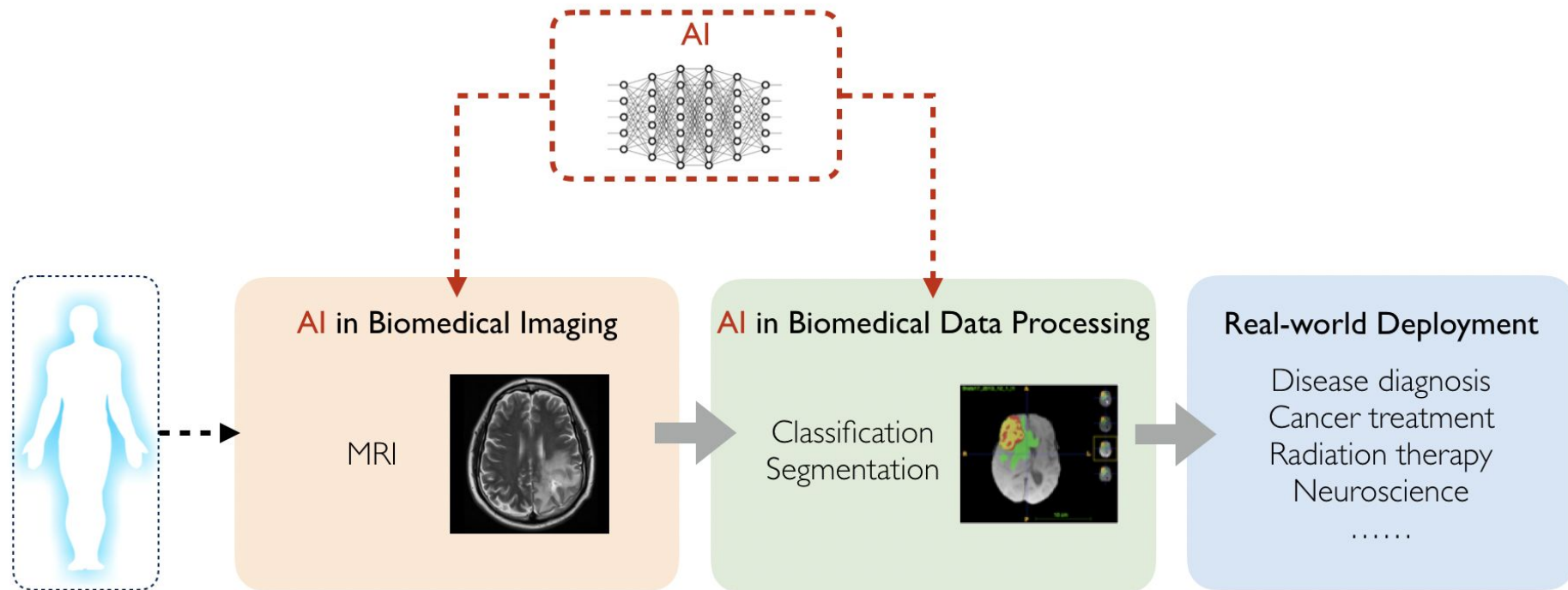


Lecture 6:

Multimodal Foundation Models

In this class:

- Part I: AI in biomedical imaging
- **Part II: AI in biomedical data processing**



Today's agenda

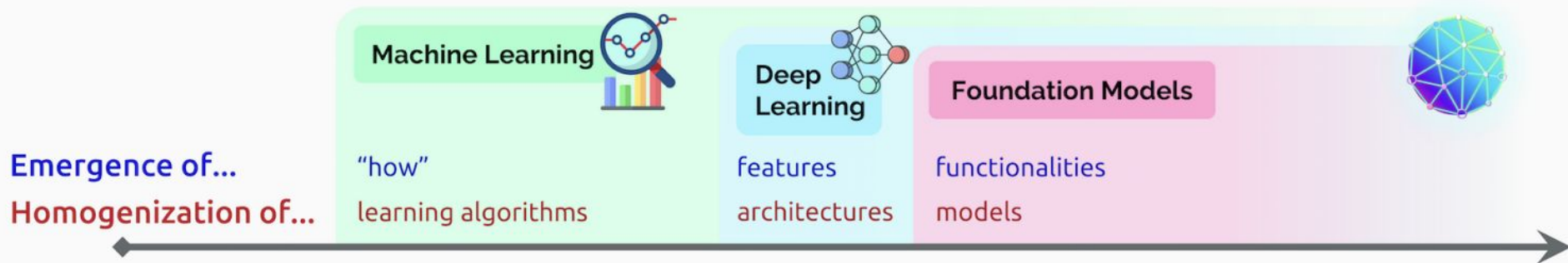
- Foundation model
- Language foundation model
- Vision foundation model
- Multimodal foundation model
- Generalist Medical AI (GMAI)

Today's agenda

- Foundation model
- Language foundation model
- Vision foundation model
- Multimodal foundation model
- Generalist Medical AI (GMAI)

Foundation model

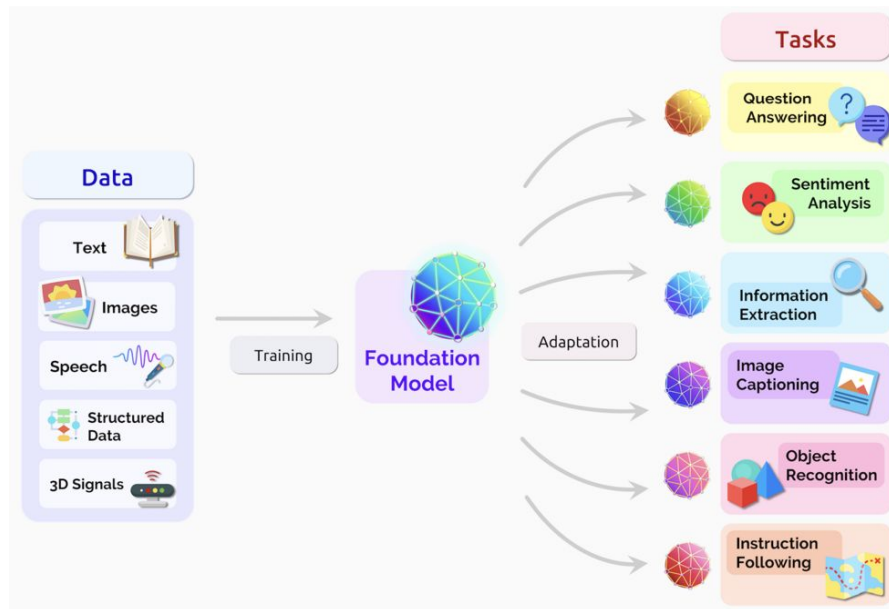
- A new paradigm for building AI systems



Bommasani, et al., On the Opportunities and Risks of Foundation Models, arXiv 2022.

Foundation model

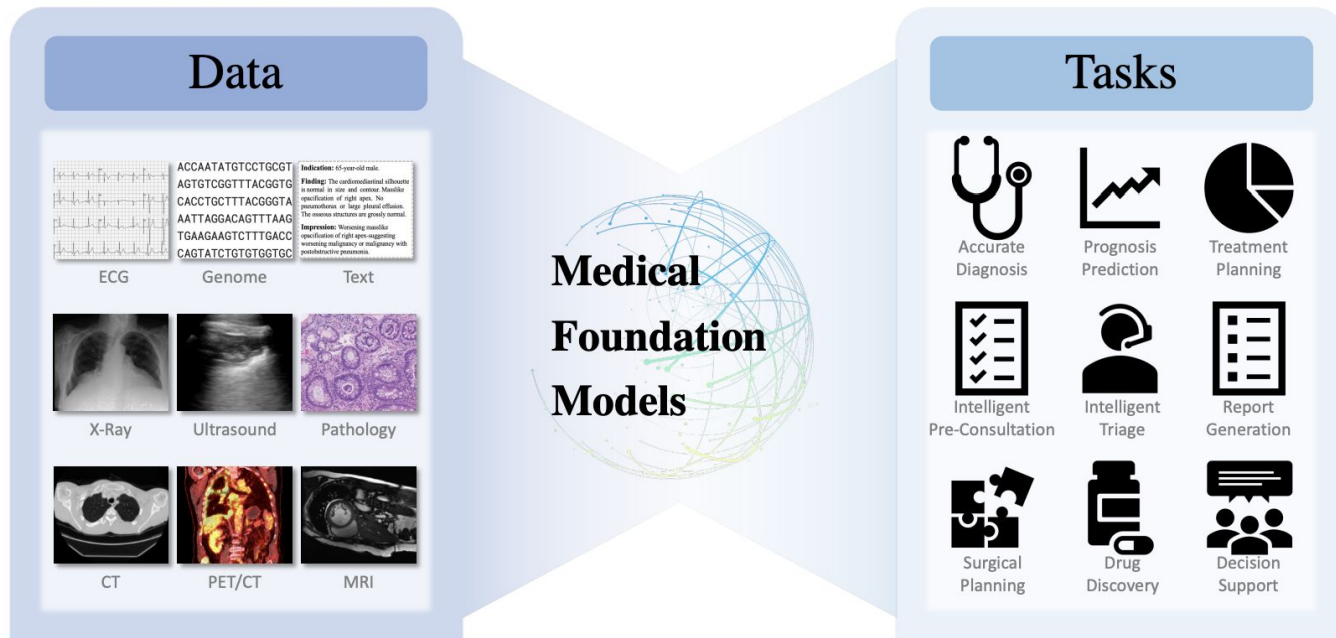
- A new paradigm for building AI systems
 - Train one model on a huge amount of **multimodal** data using **self-supervision** at scale
 - Can adapt to a wide range of downstream tasks



Bommasani, et al., On the Opportunities and Risks of Foundation Models, arXiv 2022.

Medical foundation model

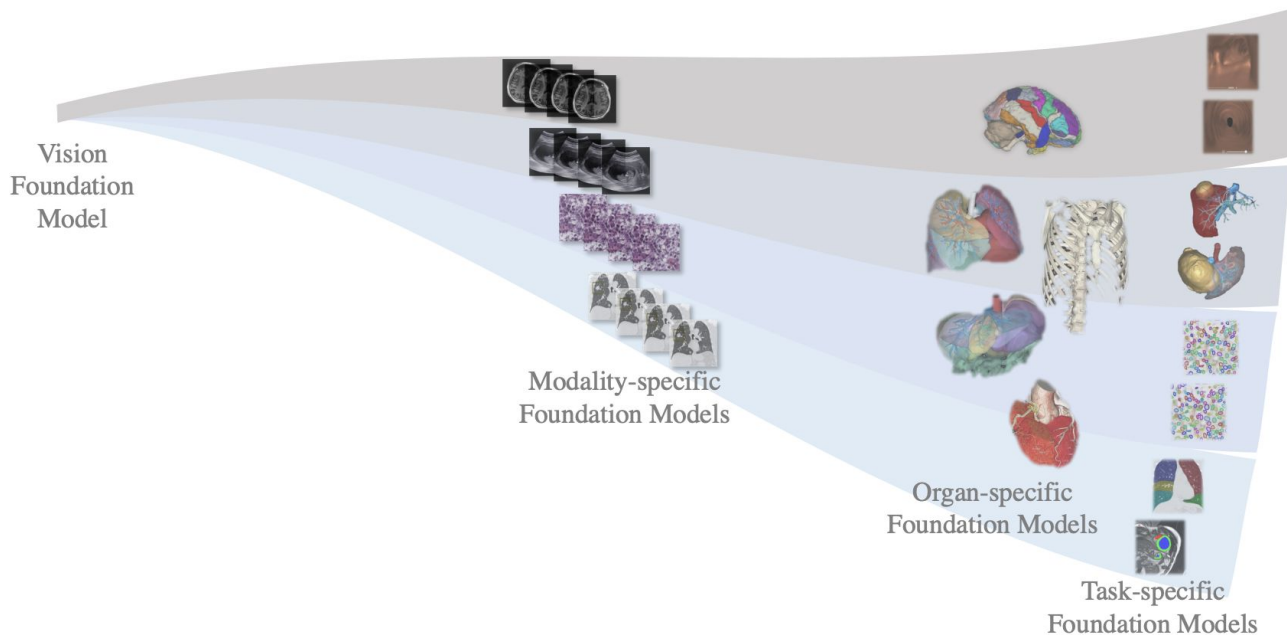
- Models trained on multiple medical data modalities can enable comprehensive clinical solutions



Zhang, et al., *On the Challenges and Perspectives of Foundation Models for Medical Image Analysis*, arXiv 2023.

Medical foundation model

- Modality-specific and organ-specific foundation models



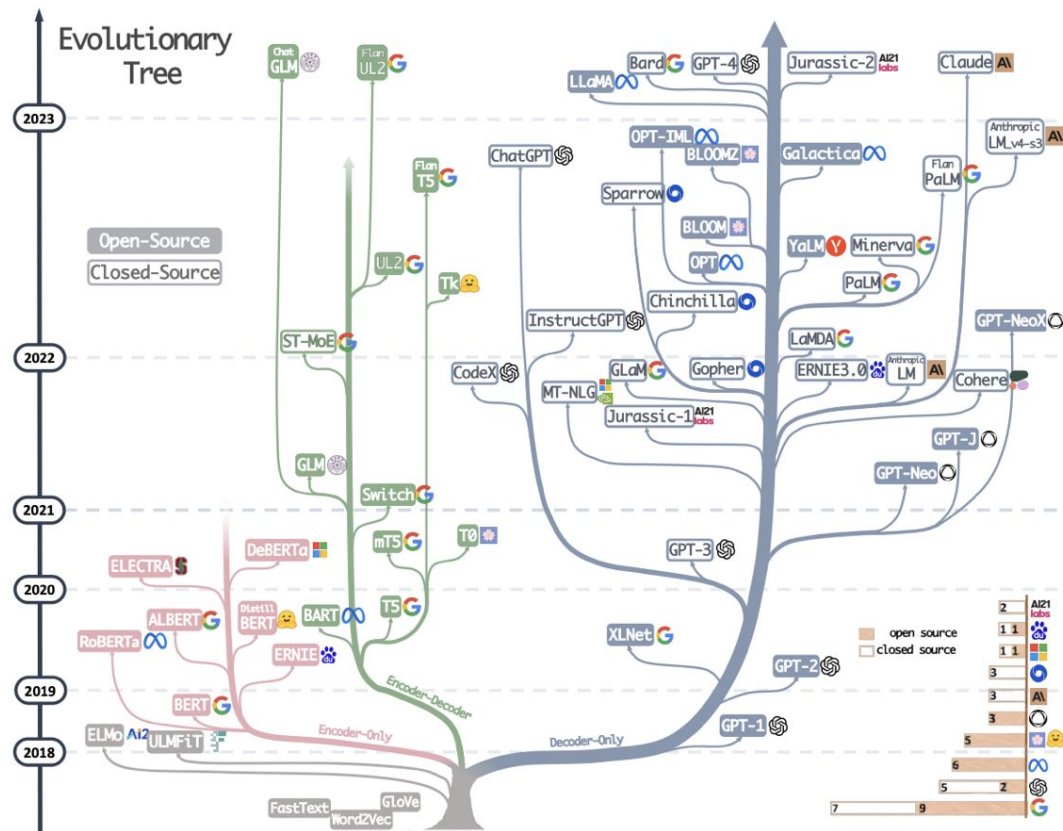
Zhang, et al., *On the Challenges and Perspectives of Foundation Models for Medical Image Analysis*, arXiv 2023.

Today's agenda

- Foundation model
- Language foundation model
- Vision foundation model
- Multimodal foundation model
- Generalist Medical AI (GMAI)

Language foundation model

- Large Language Models (LLMs)
 - Zero-shot generalization ability



Yang, et al., Harnessing the power of LLMs in practice: a survey on ChatGPT and beyond, arXiv 2023.

Language foundation model

- Large Language Models (LLMs)
 - BERT-style language models: encoder-decoder or encoder-only
 - GPT-style language models: decoder-only

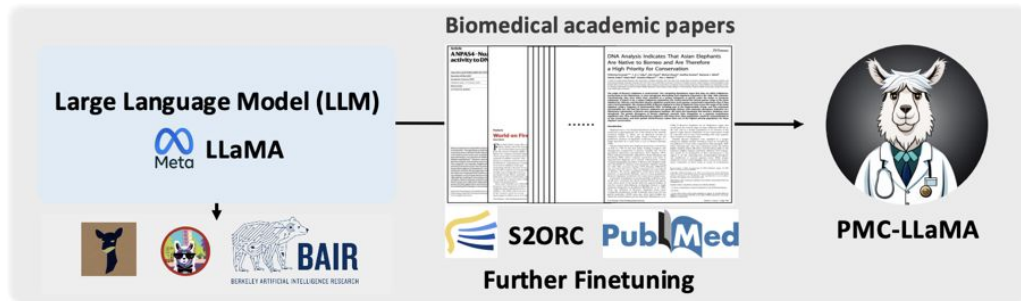
	Characteristic	LLMs
Encoder-Decoder or Encoder-only (BERT-style)	Training: Masked Language Models Model type: Discriminative Pretrain task: Predict masked words	ELMo [80], BERT [28], RoBERTa [65], DistilBERT [90], BioBERT [57], XLM [54], Xlnet [119], ALBERT [55], ELECTRA [24], T5 [84], GLM [123], XLM-E [20], ST-MoE [133], AlexaTM [95]
Decoder-only (GPT-style)	Training: Autoregressive Language Models Model type: Generative Pretrain task: Predict next word	GPT-3 [16], OPT [126], PaLM [22], BLOOM [92], MT-NLG [93], GLaM [32], Gopher [83], chinchilla [41], LaMDA [102], GPT-J [107], LLaMA [103], GPT-4 [76], BloombergGPT [117]

Yang, et al., *Harnessing the power of LLMs in practice: a survey on ChatGPT and beyond*, arXiv 2023.

Medical language foundation model

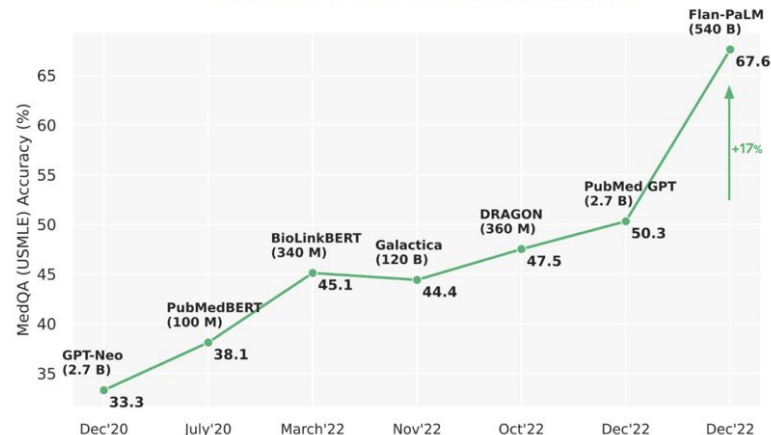
- Medical LLMs

- Train from scratch on medical language
 - BioGPT
 - BioMedLM (2.7-billion)
- Fine-tune from LLMs of natural language
 - PMC-LLaMA (7-billion)
 - MedPaLM (540-billion)



Luo, et al., *BioGPT: generative pre-trained transformer for biomedical text generation and mining*, *Briefings in Bioinformatics* 2022.
Singhal, et al., *Large language models encode clinical knowledge*, *arXiv* 2022.
Wu, et al., *PMC-LLaMA: Further finetuning LLaMA on medical papers*, *arXiv* 2023.

Automated and Human Evaluation



Q: How long does it take for newborn jaundice to go away?

Med-PaLM: Newborn jaundice is when a newborn baby's skin and eyes look yellow. It is caused by a buildup of a substance called bilirubin in the baby's blood. Bilirubin is a byproduct of the normal process of red blood cells breaking down. Newborn jaundice is common and typically harmless, but it can be a sign of a more serious condition. The jaundice typically goes away on its own within a few weeks. In some cases, it may last longer or require treatment. The time it takes for the jaundice to go away can vary depending on the cause and the severity of the jaundice. If the jaundice is severe or lasts longer than a few weeks, the doctor may recommend testing or treatment to determine the cause and prevent complications.

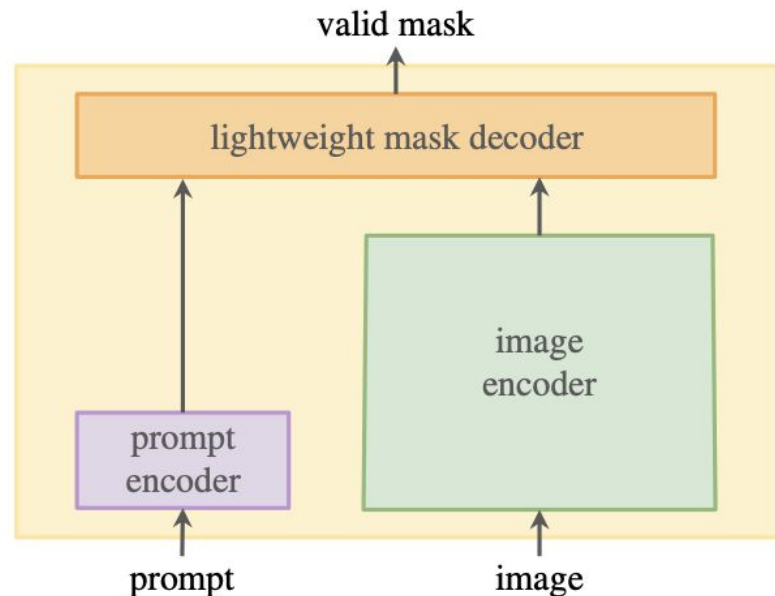
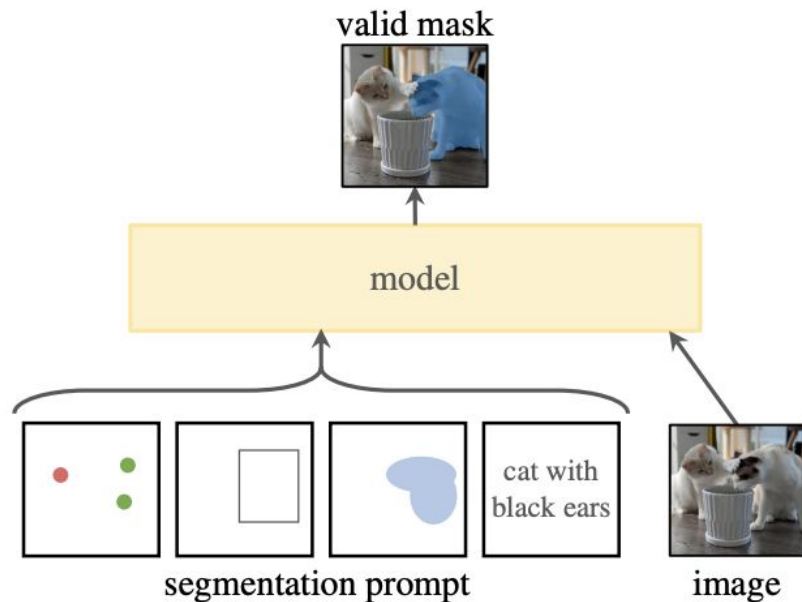
Med-PaLM performs encouragingly on consumer medical question answering

Today's agenda

- Foundation model
- Language foundation model
- Vision foundation model
- Multimodal foundation model
- Generalist Medical AI (GMAI)

Vision foundation model

- Segment Anything Model (SAM) for image segmentation
 - Enable zero-shot transfer to a range of tasks via prompt engineering



Kirillov et al. Segment Anything. arXiv 2023.

Medical vision foundation model

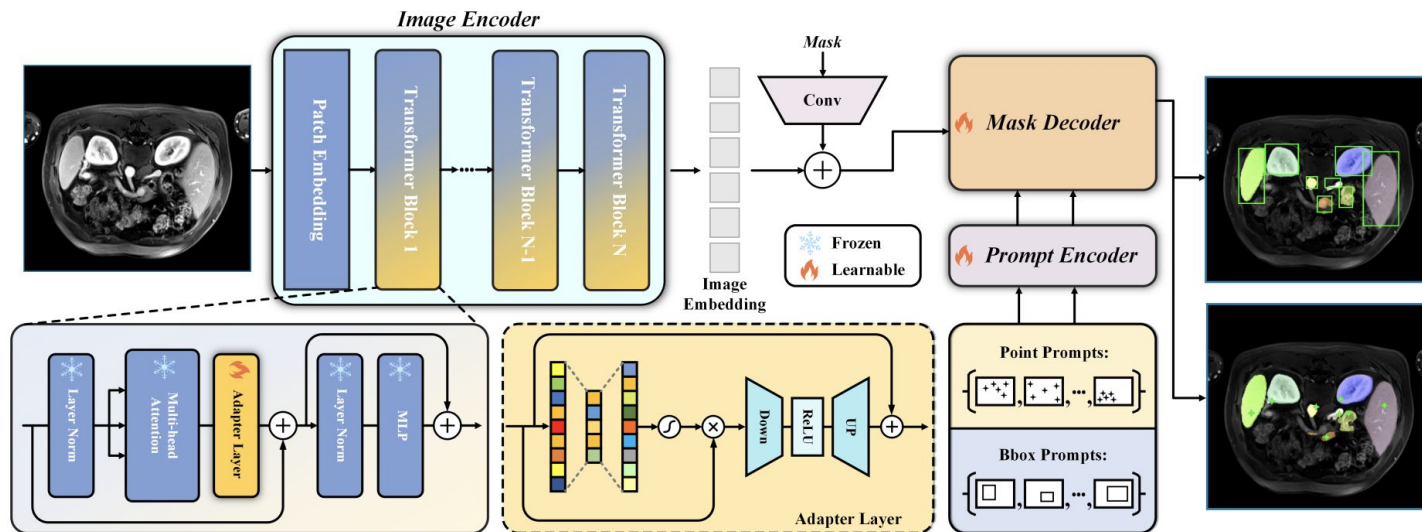
- Fine-tune SAM for medical images
 - Support multiple prompts

Model	Dataset (size)	Encoder (or Adapter)	Prompts mode			Decoder
			<i>Point</i>	<i>Bbox</i>	<i>Mask</i>	
SAM-U [16]	6000 masks	✗	✗	✓	✗	✗
SAMed [17]	3779 masks	✓	✗	✗	✗	✓
AutoSAM [18]	ACDC [19]	✗	✗	✗	✗	✓
MedSAM [20]	~1.1M masks	✗	✗	✓	✗	✓
MSA [21]	5 datasets	✓	✓	✗	✗	✓
SAM-Med2D (Ours)	~19.7M masks	✓	✓	✓	✓	✓

Cheng et al. SAM-Med2D. arXiv 2023.

Medical vision foundation model

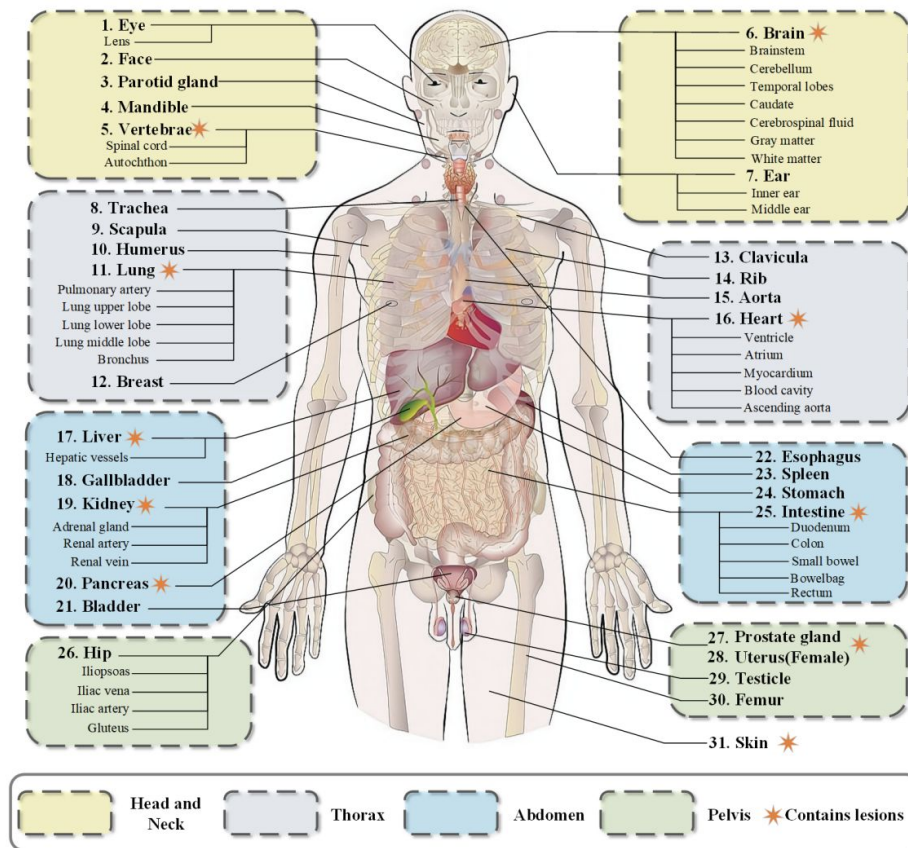
- Fine-tune SAM for medical images
 - Incorporate learnable adapter layers in each Transformer block to acquire domain-specific knowledge in the medical field
 - Fine-tune the prompt encoder using point, Bbox, and mask information, and update mask decoder



Cheng et al. SAM-Med2D. arXiv 2023.

Medical vision foundation model

- Fine-tune SAM for medical images
 - Datasets used cover 31 major organs



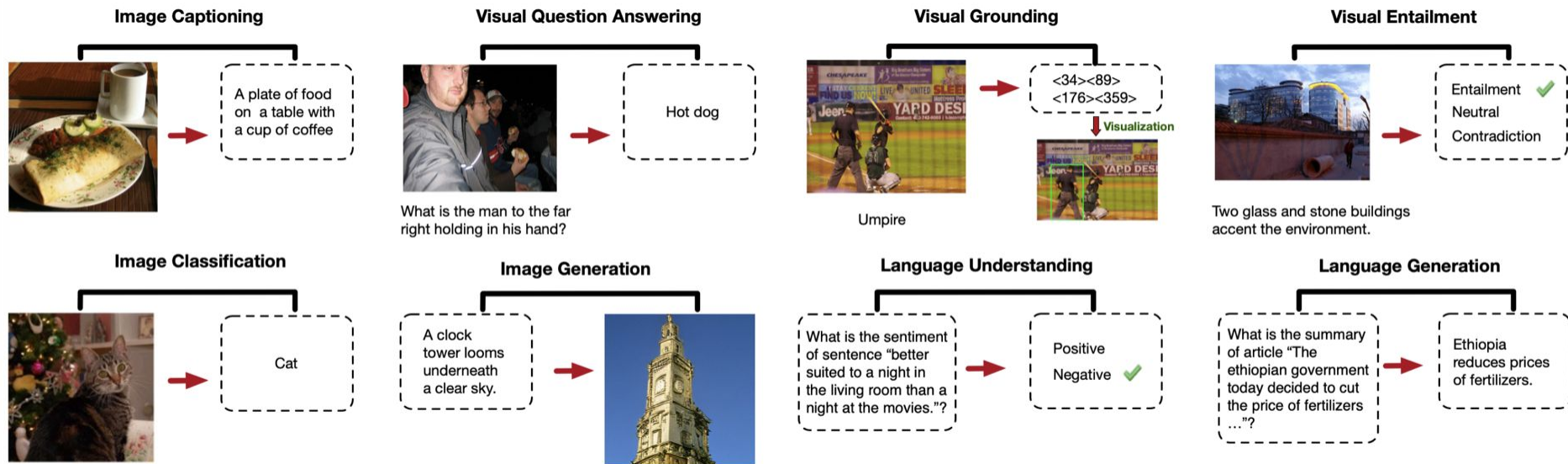
Cheng et al. SAM-Med2D. arXiv 2023.

Today's agenda

- Foundation model
- Language foundation model
- Vision foundation model
- **Multimodal foundation model**
- Generalist Medical AI (GMAI)

Multimodal foundation model

- Multimodal pre-training paradigm for general vision-language understanding
 - Learn transferable cross-modal representations from various self-supervised tasks
 - Generalize to advance various downstream tasks through fine-tuning



Wang, et al., OFA: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework, ICML 2022.

Multimodal foundation model

- Multimodal pre-training paradigm for general vision-language understanding
 - Learn transferable cross-modal representations from various self-supervised tasks
 - Generalize to advance various downstream tasks through fine-tuning

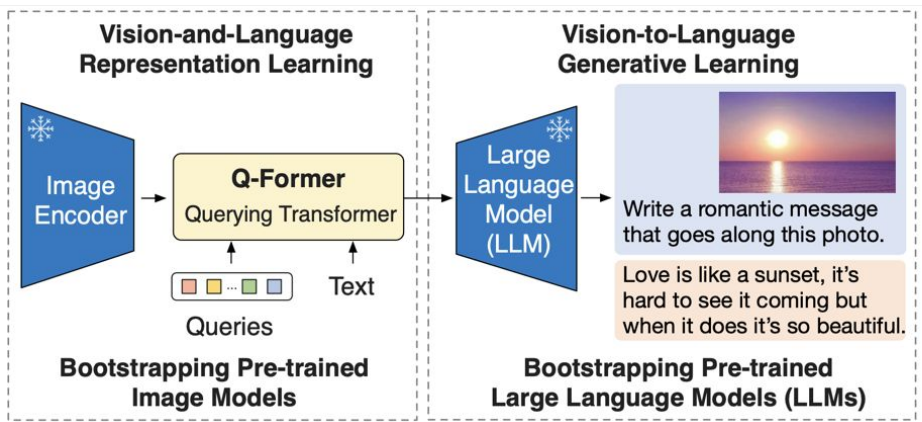
Method	Computer Vision				Natural Language Processing			Image-Text				Video-Text		
	Image Cls.	Video Cls.	Det.	Seg.	Text Cls.	QA	Summarization	Retrieval	QA	Captioning	VG	Retrieval	QA	Captioning
BEiT-3	✓		✓	✓				✓	✓	✓				
EVA	✓	✓	✓	✓										
CLIP	✓							✓				✓		
ALBEF								✓	✓		✓			
BLIP								✓	✓	✓		✓	✓	
VATT	✓	✓										✓		
Florence	✓	✓	✓					✓	✓			✓		
CoCa	✓	✓						✓	✓	✓		✓		
VideoCoCa		✓												
Flamingo		✓						✓	✓	✓			✓	✓
GIT2	✓				✓				✓	✓			✓	✓
FLAVA	✓				✓	✓	✓		✓					
OFA	✓				✓	✓	✓	✓	✓	✓	✓			
OmniVL	✓	✓						✓	✓	✓		✓	✓	✓
mPLUG 2.0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Xu, et al., mPLUG-2: A Modularized multi-modal foundation model across text, image and video, arXiv 2023.

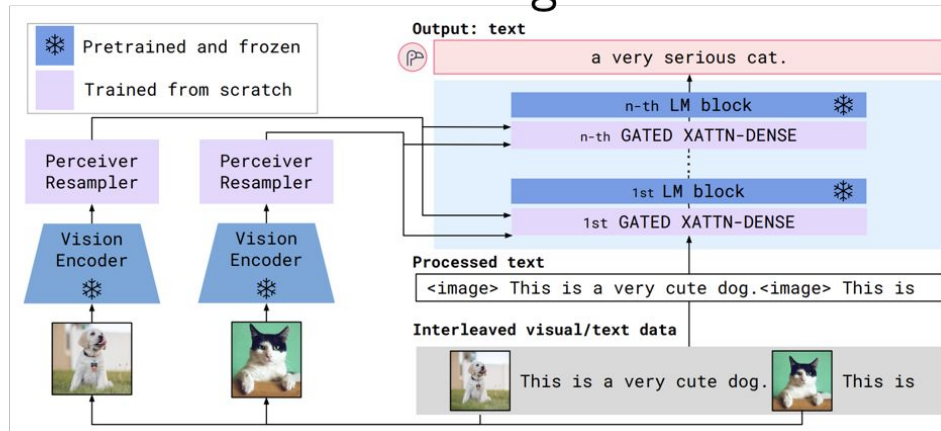
Multimodal foundation model based on LLM

- Enhance vision-language understanding using LLM
 - Learn connection module between visual encoder and LLMs
 - BLIP-2, Flamingo, FROMAGe, MiniGPT-4 ...

BLIP-2



Flamingo




Alayrac, et al., *Flamingo: a Visual Language Model for Few-Shot Learning*, NeurIPS 2022.

Li, et al., *BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models*, arXiv 2023.


Multimodal foundation model based on LLM

- BLIP-2:
 - Instructed zero-shot image-to-text generat




Explain the advantages of this product.

The audi e-tron quattro concept is a plug-in hybrid electric sports car that has a range of 310 miles and can accelerate from 0 to 60 mph in just 4 seconds.




Tell me something about the history of this place.

The great wall of china was built in 221 bc by emperor qin shi Huang to protect the emperor's capital from invaders from the north.



Write down the facts that you know about this flower.

It is a genus of flowering plants in the orchid family, orchidaceae. It is native to eastern asia and south east asia. It is a perennial flowering plant that grows up to a height of 1 m.



Is this photo unusual?

Yes, it's a house that looks like it's upside down.

How could someone get out of the house?

It has a slide on the side of the house.




What are shown in the photo?

A man and a chicken.

What does the man feel and why?

He is scared of the chicken because it is flying at him.



What are the ingredients I need to make this?








Pepperoni, mozzarella cheese, pizza sauce, olive oil, salt, pepper, basil.

What is the first step?

Place the pizza dough on a baking sheet, brush with olive oil, sprinkle with salt, pepper, and basil.


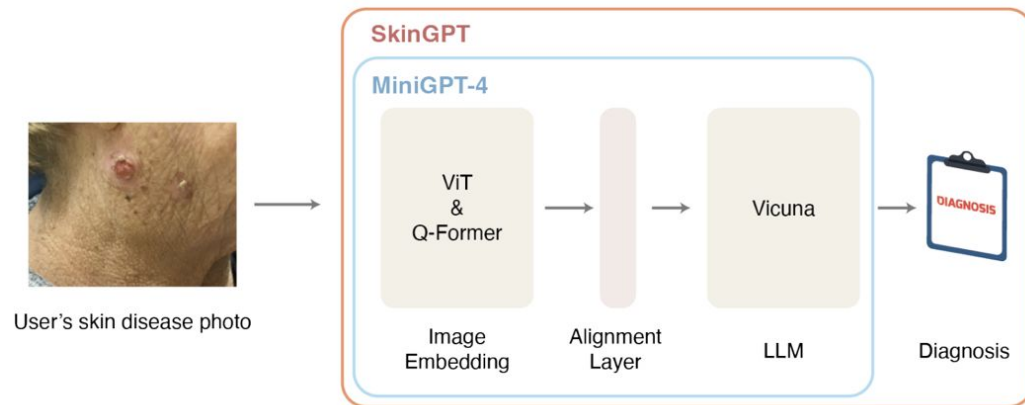
Multimodal foundation model based on LLM

- Flamingo:
 - Multi-image visual dialogue


 <p>P This is a picture of two teddy bears on the moon.</p> <p>What are they doing?</p> <p>P They are having a conversation.</p> <p>What object are they using?</p> <p>P It looks like a computer.</p> <p>Is this surprising?</p> <p>P Yes, it is surprising.</p> <p>Why is this picture surprising to you?</p> <p>P I think it is surprising because teddy bears are not usually found on the moon.</p>	   <p>P What is the common thing about these three images?</p> <p>P They are all flamingos.</p> <p>What is the difference between these three images?</p> <p>P The first one is a cartoon, the second one is a real flamingo, and the third one is a 3D model of a flamingo.</p>	 <p>P This is an apple with a sticker on it.</p> <p>What does the sticker say?</p> <p>P The sticker says "iPod".</p> <p>Where is the photo taken?</p> <p>P It looks like it's taken in a backyard.</p> <p>Do you think it is printed or handwritten?</p> <p>P It looks like it's handwritten.</p> <p>What color is the sticker?</p> <p>P It's white.</p>	 <p>P This is a cityscape. It looks like Chicago.</p> <p>What makes you think this is Chicago?</p> <p>P I think it's Chicago because of the Shedd Aquarium in the background.</p>  <p>What about this one? Which city is this and what famous landmark helped you recognise the city?</p> <p>P This is Tokyo. I think it's Tokyo because of the Tokyo Tower.</p>
---	---	---	--


Medical multimodal foundation model based on LLM

- SkinGPT-4
 - Fine-tune from MiniGPT-4 with skin disease images
 - Enable interactive dialogue with multimodal inputs




What is wrong with my skin?

Plaque, Crust, Erythema 






Please tell us more about this skin disease, what the causes are, and how to treat it.

This is a type of skin cancer that occurs due to excessive sun exposure, age, and other factors. Treatment includes surgical excision, radiation therapy, and topical medications such as chemotherapy creams. Prevention includes sun protection, regular skin exams, and avoidance of risk factors. 

Zhou, et al., SkinGPT-4: An Interactive Dermatology Diagnostic System with Visual Large Language Model, arXiv 2023.

Medical multimodal foundation model based on LLM

- SkinGPT-4

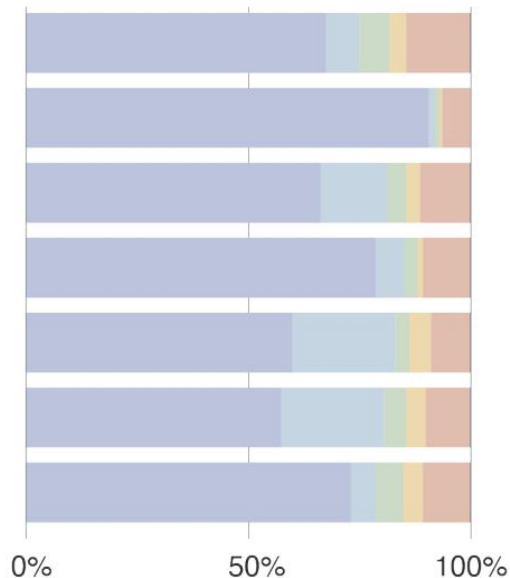
SkinGPT	MiniGPT-4	Dermatology Physician
		
<p>What is wrong with my skin?</p>	<p>What is wrong with my skin?</p>	<p>What is wrong with my skin?</p>
<p>Plaque, Crust, Erythema, Telangiectasia, Atrophy, Xerosis, Skin cancer.</p>	<p>I'm sorry, but I cannot see the image you provided. Please ensure that the image is clear and visible to me.</p>	<p>Did you scratch it with your hand? Are there any other places?</p>
<p>Please tell us more about this skin disease, what the causes are, and how to treat it.</p>	<p>To answer your question, there are several skin conditions that can cause brown spots on the skin. These include:</p> <ol style="list-style-type: none">1. Age spots2. Brown spots caused by acne scars	<p>Yes</p>
<p>This is a type of skin cancer that occurs due to excessive sun exposure, age, and other factors. Treatment includes surgical excision, radiation therapy, and topical medications such as chemotherapy creams. Prevention includes sun protection, regular skin exams, and avoidance of risk factors.</p>	<p>Please tell us more about this skin disease, what the causes are, and how to treat it.</p>	<p>Please tell us more about this skin disease, what the causes are, and how to treat it.</p>
<p>Is it serious?</p>	<p>Melasma is a common skin condition that causes dark, irregularly shaped patches to appear on the skin, typically on the face. It is more common in women, particularly those of childbearing age, and is often referred to as the "mask of pregnancy" because it is more common during pregnancy. However, it can also occur in men and in non-pregnant women.</p>	<p>It is highly probable that you have solar keratosis. It is recommended that you visit a hospital for a proper diagnosis. In most cases, laser treatment can effectively treat the condition. It is important to practice sun protection and avoid scratching the affected area.</p>
<p>Basal cell carcinoma is a type of skin cancer that can be serious if left untreated, as it can spread to other parts of the body and cause tissue damage. However, if caught early and treated properly, it is highly curable. Regular skin exams and sun protection can help prevent the development of basal cell carcinoma.</p>	<p>The exact cause of melasma is not fully understood, but it is thought to be related to hormonal fluctuations and exposure to sunlight.</p>	<p>Is it serious?</p>
		<p>Early treatment of actinic keratoses can lead to complete clearance or elimination of the condition. However, if left untreated, some spots may progress into squamous cell carcinoma. Nevertheless, if the cancer is detected and treated promptly, it is usually not life-threatening.</p>

Medical multimodal foundation model based on LLM

- Clinical evaluation of SkinGPT-4 by certified offline and online dermatologists

Evaluation of SkinGPT-4 (150 cases) Strongly agree Agree Neutral Disagree Strongly disagree

1. SkinGPT-4's diagnosis is correct or relevant.
2. SkinGPT-4's description is informative.
3. SkinGPT-4's suggestions are useful.
4. SkinGPT-4 can help doctors with diagnosis.
5. SkinGPT-4 can help patients to understand their disease better.
6. If SkinGPT-4 can be deployed locally, it protects patients' privacy.
7. Willingness to use SkinGPT-4.

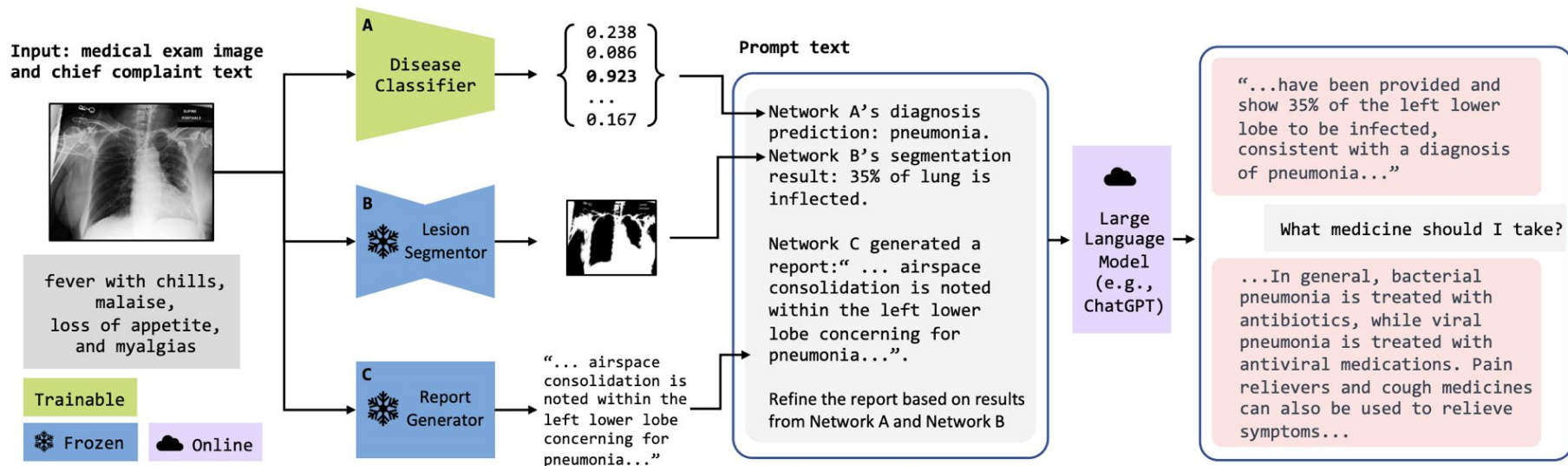


Zhou, et al., SkinGPT-4: An Interactive Dermatology Diagnostic System with Visual Large Language Model, arXiv 2023.

Medical multimodal foundation model based on LLM

- ChatCAD

- Use LLMs to enhance the output of multiple CAD networks by summarizing and reorganizing information presented in natural language text format
- Leverage strengths of LLMs' medical domain knowledge and logical reasoning



Wang, et al., ChatCAD: Interactive computer-aided diagnosis on medical image using large language models, arXiv 2023.

Medical multimodal foundation model based on LLM

- ChatCAD
 - Interactive CAD with LLM



<Network x:...>
Revise the report based on results
from Network x.

<summarized report>

<questions from patient>

What medicine should I take?

For bacterial pneumonia, antibiotics are the most common form of treatment. Some commonly used antibiotics for pneumonia include penicillin, macrolides, fluoroquinolones, and beta-lactams.

If your pneumonia is caused by a virus, antibiotics are not effective. Instead, antiviral medications, such as oseltamivir (Tamiflu) or zanamivir (Relenza), if your pneumonia is caused by the flu.

What's airspace consolidation

Airspace consolidation may refer to the appearance of a particular area of the lung where there is a buildup of fluid, air, or other material that is causing the lung tissue to collapse. This can result in a decreased amount of air in the lung, which can make it difficult for the patient to breathe. Airspace consolidation can be a sign of various lung conditions, such as pneumonia, lung abscess...

I lost smell for more than a week, why?

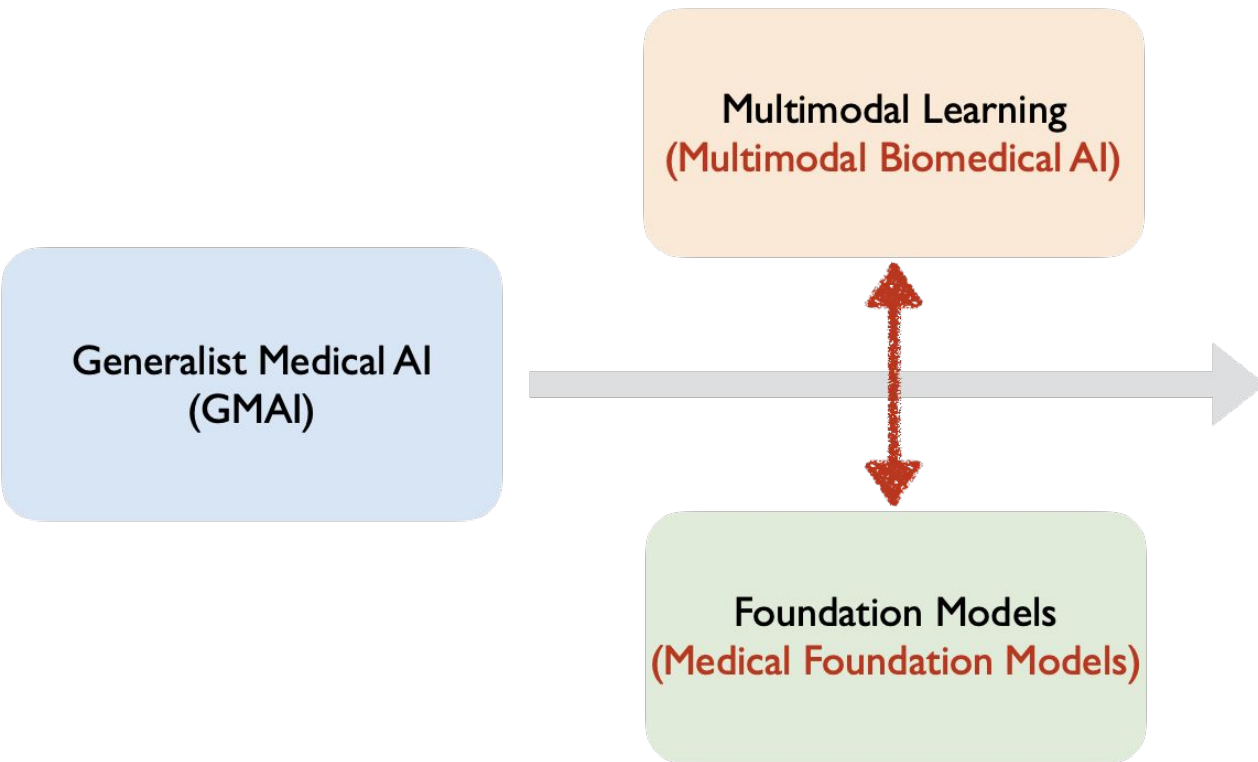
The loss of smell may be due to inflammation and swelling in the nasal passages, which can interfere with the ability to detect odors. Additionally, the infection can also spread to the olfactory nerve, which is responsible for transmitting smells from the nose to the brain.

Wang, et al., ChatCAD: Interactive computer-aided diagnosis on medical image using large language models, arXiv 2023.

Today's agenda

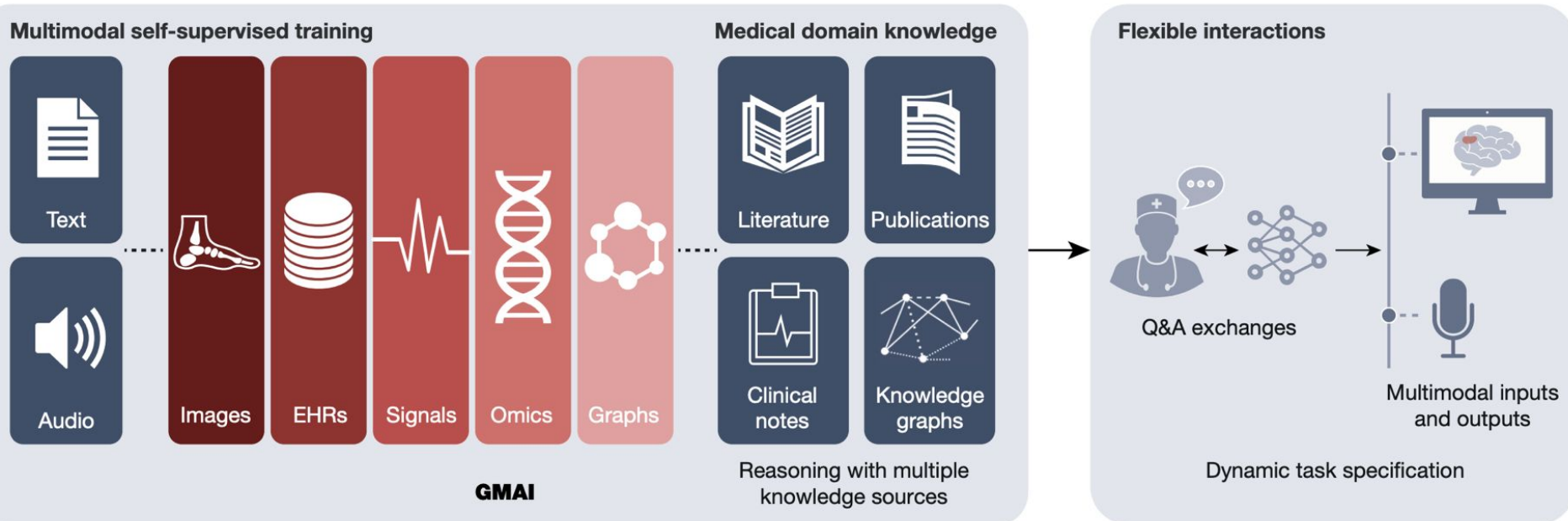
- Foundation model
- Language foundation model
- Vision foundation model
- Multimodal foundation model
- Generalist Medical AI (GMAI)

Generalist Medical AI (GMAI)



Generalist Medical AI (GMAI)

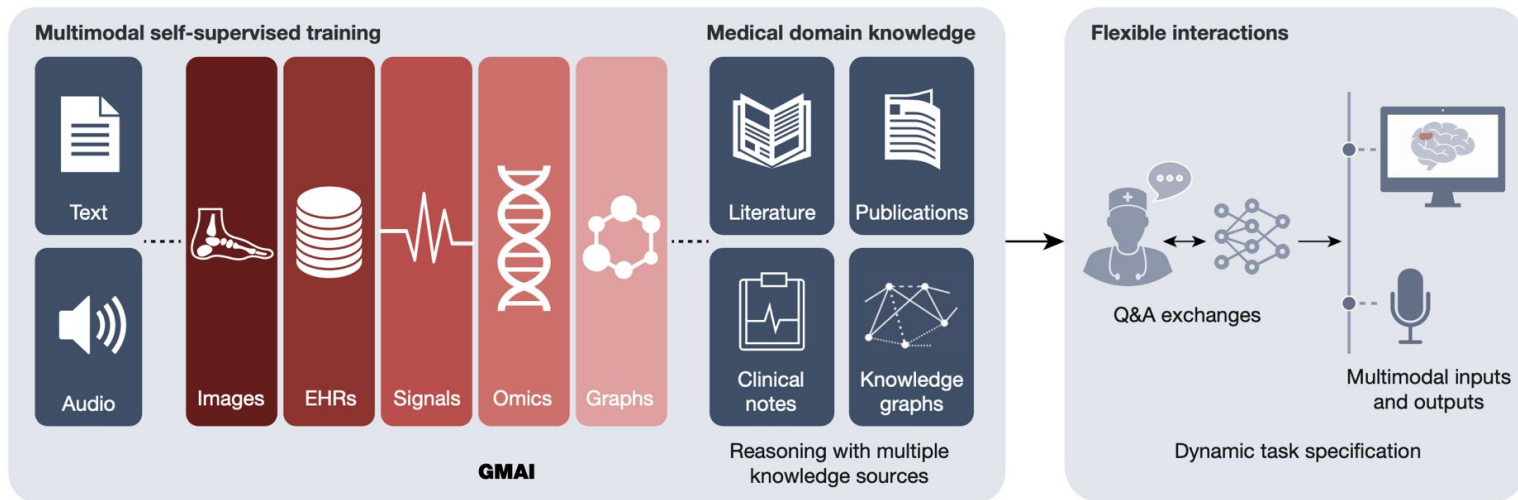
- Generalist paradigm: medical **foundation models** on **multimodal** architectures
 - Promise to solve more diverse and challenging tasks than current medical AI models



Moor, et al., *Foundation models for generalist medical artificial intelligence*, Nature 2023.

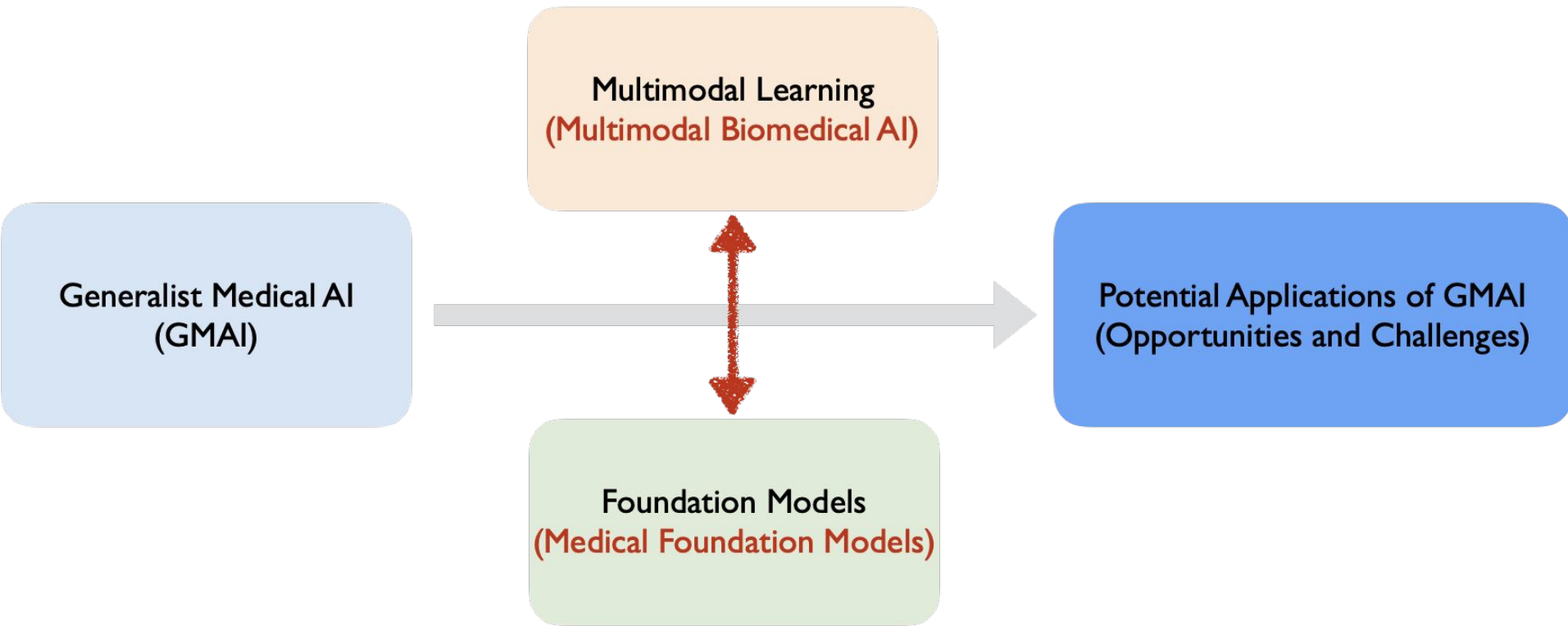
Generalist Medical AI (GMAI)

- Generalist paradigm: medical **foundation models** on **multimodal** architectures
 - Adapt to new tasks with dynamic task specification
 - Flexible combinations and interactions of data modalities
 - Represent medical domain knowledge for reasoning new tasks



Moor, et al., *Foundation models for generalist medical artificial intelligence*, Nature 2023.

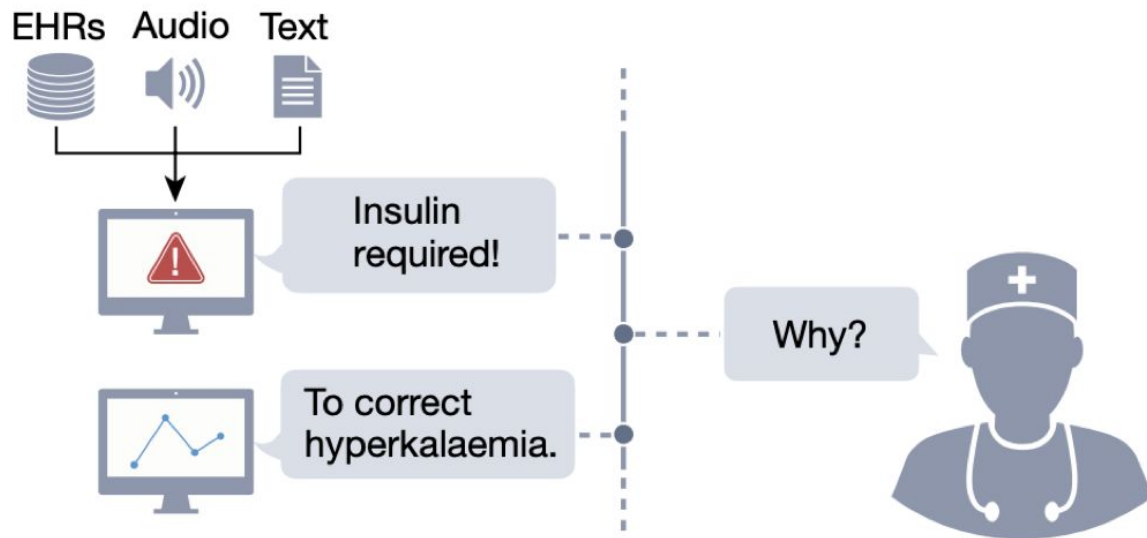
Generalist Medical AI (GMAI)



Potential applications of GMAI

- Generalist paradigm: medical **foundation models** on **multimodal** architectures
 - Bedside decision support

a Bedside decision support

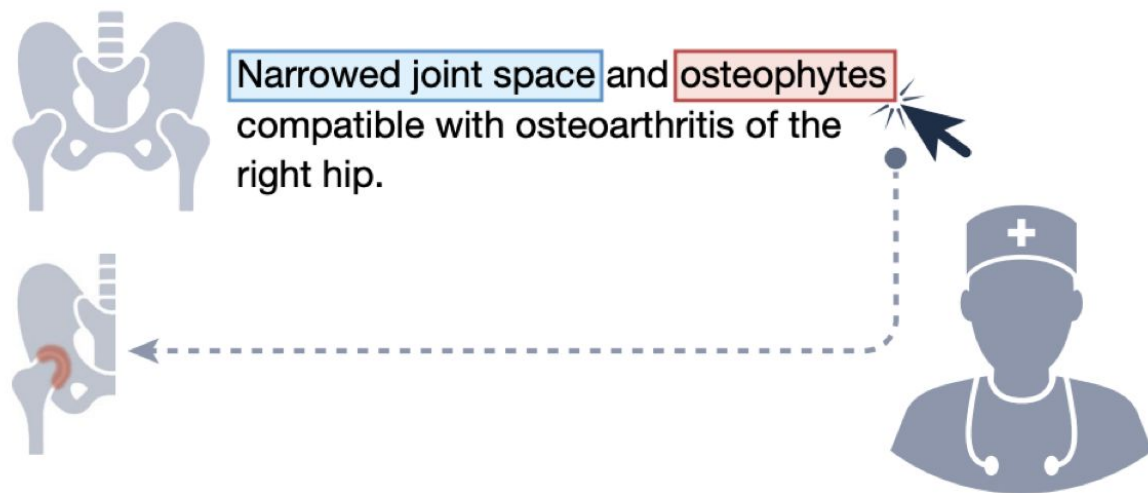


Moor, et al., Foundation models for generalist medical artificial intelligence, Nature 2023.

Potential applications of GMAI

- Generalist paradigm: medical **foundation models** on **multimodal** architectures
 - Grounded radiology reports

b Grounded radiology reports

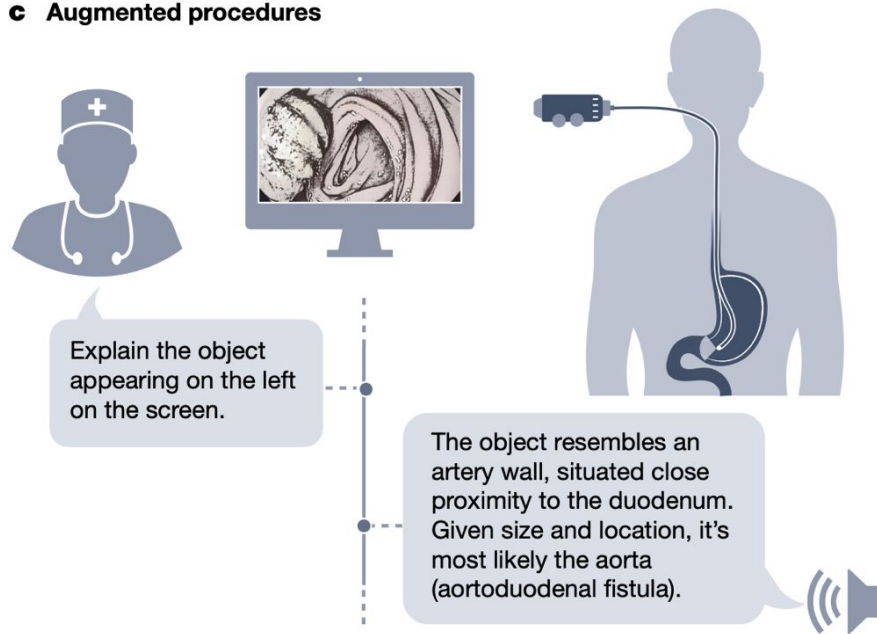


Moor, et al., Foundation models for generalist medical artificial intelligence, Nature 2023.

Potential applications of GMAI

- Generalist paradigm: medical **foundation models** on **multimodal** architectures
 - Augmented procedures

■ Augmented procedures



Moor, et al., *Foundation models for generalist medical artificial intelligence*, Nature 2023.

Potential applications of GMAI

- Generalist paradigm: medical **foundation models** on **multimodal** architectures
 - Bedside decision support
 - Grounded radiology reports
 - Augmented procedures
 - Interactive note-taking
 - Chatbots for patients
 - Text-to-protein generation
 -



Chatbots for patients



Interactive note-taking



Augmented procedures

...



Grounded radiology reports



Text-to-protein generation



Bedside decision support

Moor, et al., Foundation models for generalist medical artificial intelligence, Nature 2023.

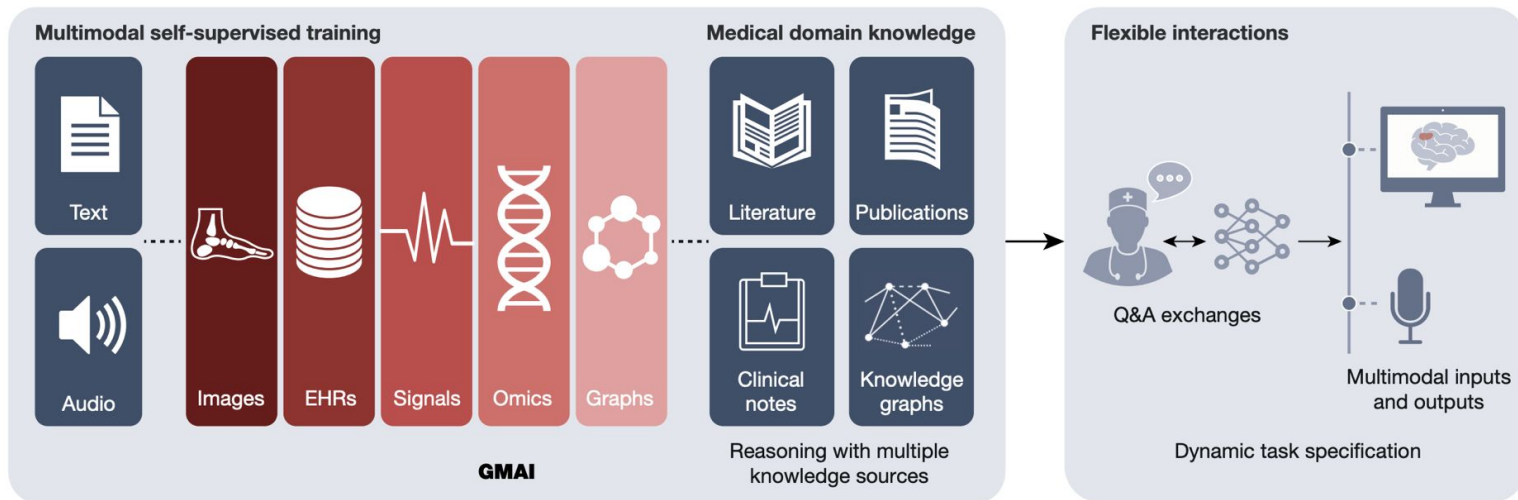
Challenges of GMAI

- Validation
- Verification
- Social biases
- Privacy
- Scale
-

Moor, et al., Foundation models for generalist medical artificial intelligence, Nature 2023.

Generalist Medical AI (GMAI)

- Generalist paradigm: medical **foundation models** on **multimodal** architectures
 - Adapt to new tasks with dynamic task specification
 - Flexible combinations and interactions of data modalities
 - Represent medical domain knowledge for reasoning new tasks



Moor, et al., Foundation models for generalist medical artificial intelligence, Nature 2023.

All lectures are done!

Any feedback and suggestions?

Date	Lecture #	Topic	Papers	Instructor / Presenter
Tue 8/29	1	Introduction and course overview		Liyue Shen
Thu 8/31	2	Biomedical imaging with deep learning [Fundamental]		Liyue Shen
Tue 9/5	3	Implicit neural representation learning [Advanced]		Liyue Shen
Thu 9/7	4	Generative diffusion models [Advanced]		Liyue Shen
Tue 9/12	5	Medical image analysis [Fundamental]		Liyue Shen
Thu 9/14	6	Multimodal foundation models [Advanced]		Liyue Shen
Mon 9/18		Drop/add deadline for full term classes		
Tue 9/19	7	Implicit neural representation learning		
Thu 9/21	8	Implicit neural representation learning		
Tue 9/26	9	Implicit neural representation learning		
Thu 9/28	10	Implicit neural representation learning		
Tue 10/3	11	Generative diffusion models		
Thu 10/5	12	Generative diffusion models		
Tue 10/10	13	Generative diffusion models		
Thu 10/12	14	Generative diffusion models		
Tue 10/17		No class (fall study break)		
Thu 10/19	15	Self-supervised learning		
Tue 10/24	16	Self-supervised learning		
Thu 10/26	17	Multimodal learning		
Tue 10/31	18	Multimodal learning		
Thu 11/2	19	Transformer and LLM		
Tue 11/7	20	Transformer and LLM		