做氯气的物质来攻击病菌或病原体。因此, 如果血常规中白细胞 [AI] 用作评估体内免疫系统的功能是否正常, 如果值偏低, 则表示免疫力低下, 易发生感染或其他疾病		
9/12	白细胞是一种血液中的免疫系统组成部分, 其数量可以用来评估是否有感染或其他疾病的存在的。如果白细胞比较低, 这可能意味着身体免疫力下降, 容易出现一些感染性疾病或自身免疫性疾病的可能性增加	9/12	白细胞是一种血液中的免疫系统组成部分, 其数量可以用来评估是否有感染或其他疾病的存在的。如果白细胞比较低, 这可能意味着身体免疫力下降, 容易出现一些感染性疾病或自身免疫性疾病的可能性增加	10/12	白细胞低可能是由多种原因引起的, 包括感染、药物副作用、疾病等。在血常规检查中显示白细胞低于正常范围, 建议您咨询医生以获取更多信息并进行进一步的检查和评估,	11/12	白细胞低可能是由多种原因引起的, 包括感染、药物副作用、疾病等。在血常规检查中显示白细胞低于正常范围, 建议您咨询医生以获取更多信息并进行进一步的检查和评估,	12/12	白细胞低可能是由多种原因引起的, 包括感染、药物副作用、疾病等。在血常规检查中显示白细胞低于正常范围, 建议您咨询医生以获取更多信息并进行进一步的检查和评估,

### RL Models



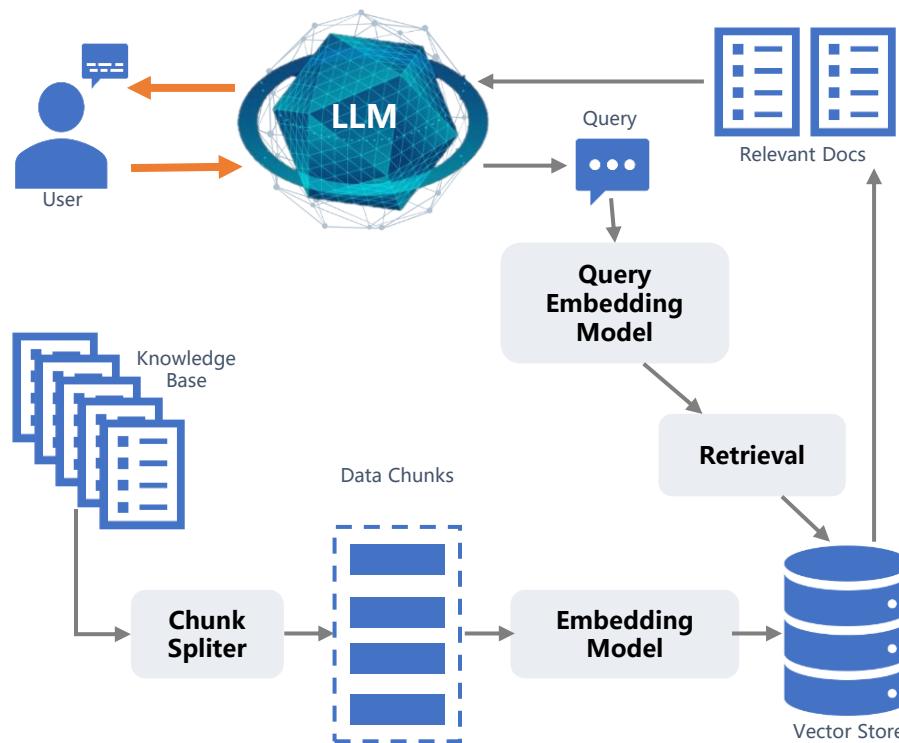
# Our Related Work: Large Language Models



# Our Related Work: Large Language Models

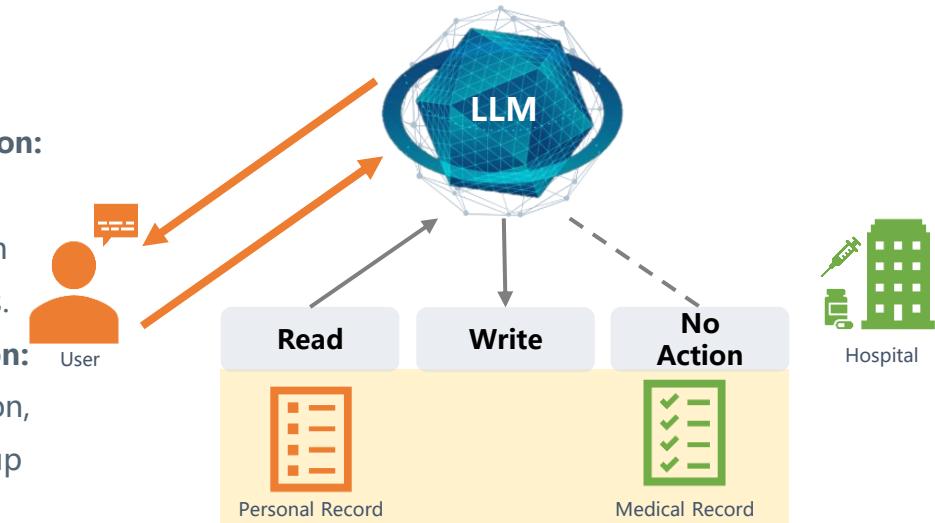
## Customized Knowledge Base

The model can search within a specified knowledge base based on the question posed. It summarizes the content of the related documents retrieved and generates a response to the question.



## Long-range Memory

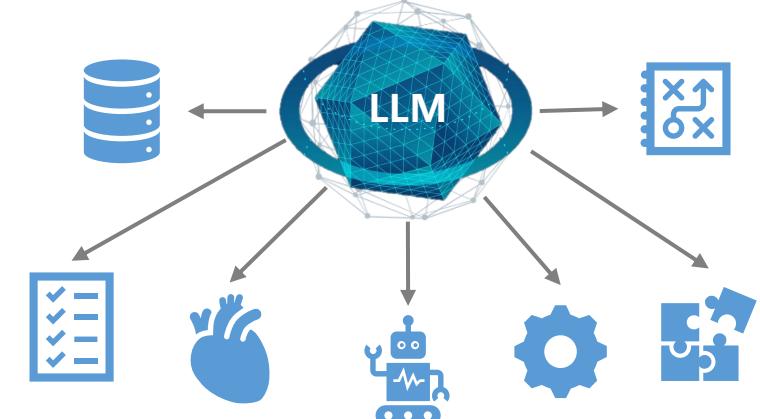
- Storing personal information:** No maximum input length restriction, retains long-term health records of individuals.
- Storing medical information:** Maintains patient information, supports automatic follow-up scenarios.



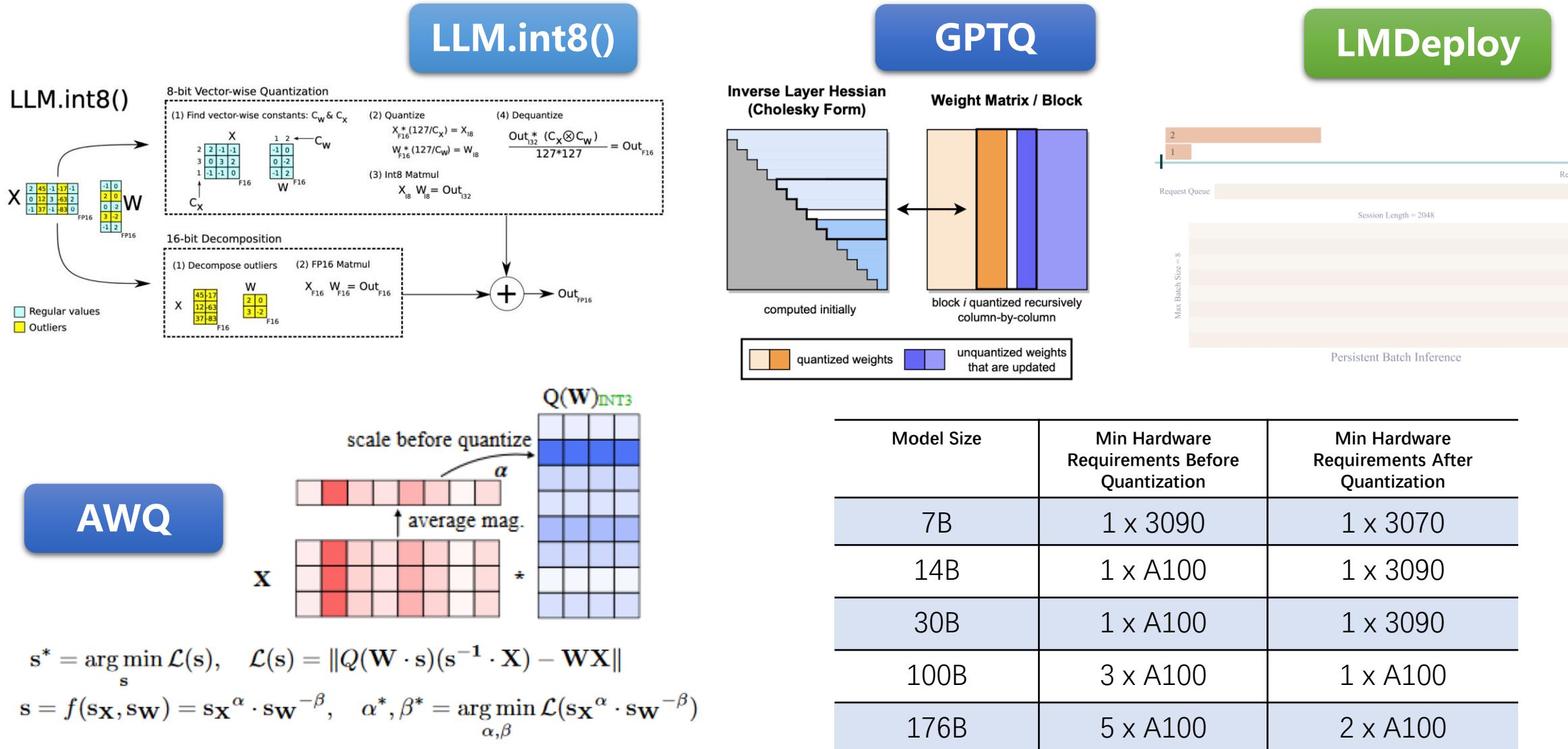
## Multi-agent Scheduling

In current complex AI medical diagnostic and treatment processes, there often exist multiple specialized small models, such as data quality control, vessel segmentation, lesion detection, and critical judgment, etc.

Having LLM invoke its own capabilities to complete simple tasks, and coordinating the scheduling of specialized small models for complex tasks, will make complex reasoning across multiple departments and modalities in the diagnostic and treatment phases possible.



# Our Related Work: Large Language Models



# Our Related Work: Large Language Models



Detailed and accurate “popular science” drug consultation and Q&A capabilities

Ability to recommend drugs through light consultation covering common diseases and symptoms

Medication consultation  
LLM

Specific training and optimization for authoritative drug data

Powerful context-related capabilities and multi-turn dialogue capabilities

## ■ Multi-terminal product application: mobile + PC web



## ■ Medication consultation cases

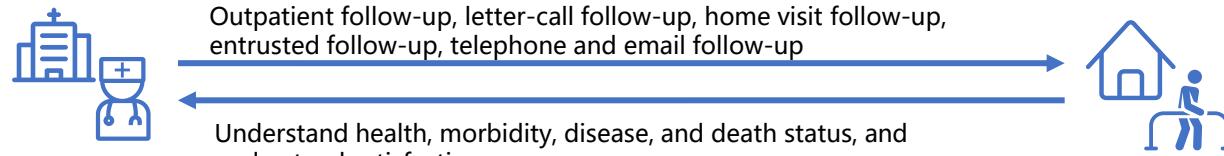


- The LLM gives medication consultation powerful multi-round dialogue and context-related capabilities.
- Authoritative drug data enables LLM to gain powerful drug understanding and reasoning capabilities
- Simulate conversations between real doctors and pharmacists, eliminating the need to prompt for drug names and related details in each conversation

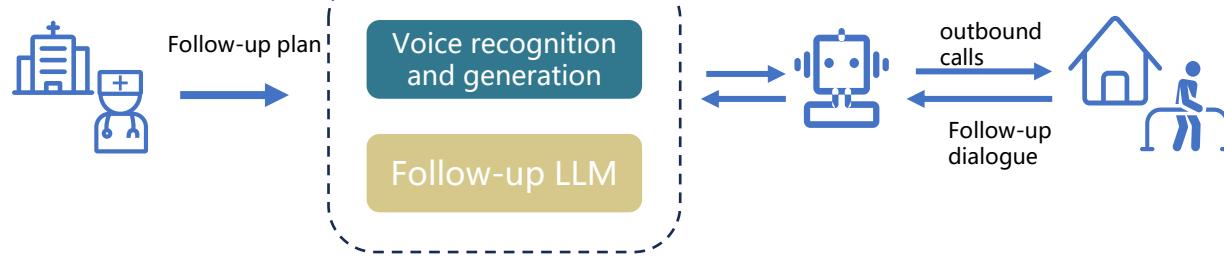
# Our Related Work: Large Language Models

Based on LLMs and intelligent voice technology, it realizes functions such as batch outbound calls, human-computer interaction, and automatic statistics, helping hospitals complete follow-up of scientific research, specialized disease and caring work, reducing hospital follow-up costs and improving the work efficiency of medical staff.

## Traditional Follow-up pipelines



## AI for Follow-up



More understandable questions

Analyze non-standard and out-of-range answer content

Understand and connect contextual questions and answers

Thinking and answering out-of-scope questions

More humane follow-up interaction

Stronger conversational understanding

More comprehensive information collection



# Our Related Work: Foundation Models for Science



**Core scientific question: Can a general AI model for protein engineering be constructed to directly design protein products that meet application requirements (end-to-end)?**

1. First, use the sequence and structure information that already exists in nature for pre-training. Our model: ~600 million parameters, billions of protein training data.



**Human Corpus Pre-training  
(gpt 3.0: 175 billion parameters, hundreds of billions of training data)  
Analogy: Chat GPT**

2. Then, use the labeled related data set (such as ocean data) for supervised learning.



**Standard Dataset Optimization  
(supervised learning)**

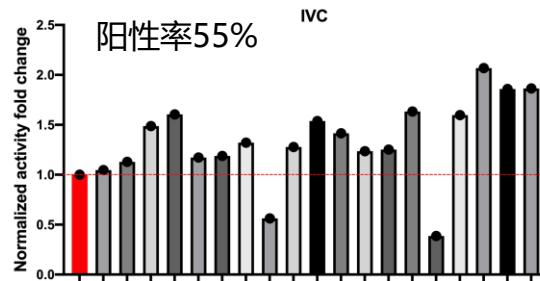
3. Finally, use a small amount of wet experimental data (dozens) for model tuning.



**Reinforcement Learning for Expert Decision Making**

# Our Related Work: Foundation Models for Science

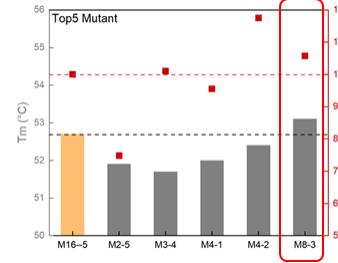
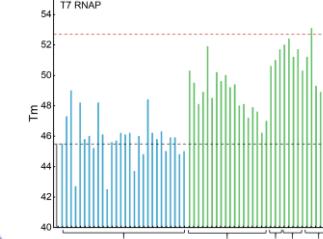
## Crispr Case 12 (gene editing proteins, improve activity)



One round of design + experimental iteration, the positive rate of the design is 55%. The editing efficiency of the best mutant is increased by 2 times, achieving the best of its kind.

## T7RNAP (RNA drug polymerase, improves thermostability)

Improve high temperature stability and reduce immunogenic impurity products

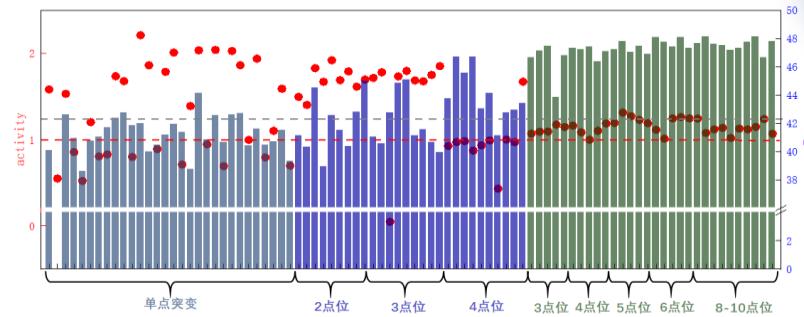


It took three and a half months and two rounds of iterations to develop a multi-site mutant with a melting temperature increased by 7 °C. Its stability and activity are comparable to those of the engineering protein developed by the leading company.

## Case 1 Case 2

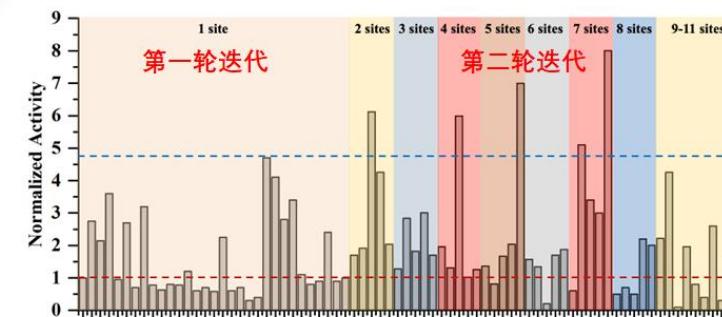
## Case 3 Case 4

### An IVD enzyme (improves thermostability)



Three-round design + wet experiment, the positive rate of the design (30% in the first round), the second round (60%), and the third round (100%). The melting temperature of the best multi-site mutant was 6°C higher than that of the wild type (WT).

### Ago protein (nucleic acid detection/tumor detection enzyme, improve activity)



Two-round design + wet experiment, the positive rate of the design (first round 40%), second round (90%). The optimal mutant has an 8-fold increase in activity compared to the wild type (WT).

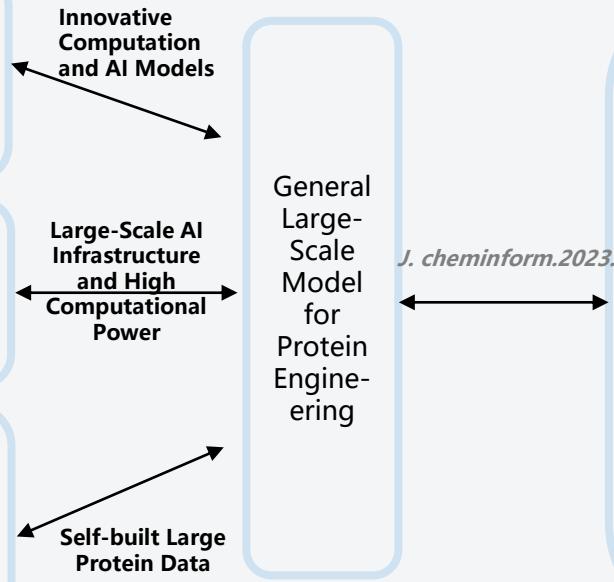
# Our Related Work: Foundation Models for Science

Institute of Natural Sciences,  
Shanghai Jiao Tong University  
*JCP*, 157 (2022), 144102  
*JCP*, 156 (2022), 014114  
*NeurIPS*, 2022

Shanghai AI Lab  
Large-scale Natural Language Model + High Computational Power  
*Brie. in Bio.*, 2022

Integration of public data, scientific research  
collaborative data, enterprise cooperative data,  
totaling: **4.7 billion** unique protein sequence  
**proprietary database**.  
vs the largest public database BFD (**2 billion**)

*Nat. Comm.*, 13, 7861 (2022)



## Research cooperation

1. Stability of a RNA polymerase (本课题组, 阳性)
2. Working efficiency of a gene editing protein (上科大免疫所, 阳性)
3. Stability/activity of a nucleic acid detection enzyme (交大微生物国重室, 阳性)
4. Stability of a kidney function test enzyme (本课题组, 阳性)
5. Activity of non-natural nucleic acid polymerase (中科院基础医学与肿瘤所, 阴性)
6. Activity of artemisinin β-farnesene synthase (湖北大学酶工程国重室, 阴性)

## Enterprise cooperation:

1. Alkaline resistance of an antibody (金赛药业)
2. Substrate selectivity of an engineered enzyme (瀚海新酶)
3. Stability of a pharmaceutical enzyme (康宁杰瑞)
4. Activity of Nucleic Acid Polymerase (中元汇吉)
5. reverse transcriptase (近岸蛋白)

Model type	Large-scale model training in Natural Language Processing			
Development Team	Facebook (USA)		Salesforce Research (USA)	SJTU (Research Group)
Model Name and Parameter Quantity	ESM-1V (650 Million Parameters)	ESM2 (150 Million Parameters)	Progen (Xlarge) (6.4 Billion Parameters)	TemPL (650 Million Parameters)
Zero-Shot Model Accuracy (Correlation coefficient between predictions and laboratory results, for 22 protein TM datasets)	0.350	0.384	0.256	0.467

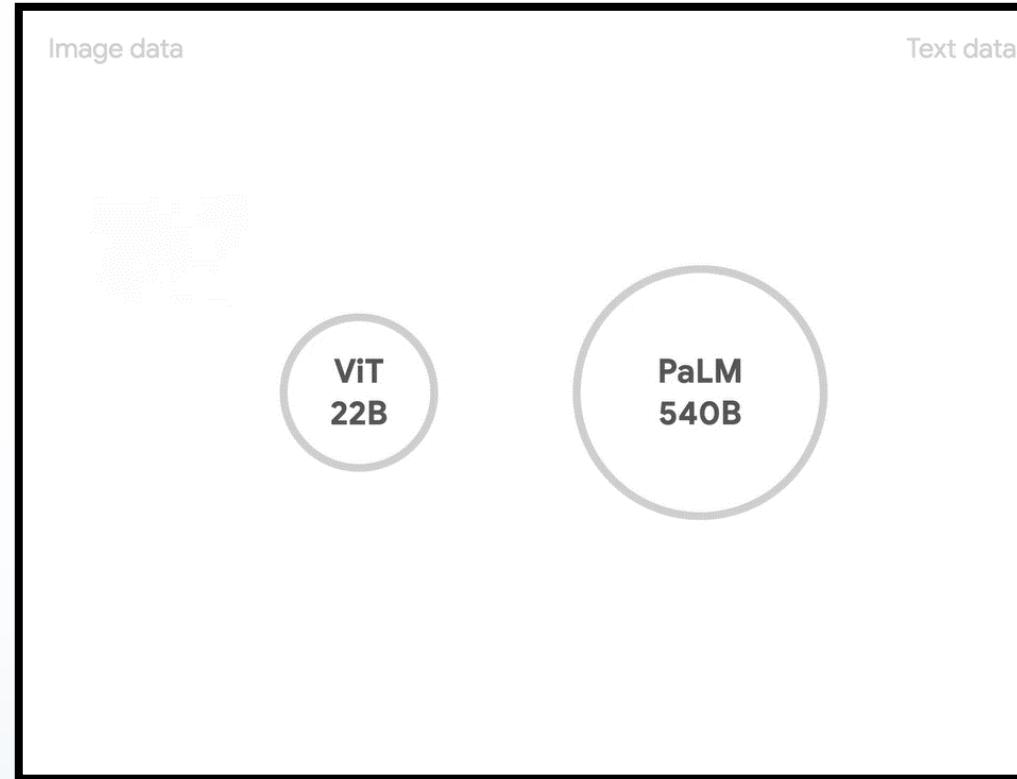
Performance Comparison with Internationally Renowned Teams or Companies

<https://arxiv.org/abs/2304.03780>

# The Future Direction of the Medical Foundational Model



## PaLM-E: A concrete multimodal language model.



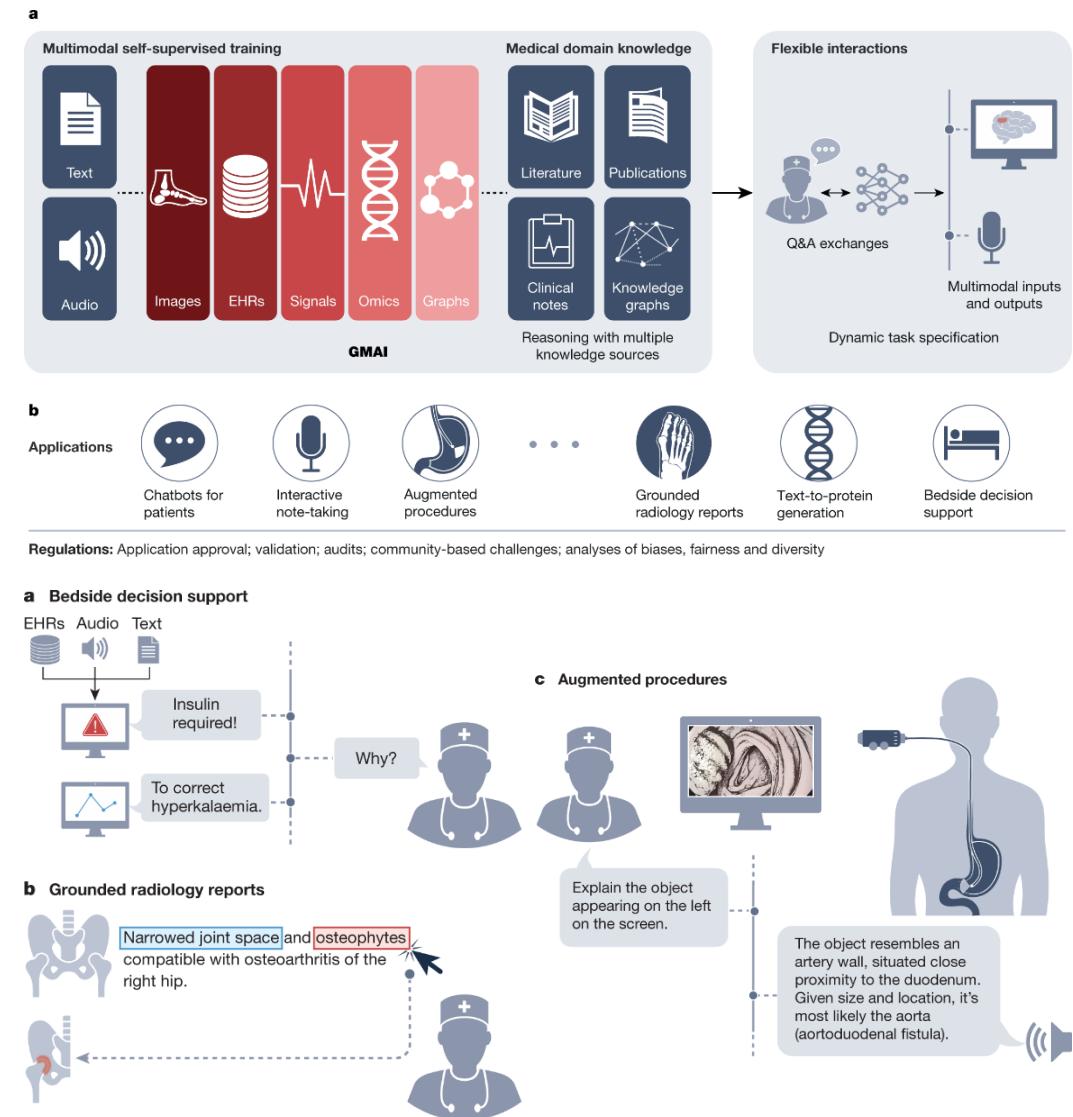
<https://palm-e.github.io>

# The Future Direction of the Medical Foundational Model

# nature

## Foundation models for generalist medical artificial intelligence

- Propose a new paradigm for medical AI, which refer to as generalist medical AI (GMAI)
- Be capable of carrying out a diverse set of tasks using very little or no task-specific labelled data.
- Flexibly interpret different combinations of medical modalities, including data from imaging, electronic health records, laboratory results, genomics, graphs or medical text
- In turn produce expressive outputs such as free-text explanations, spoken recommendations or image annotations that demonstrate advanced medical reasoning abilities



# The Future Direction of the Medical Foundational Model

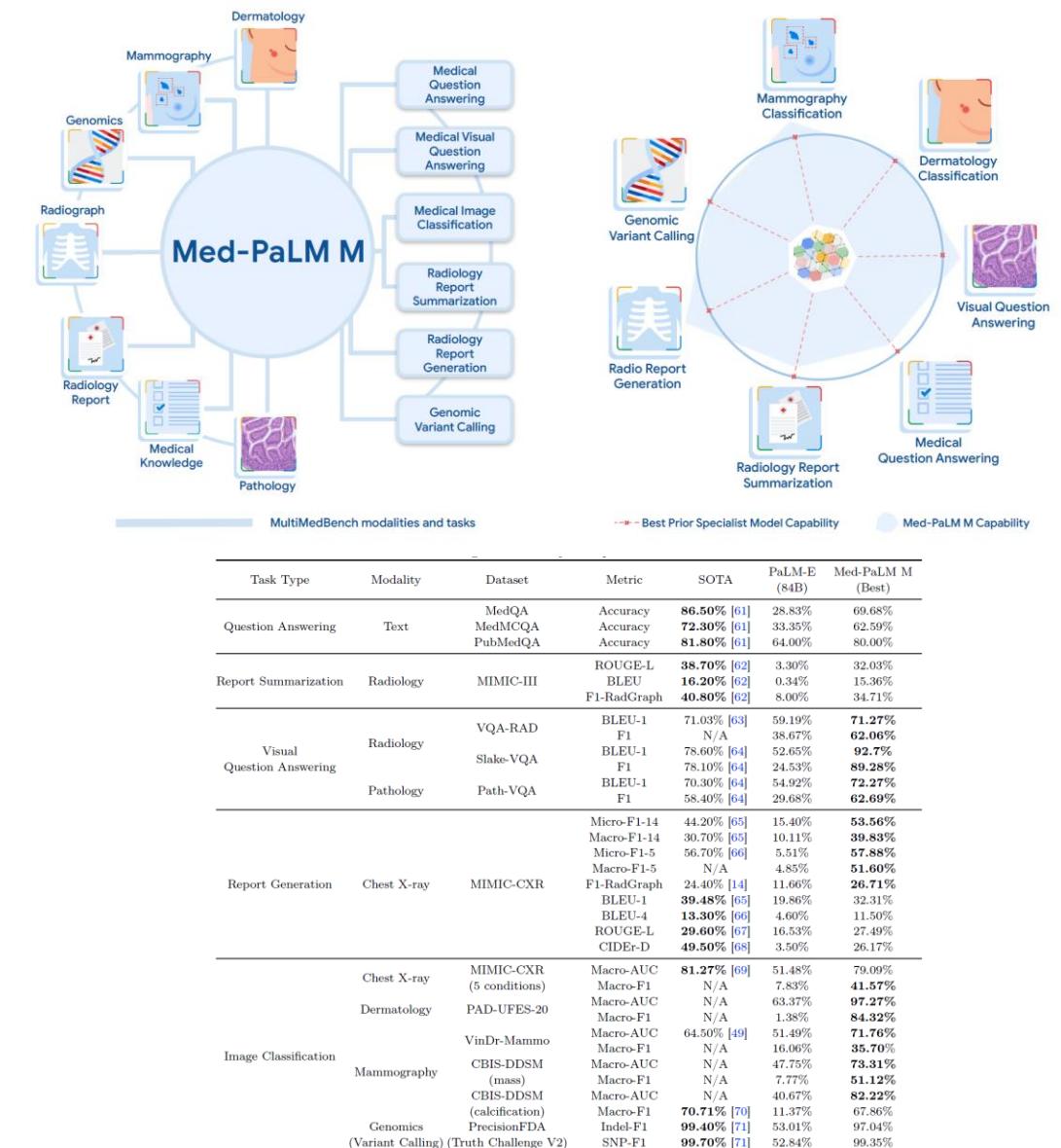


## Towards Generalist Biomedical AI

- (1) MultiMedBench, a new multimodal biomedical benchmark, is curated.
- (2) MultiMedBench encompasses 14 diverse tasks such as medical question answering, mammography and dermatology image interpretation, radiology report generation and summarization, and genomic variant calling.
- (3) Then introducing Med-PaLM M, large multimodal generative model that flexibly encodes and interprets biomedical data including clinical language, imaging, and genomics with the same set of model weights.
- (4) In a side-by-side ranking on 246 retrospective chest X-rays, clinicians express a pairwise preference for Med-PaLM M reports over those produced by radiologists in up to 40.50% of cases, suggesting potential clinical utility.

D. Driess, M. Schaeckermann, A. Karthikesalingam, V. Natarajan, et al. Towards Generalist Biomedical AI.  
arXiv:2307.14334. 2023

Shanghai AI Laboratory |



# About us

## PUBLICATION

### THE LANCET Digital Health



### nature communications

### scientific data

论文标题	期刊名称
Spatially aware graph neural networks and cross-level molecular profile prediction in colon cancer histopathology: a retrospective multi-cohort study	The Lancet Digital Health
Predicting unseen antibodies' neutralizability via adaptive graph neural networks	Nature Machine Intelligence
Mining Multi-Center Heterogeneous Medical Data with Distributed Synthetic Learning	Nature Communications
Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning	npj Precision Oncology
MedFMC: A Real-world Dataset and Benchmark For Foundation Model Adaptation in Medical Image Classification	Scientific Data
A Large-scale Synthetic Pathological Dataset for Deep Learning-enabled Segmentation of Breast Cancer	Scientific Data

**OpenMEDLab**

The diagram illustrates the OpenMEDLab platform as a central hub connected to various medical datasets and services. The central circle is labeled "OpenMEDLab". Surrounding it are several smaller circles, each representing a different dataset or service, such as "MedFM CT", "MedFM NLP", "MedFM US", "MedFM DR", "MedFM ECG", "MedFM Proteins", "MedFM MR", and "MedFM More". To the right of the diagram is a QR code.

<https://github.com/openmedlab>



Kang Li Guotai Wang Qicheng Lao



Dequan Wang Xiaofan Zhang Liang Hong

WCH, BUPT  
and UESTC

SJTU

Shanghai AI  
Lab



Junjun He Xiaosong Wang



Teams of  
PIs

Guangzhou &  
Zhejiang Lab



Yixue Li Feng Gao Wentao Zhu



CHUK, FDU,  
and CPII



Qi Dou Lijian Xu Yi Guo



Overseas



Mu Zhou Dimitris Metaxas Yukun Zhou Xiaoxiao Li

# Q & A

<https://www.shlab.org.cn>

Shanghai Artificial Intelligence Laboratory

