



中国科学院新疆理化技术研究所  
Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences

# The Power of AI for General Applications in Medicine

Mieradilijiang Maimaiti  
(2025.04.03, Hangzhou)

# Outline

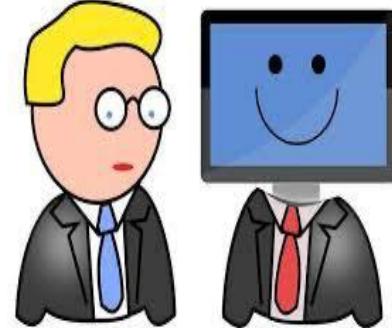
- Introduction of LLM
- Medical LLM
- AI for protein structure prediction
- AI for new drug discovery & development
- AI in medical diagnostics & healthcare
- LLM agent hospital

# Introduction of LLM

# Large-scale Language Model

## Input experience

Teach computers knowledge via rules for symbolic processing.



## Mechanisms of the human brain



Decipher brain mechanisms to replicate human intelligence.

## Learn from data

Through data-driven approaches, ML methods emulate human intelligence.



[Learn From Data](#)

**LLM = Learn from data + Inspired by human brain + Input experience**

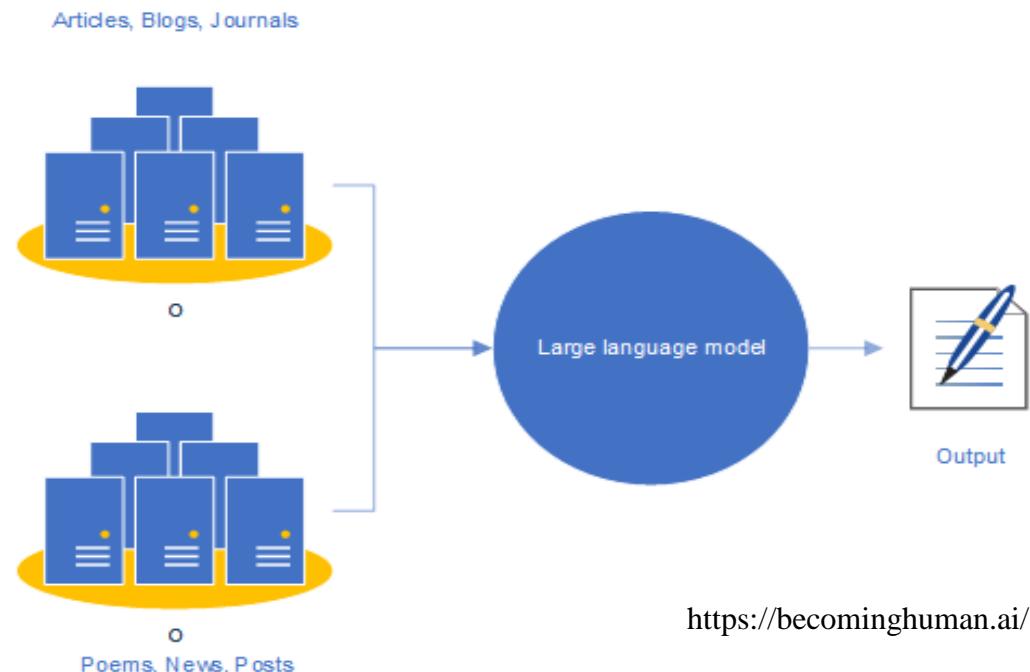
# What Are Large Language Models?

Large **language models** (LLMs) are Transformer-based models trained on extensive unlabeled text via self-supervised or semi-supervised learning, designed to produce natural, context-relevant text in various styles and formats.

**LLM**

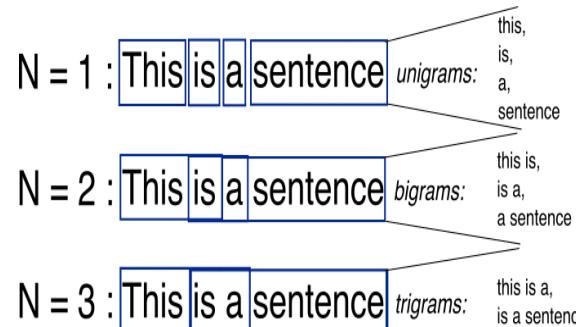


**Large-scale Pre-trained  
Language Model (PLM)**

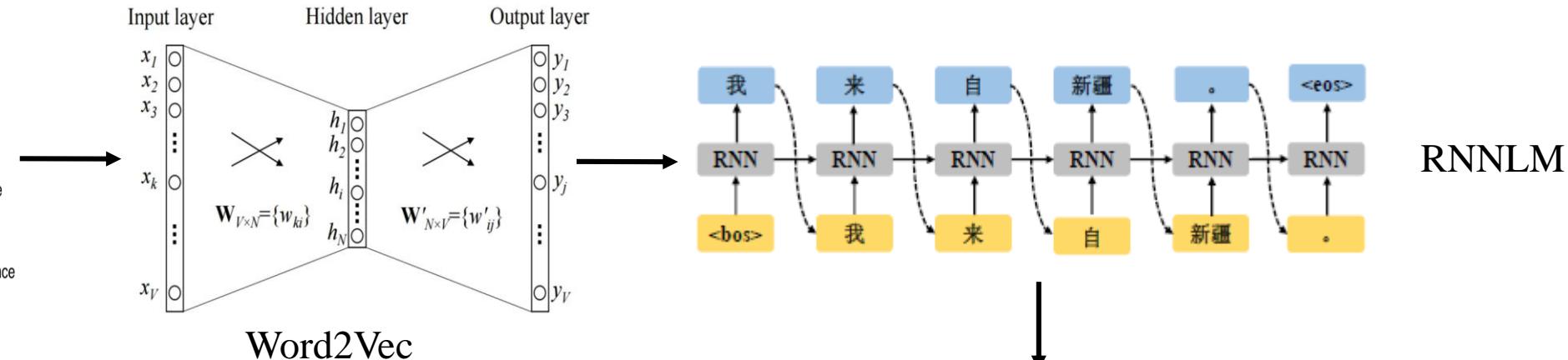


<https://becominghuman.ai/>

# Language Models



N-gram LM

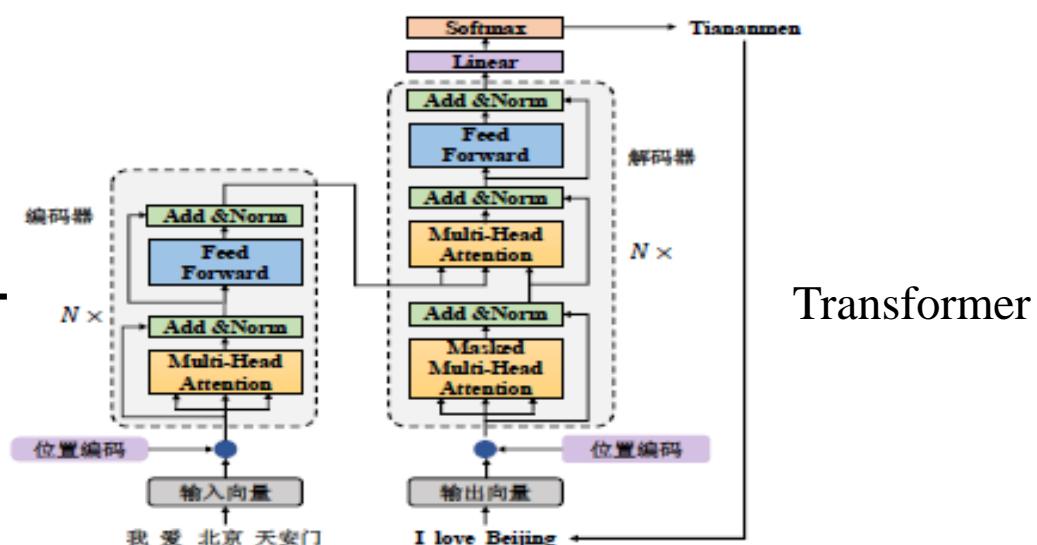


RNNLM

LLM

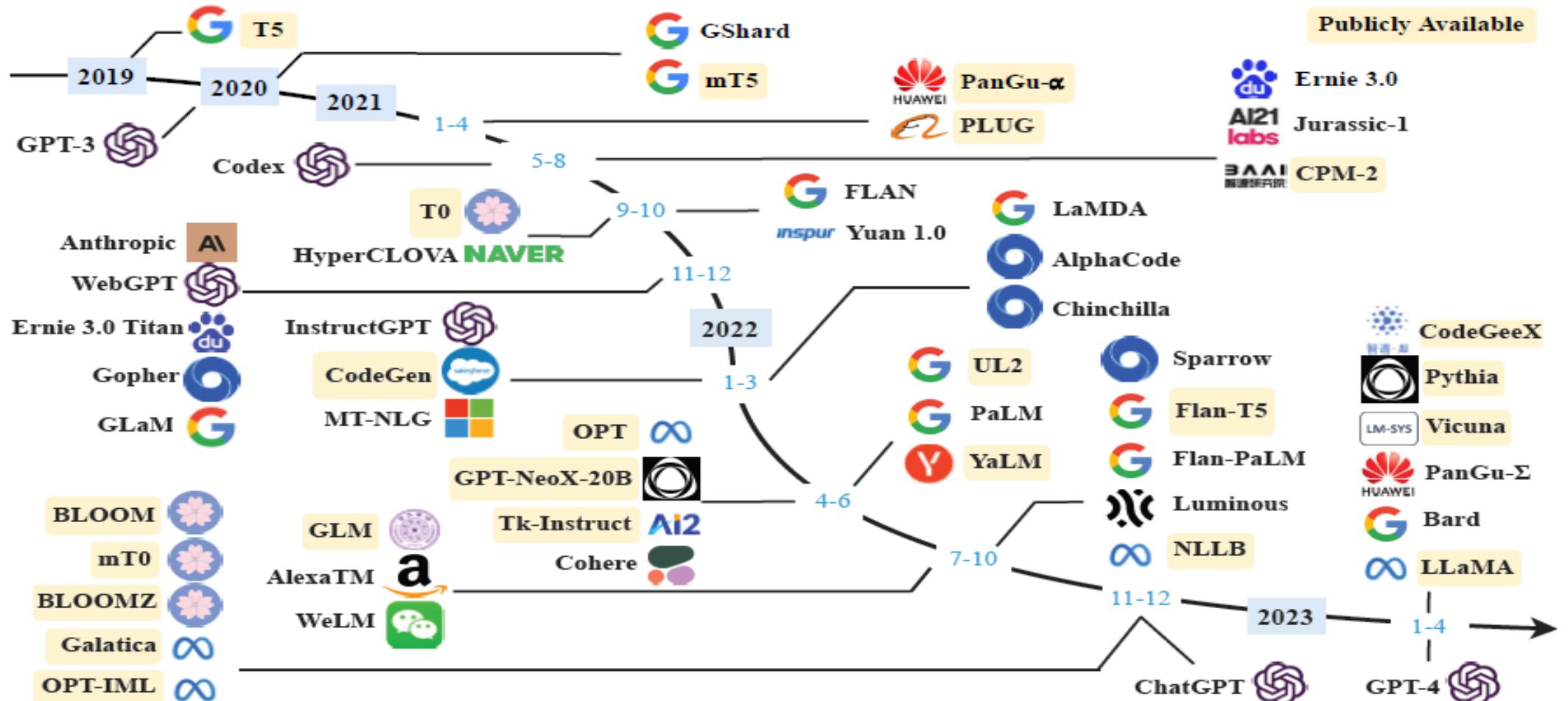


PLM



Transformer

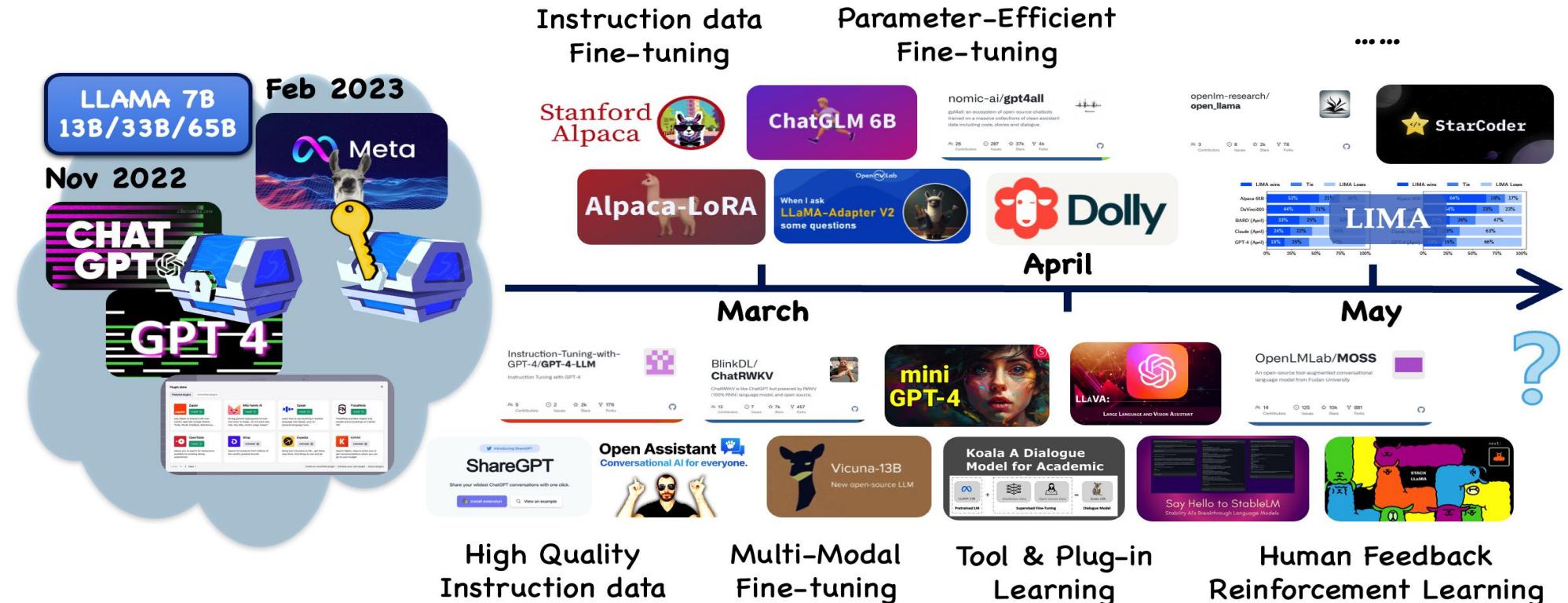
# The timeline of existing LLM



(Xin Zhao et al., 2023)

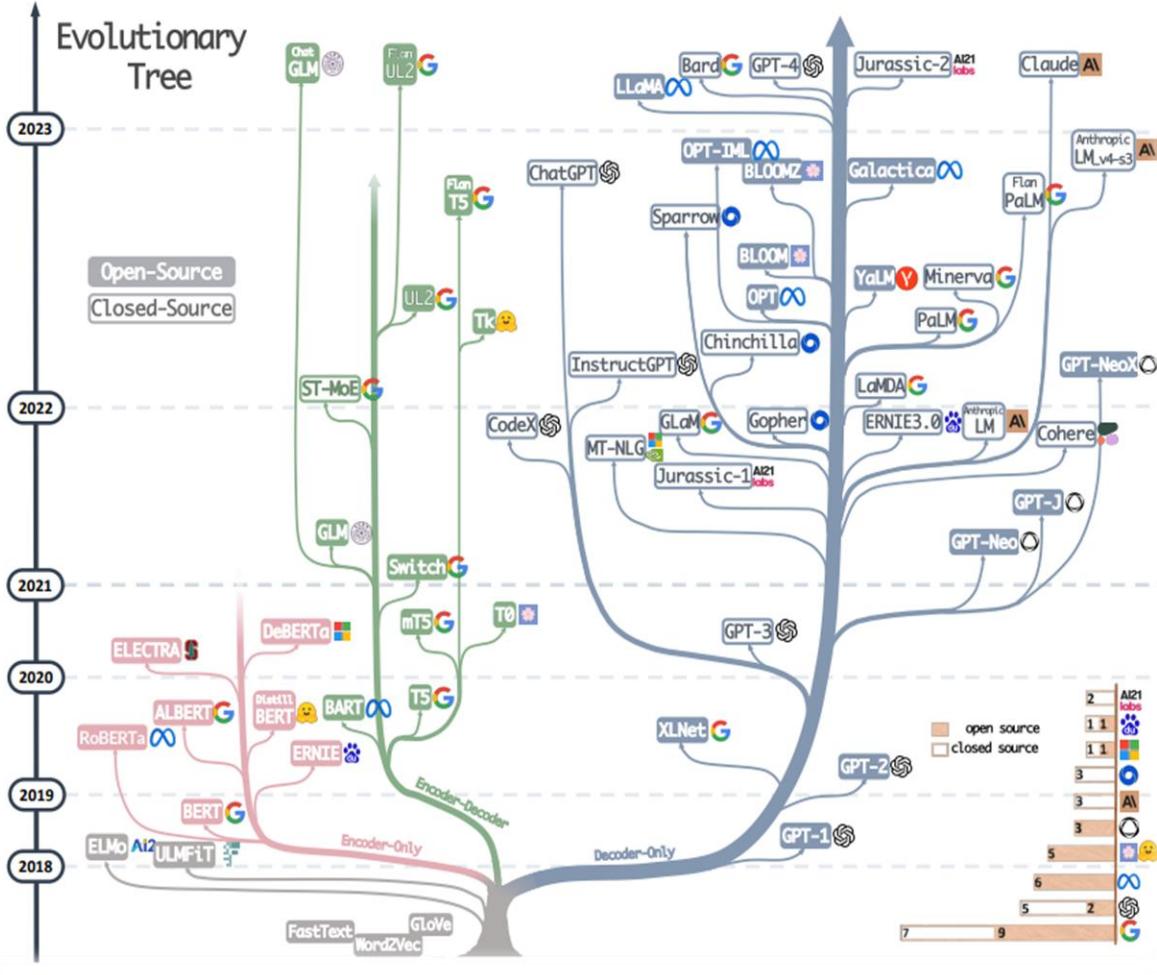
# Basic Fact - foundational models is rapidly evolving

A bunch of instruction fine-tuned models keep emerging.



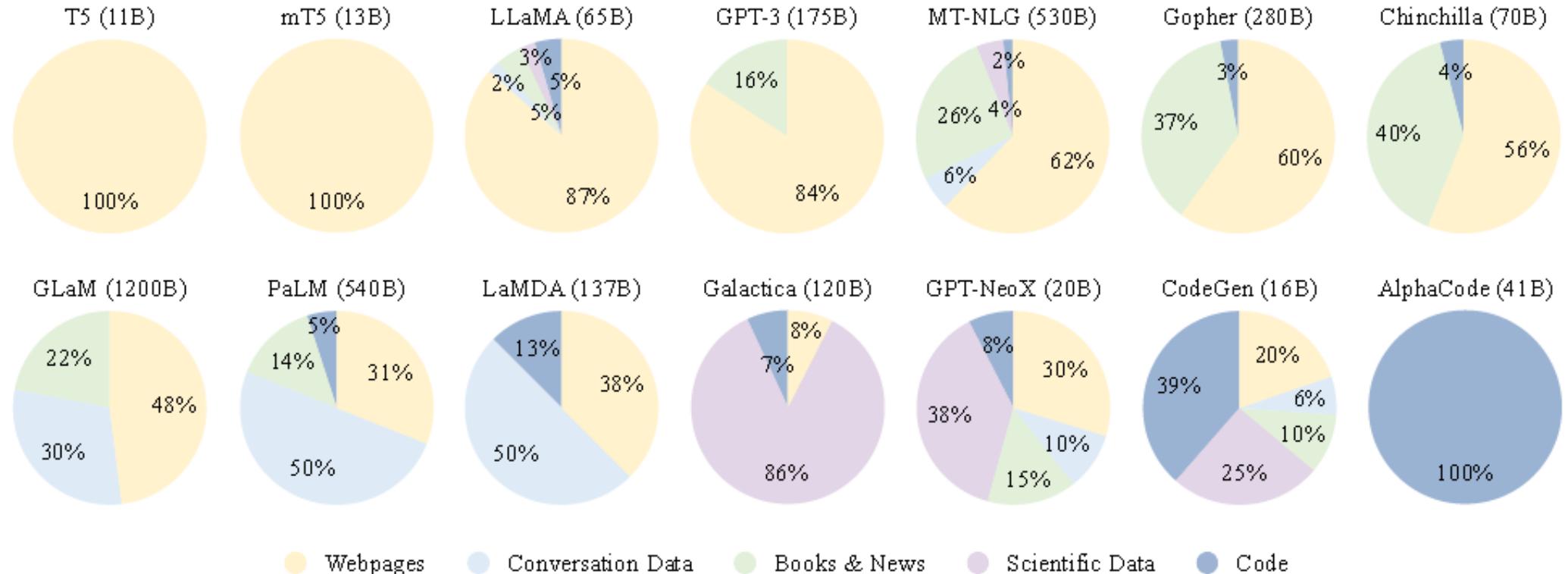
(Ding et al., CCL2023)

# The evolutionary of LLM



- **GPT** "Improving Language Understanding by Generative Pre-Training". 2018
- **GPT-2** "Language Models are Unsupervised Multitask Learners". 2018.
- **GPT-3** "Language Models are Few-Shot Learners". NeurIPS 2020.
- **OPT** "OPT: Open Pre-trained Transformer Language Models". 2022.
- **PaLM** "PaLM: Scaling Language Modeling with Pathways". Aakanksha Chowdhery et al arXiv 2022.
- **BLOOM** "BLOOM: A 176B-Parameter Open-Access Multilingual Language Model". 2022.
- **MT-NLG** "Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model". 2021.
- **GLaM** "GLaM: Efficient Scaling of Language Models with Mixture-of-Experts". ICML 2022.
- **Gopher** "Scaling Language Models: Methods, Analysis & Insights from Training Gopher", 2021.
- **chinchilla** "Training Compute-Optimal Large Language Models". 2022.
- **LaMDA** "LaMDA: Language Models for Dialog Applications". 2021.
- **LLaMA** "LLaMA: Open and Efficient Foundation Language Models". 2023.
- **GPT-4** "GPT-4 Technical Report". 2023.
- **BloombergGPT** BloombergGPT: A Large Language Model for Finance, 2023,
- **GPT-NeoX-20B**: "GPT-NeoX-20B: An Open-Source Autoregressive Language Model". 2022.

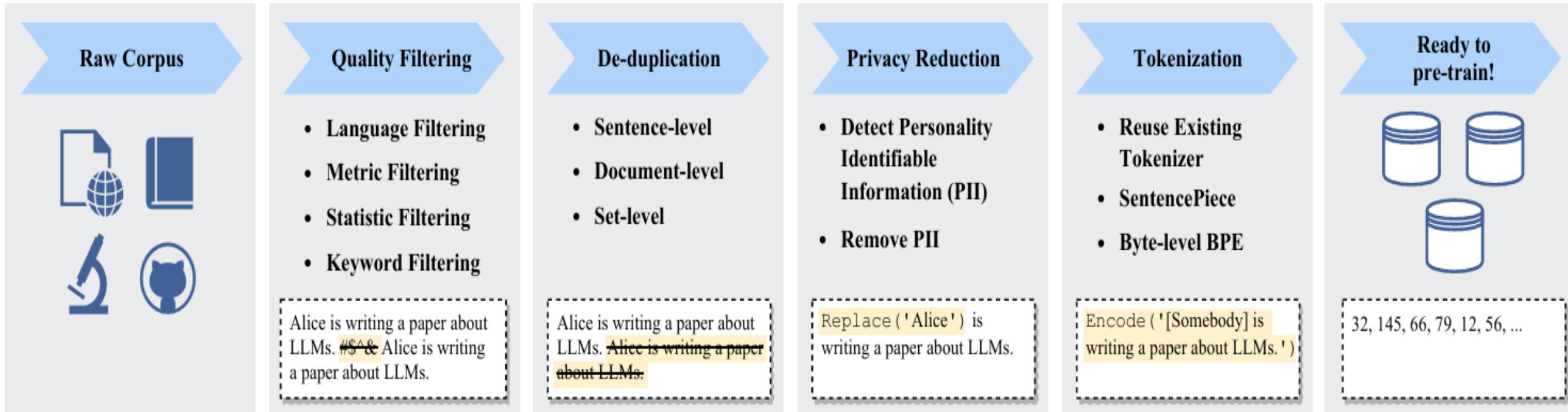
# Dataset Distribution of GPT (Generative Pre-Trained Transformer)



The LLM training (ChatGPT) used **320TB** of text data, approximately **300 billion** words/characters.  
 $320\text{TB} / 5\text{TB} = 64$  people.

(Xin Zhao et al., 2023)

# Dataset Preprocessing



(Xin Zhao et al., 2023)

# The Main Architecture of GPT --- RLHF

Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

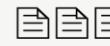
Explain reinforcement learning to a 6 year old.



We give treats and punishments to teach...



SFT



A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.

Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

Explain reinforcement learning to a 6 year old.

- A In reinforcement learning, the agent is...
- B Explain rewards...
- C In machine learning...
- D We give treats and punishments to teach...



D > C > A > B

RM

D > C > A > B

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

Step 3

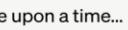
Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.

Write a story about otters.



PPO



Once upon a time...



$r_k$

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

2月28日晚20:00

## ChatGPT类算法的壁垒、开源和算力私有化

### 分享内容

- 从数据、算法、算力、环境等方面讨论国内企业研发ChatGPT的壁垒。
- 探究ChatGPT开源的意义、可能性、所需要的生态环境、难点以及如何克服这些难点。
- 围绕算力私有化讨论ChatGPT带给芯片、算法、投资人、云公司的机会是哪些。

### 主讲嘉宾

- 北京航空航天大学 自动化科学与电气工程学院教授 秦曾昌教授
- 千芯科技董事长 前华为系自然语言处理(NLP)企业首席科学家 陈巍博士
- 阿里巴巴达摩院 语言技术实验室自然语言处理(NLP)高级算法工程师 木拉丁博士

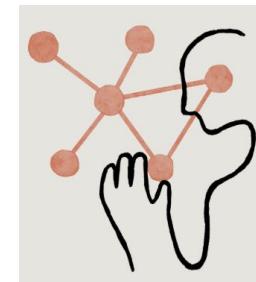
2023年2月28日 晚20:00-21:30 扫码报名



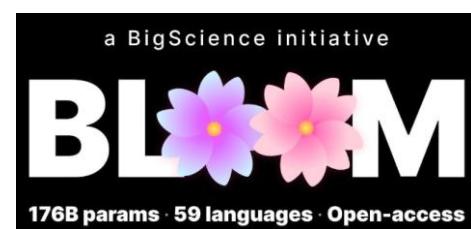
创业芯 ChipStar 工作室 星

创业芯2023

# The existing open source LLMs



KING OF OPEN-SOURCE LLMs



InternLM HOT

# LLM Everywhere --- Conversation, QA, … , MT, Coding, Reading Paper

**自主智能实验室**

西湖大学工学院自主智能实验室的研究工作包含哪些方向？简单总结一下~

西湖大学工学院自主智能实验室的核心研究方向可概括为以下五大方向：

1. 智能机器人系统
  - 医疗/救援机器人开发
  - 自主导航与精准控制（如无人机、自动驾驶）
  - 群体机器人协同技术
2. 智能感知与决策
  - 多模态环境感知（视觉/触觉/听觉融合）
  - 强化学习与自适应决策算法
  - 小样本学习与迁移学习
3. 人机交互与边缘智能
  - 自然语言交互系统
  - 脑机接口技术
  - 轻量化AI与边缘计算
4. 智能医疗应用
  - 手术辅助机器人
  - 医疗影像AI分析
  - 生物医学大数据处理
5. 可信AI与安全
  - 对抗攻击防御
  - 可解释AI技术
  - 系统安全验证

特色：注重“基础研究-技术创新-产业应用”全链条发展，在医疗机器人、自动驾驶、工业检测等领域有突出应用成果。实验室配备先进机器人实验平台和计算资源，与国内外顶尖机构保持密切合作。

自主智能大方向最有挑战性的open problem 是哪些？简单总结一下~ 不用输出思考过程~

自主智能领域中最具挑战性的开放问题（Open Problems）可以总结如下：

1. **通用人工智能（AGI）的实现**：如何创建能够执行任何智力任务的系统，而不仅仅是特定领域的任务。
2. **可解释性和透明度**：提高AI决策过程的透明度，特别是在关键领域如医疗和法律中，使得非专家也能理解和信任这些决策。
3. **安全性与鲁棒性**：确保AI系统在面对对抗攻击、异常数据或未知环境时依然能稳定运行，不会产生意外行为。
4. **多模态学习与融合**：开发能够有效整合并理解来自不同类型传感器（例如视觉、听觉、触觉等）的数据的方法，以增强系统的感知能力。
5. **长期规划与因果推理**：使机器能够在复杂环境中做出长远规划，并基于因果关系而非仅仅相关性来作出决策。
6. **人机协作与伦理决策**：在保证效率的同时解决人类与AI协作中的道德问题，比如自动驾驶汽车遇到不可避免事故时的选择。
7. **数据效率和小样本学习**：减少对大量标注数据的依赖，让模型能在少量示例的情况下学习新技能或适应新情况。
8. **多智能体系统的协调**：研究如何高效地管理和协调多个智能体共同工作，解决合作和竞争中的复杂动态。
9. **能源效率与计算资源优化**：设计更加节能高效的算法和硬件，以便于在资源受限的设备上部署复杂的AI模型。
10. **自适应能力与持续学习**：构建能够在没有人为干预的情况下自我调整并从新环境中学习的系统。

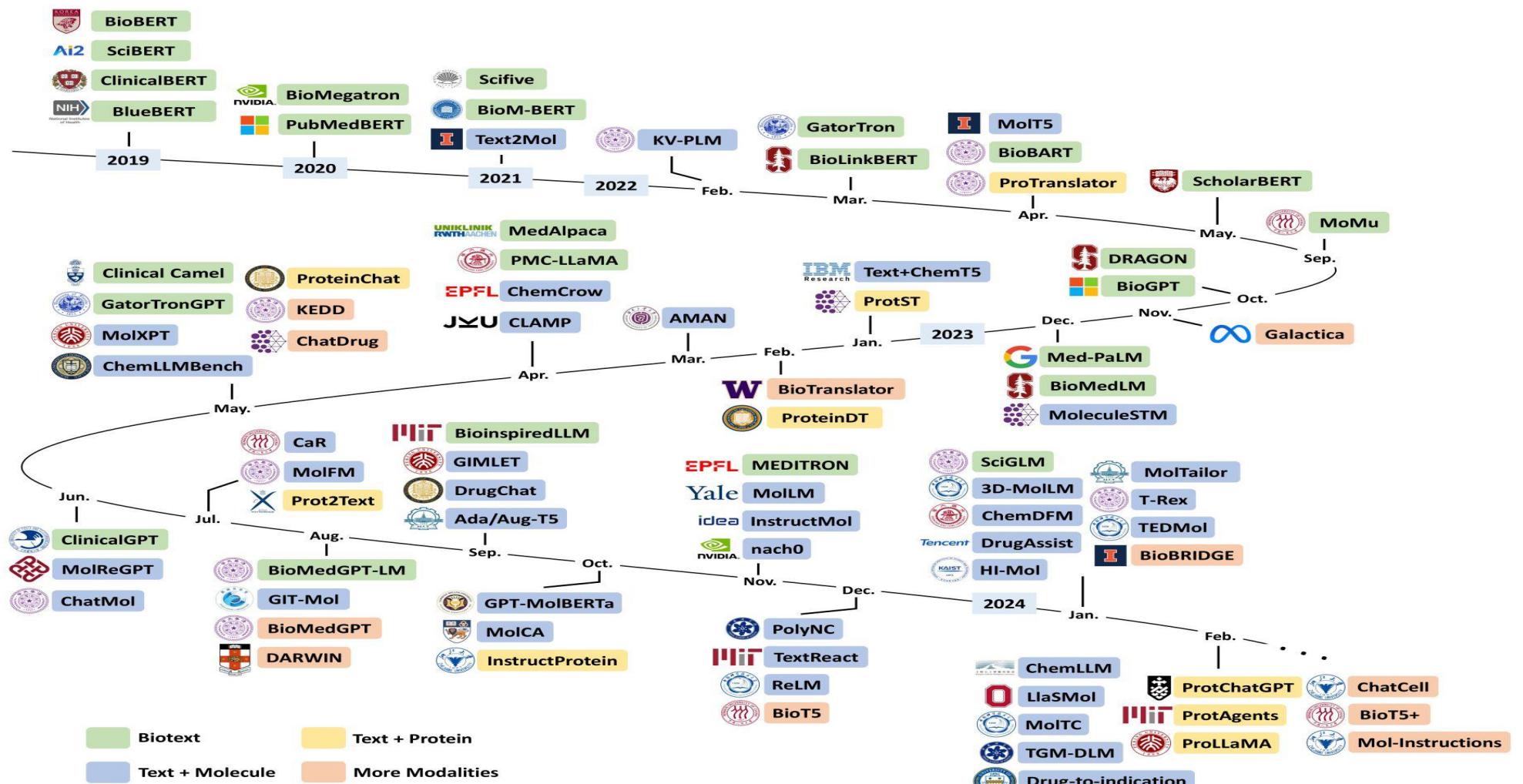
这些问题代表了当前研究者们正在努力克服的一些主要障碍，也是推动自主智能技术进一步发展的关键所在。

# Outline

- Introduction of LLM
- Medical LLM
- AI for protein structure prediction
- AI for new drug discovery & development
- AI in medical diagnostics & healthcare
- LLM agent hospital

# Medical LLM

# Biomedical LLMs



(Lijun Wu et al., 2024)

# Typical Medical LLMs



IBM Watson  
Health



Google DeepMind  
Health



Microsoft Healthcare  
NExt



OpenAI GPT-4o

CHEERS  
湛庐

# 超越想象的 GPT医疗

[美] 彼得·李 Peter Lee  
[美] 凯丽·戈德伯格 Carey Goldberg 著  
[美] 伊萨克·科恩 Isaac Kohane  
芦义 译



Tencent



Alibaba Group



Baidu

解析GPT落地实践  
揭示“人机关系”的重要范式

微软研究院负责人 彼得·李 全球首部力作  
比尔·盖茨 重磅推荐  
OpenAI 创始人 萨姆·奥尔特曼 千字长序

集体盛赞

洪小文 微软全球 资深副总裁	韦青 微软（中国）公司 首席技术官	王小川 百度 创始人兼CEO	段永朝 锤子科技 创始合伙人	周健工 未来研究 创办人
----------------------	-------------------------	----------------------	----------------------	--------------------

浙江科学技术出版社

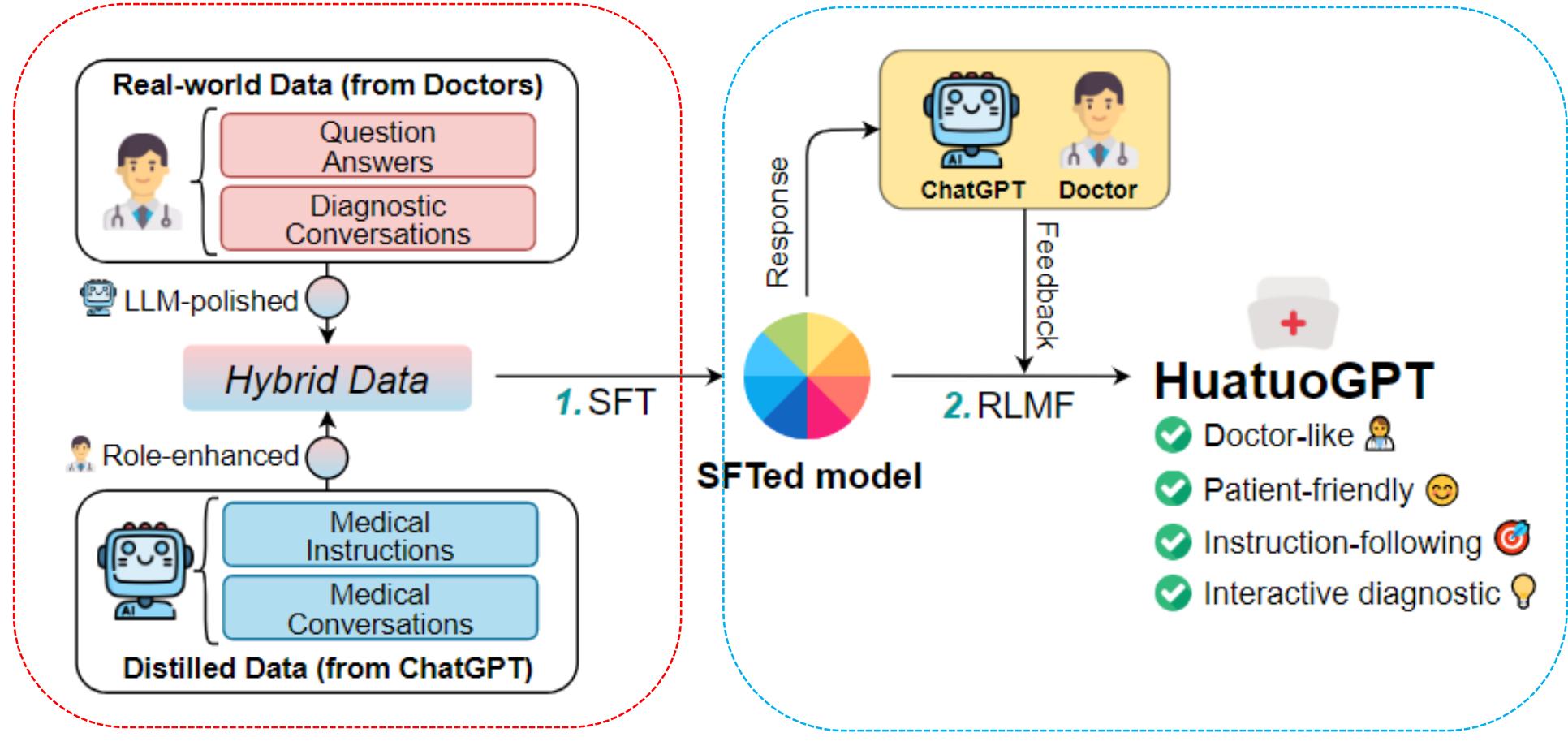
P Pearson

[Books links](#)

# Huatuo GPT



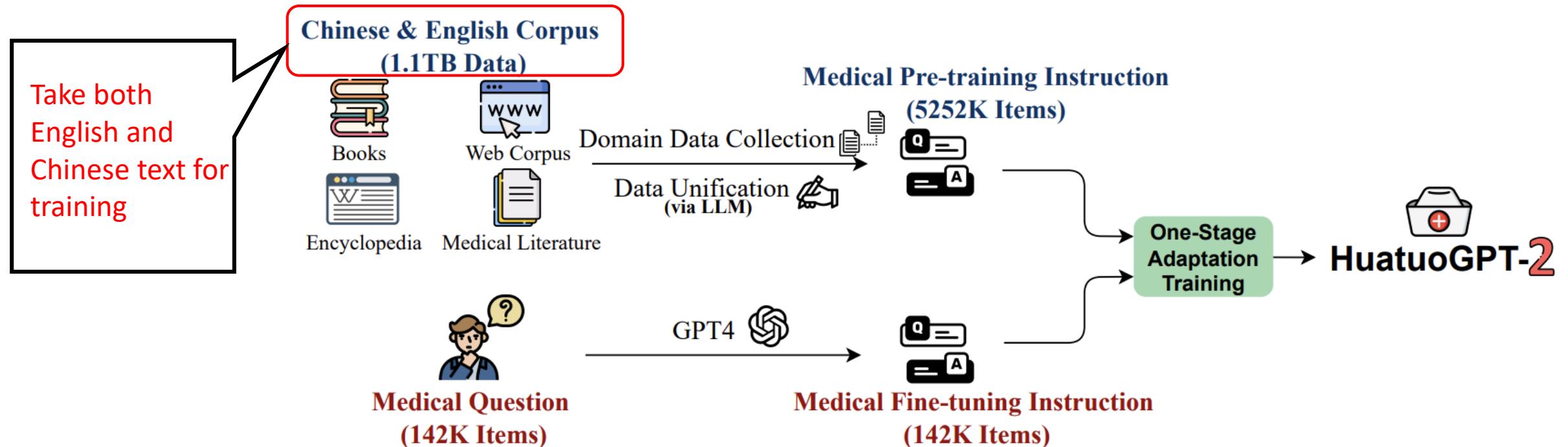
香港中文大學(深圳)  
The Chinese University of Hong Kong, Shenzhen



SFT with Hybrid Data

RL from Mixed Feedback

# Huatuo GPT-II



Junying Chen, Xidong Wang, Anningzhe Gao#, Feng Jiang, Shunian Chen, Hongbo Zhang, Dingjie Song, Wenya Xie, Chuyi Kong, Jianquan Li, Xiang Wan, **Haizhou Li, Benyou Wang#**. Huatuogpt-ii, one-stage training for medical adaption of llms. COLM 2024  
<https://arxiv.org/abs/2311.09774>

# The latest pharmacist exam in 2023 surpasses GPT-4 and 文心一言

Model	Pharmacist Licensure Examination (Pharmacy)					Pharmacist Licensure Examination (TCM)					Total Score	AVG
	Optimal Choice	Matched Selection	Integrated Analysis	Multiple Choice	Total Score	Optimal Choice	Matched Selection	Integrated Analysis	Multiple Choice	Total Score		
DISC-MedLLM	22.2	26.8	23.3	0.0	22.6	24.4	32.3	15.0	0.0	24.9	23.8	
HuatuoGPT	25.6	25.5	23.3	2.6	23.4	24.1	26.8	31.6	7.5	24.9	24.2	
ChatGLM2-6B	37.0	36.8	25.0	31.7	35.3	33.1	37.3	35.0	37.3	33.7	34.5	
ChatGLM3-6B	39.5	39.1	10.5	0.2	34.6	31.8	38.2	25.0	20.0	32.9	33.8	
Qwen-7B-chat	43.8	46.8	33.3	18.4	41.9	40.0	43.2	33.3	17.5	38.8	40.4	
Qwen-14B-chat	56.2	58.6	41.7	21.1	52.7	51.3	51.0	27.5	41.7	47.9	50.3	
Biachuan2-7B-Chat	51.2	50.9	30.0	2.6	44.6	48.1	46.0	35.0	7.5	42.1	43.4	
Biachuan2-13B-Chat	43.8	52.7	36.7	7.9	44.2	41.3	46.4	43.3	15.0	41.7	43.0	
文心一言	45.0	60.9	36.7	23.7	49.6	53.8	59.1	38.3	20.0	51.5	50.6	
ChatGPT(API)	45.6	44.1	36.7	13.2	41.2	34.4	32.3	30.0	15.0	31.2	36.2	
GPT-4(API)	65.1	59.6	46.7	15.8	57.3	40.6	42.7	33.3	17.5	38.8	48.1	
HuatuoGPT-II(7B)	41.9	61.0	35.0	15.7	47.7	52.5	51.4	41.7	15.0	47.5	47.6	
HuatuoGPT-II(13B)	47.5	64.1	45.0	23.7	52.9	48.8	61.8	45.0	17.5	51.6	52.3	
HuatuoGPT-II(34B)	66.3	75.0	48.3	34.2	65.5	63.6	71.4	50.0	27.5	62.5	64.0	

Hongbo Zhang et al., HuatuoGPT, towards Taming Language Model to Be a Doctor. <https://arxiv.org/abs/2305.15075>  
 Junying Chen et al., Huatuogpt-ii, one-stage training for medical adaption of llms. <https://arxiv.org/abs/2311.09774>

# 2024 Pharmacist Exam (Complete 960 Questions)

## Advanced LLMs (Over 20B Parameters)

模型	2024年中药	2024年西药	平均分
<i>c4ai-command-r-v01</i>	0.300	0.379	0.340
<i>Mixtral-8x7B-Instruct-v0.1</i>	0.325	0.354	0.340
<i>deepseek-llm-67B-chat</i>	0.465	0.479	0.472
<i>GPT-4o-mini</i>	0.469	0.648	0.559
<i>Claude-3-5-sonnet</i>	0.469	0.665	0.567
<i>LLaMA-3.1-70B-Instruct</i>	0.494	0.648	0.571
<i>Yi-34B-Chat</i>	0.654	0.654	0.654
<i>GPT-4o</i>	0.600	0.752	0.676
<b>HuatuoGPT2-34B</b>	0.669	0.700	<b>0.684</b>
<i>Qwen2.5-72B-it</i>	0.721	0.817	<b>0.769</b>

- Last year's HuatuoGPT-2 model can still pass this year's exam, maintaining an edge over GPT-4o.
- Qwen2.5-72B scores higher, with an average accuracy of 76.8%.
- Models like LLaMA-3-70B and Claude perform moderately in Chinese, especially in Traditional Chinese Medicine.

# Clinical doctors' manual assessments surpass GPT-4

<b>HuatuoGPT-II Win Rate</b>	<b>Win</b>	<b>Tie</b>	<b>Fail</b>
HuatuoGPT-II(7B) vs GPT-4	38	38	24
HuatuoGPT-II(7B) vs ChatGPT	52	33	15
HuatuoGPT-II(7B) vs Baichuan2-13B-Chat	63	19	18
HuatuoGPT-II(7B) vs HuatuoGPT	81	11	8

Table 6: Results of Expert Assessment on Single-round Medical Response.

<b>HuatuoGPT-II Win Rate</b>	<b>Win</b>	<b>Tie</b>	<b>Fail</b>
HuatuoGPT-II(7B) vs GPT-4	53	17	30
HuatuoGPT-II(7B) vs ChatGPT	56	11	33
HuatuoGPT-II(7B) vs Baichuan2-13B-Chat	63	19	18
HuatuoGPT-II(7B) vs HuatuoGPT	68	6	26

Table 7: Results of Expert Assessment on Multi-round Medical Dialogue.

# The largest medical QA dataset: Huatuo-26M

Dataset	Lang	Domain	Source	#Q
MedHop (Welbl et al., 2018)	English	Medical	MEDLINE	2.5K
BiQA (Lamurias et al., 2020)	English	Medical	Online Medical forum	7.4K
HealthQA (Zhu et al., 2019)	English	Medical	Medical-services website	7.5K
MASH-QA (Zhu et al., 2020)	English	Medical	Medical article website	35K
MedQuAD (Ben Abacha and Demner-Fushman, 2019)	English	Medical	U.S. National Institutes of Health (NIH)	47K
ChiMed (Tian et al., 2019)	Chinese	Medical	Online Medical forum	47K
MedQA (Jin et al., 2020)	EN&CH	Medical	Medical Exam	60K
webMedQA (He et al., 2019)	Chinese	Medical	Medical consultancy websites	63K
CliCR (Šuster and Daelemans, 2018)	English	Medical	Clinical case reports	100K
cMedQA2 (Zhang et al., 2018)	Chinese	Medical	Online Medical forum	108K
<b>Huatuo-26M</b>	<b>Chinese</b>	<b>Medical</b>	<b>Consultation records, Encyclopedia, KBs</b>	<b>26M</b>
TriviaQA (Joshi et al., 2017)	English	General	Trivia	96K
HotpotQA (Yang et al., 2018)	English	General	Wikipedia	113K
SQuAD (Rajpurkar et al., 2016)	English	General	Wikipedia	158K
DuReader (He et al., 2017)	Chinese	General	Web search	200K
Natural Questions (Kwiatkowski et al., 2019)	English	General	Wikipedia	323K
MS MARCO (Nguyen et al., 2016)	English	General	Web search	1.0M
CNN/Daily Mail (See et al., 2017)	English	General	News	1.3M
PAQ (Lewis et al., 2021)	English	General	Wikipedia	65M

Jianquan Li, Xidong Wang, Xiangbo Wu, Zhiyi Zhang, Xiaolong Xu, Jie Fu, Prayag Tiwari, Xiang Wan, Benyou Wang. Huatuo-26M, a **Large-scale Chinese Medical QA Dataset**.  
<https://arxiv.org/abs/2305.01526>

# Huatuo GPT-II ---- Demo1

长期大量服用哪种维生素会引起严重的神经感觉异常，导致进行性步态不稳、足麻木和手不灵活？



## 模型内部深度思考



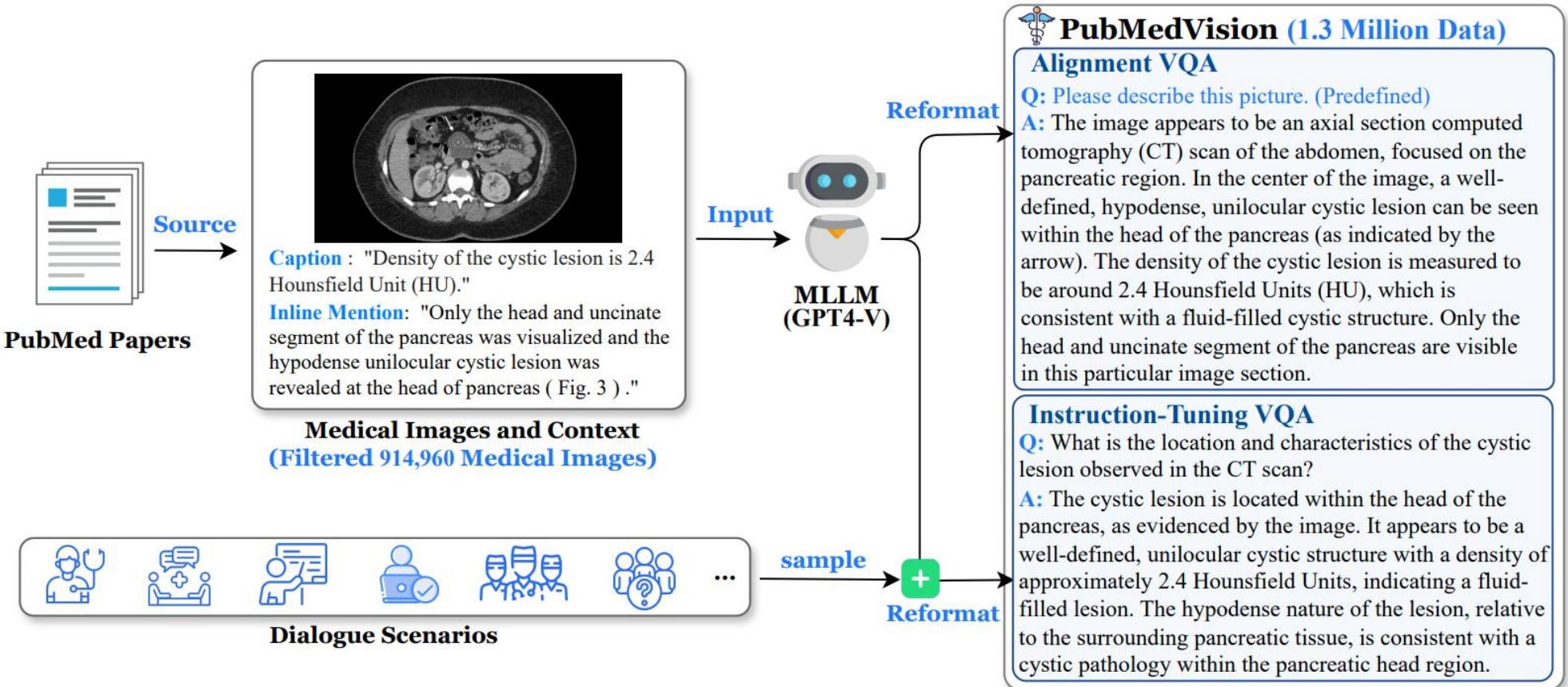
长期大量服用哪种维生素会引起严重的神经感觉异常呢？让我想想。首先，维生素A、B族维生素、C、D、E，这些都得考虑一下。不过，维生素A好像主要是和皮肤和黏膜有关，不太容易导致神经问题。维生素C和E呢，它们主要是抗氧化，应该也不会引起神经异常。维生素D呢，更多是和钙磷代谢有关，感觉也不太可能。那剩下的就是B族维生素了。

B族维生素中，维生素B1、B2、B3、B6、B12都有各自的作用，但好像B12和神经关系比较密切。不过，B12缺乏的话，通常会导致贫血和神经病变，而不是因为过量。那剩下的B1、B2、B3、B6，哪个更有可能呢？

长期大量服用维生素B6会引起严重的神经感觉异常，导致进行性步态不稳、足麻木和手不灵活。这种现象被称为维生素B6神经病，是由于过量摄入维生素B6导致的周围神经病变。因此，答案是维生素B6。



# Multi-modal Huatuo GPT: Injecting Multimodal Medical knowledge



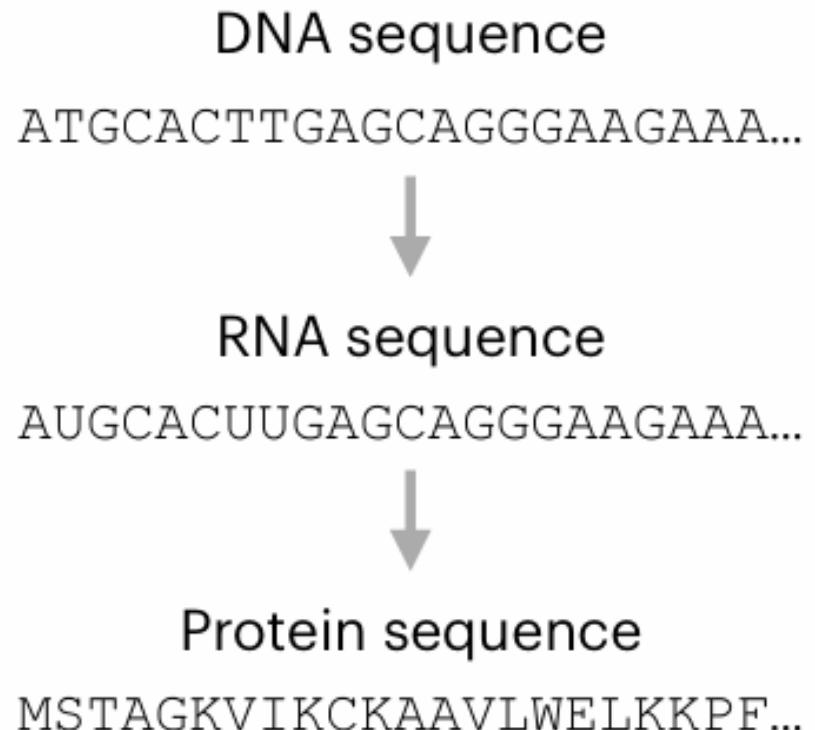


# Outline

- Introduction of LLM
- Medical LLM
- AI for protein structure prediction
- AI for new drug discovery & development
- AI in medical diagnostics & healthcare
- LLM agent hospital

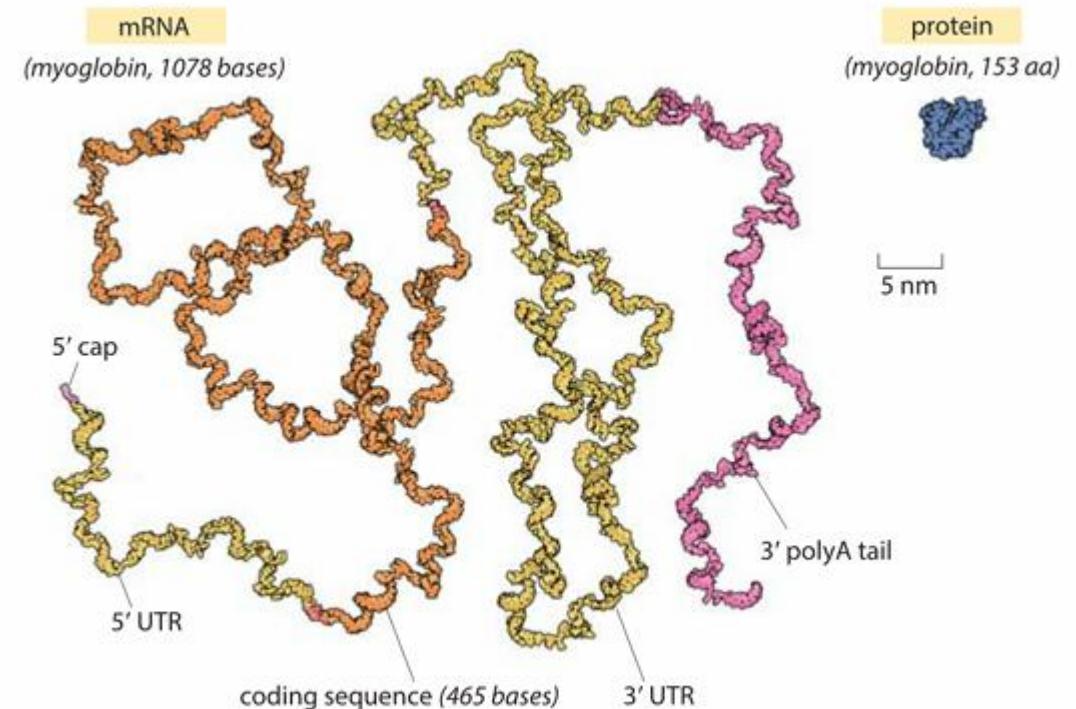
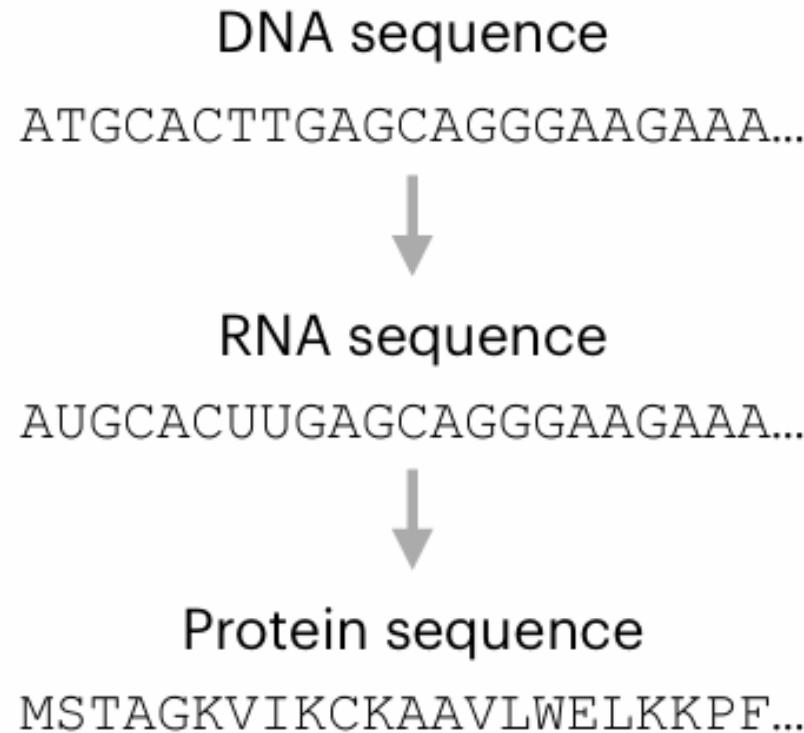
# AI for Protein Structure Prediction

# Protein ---The central dogma of molecular biology



- Human genome:
- Contains around 3 billion base pairs
  - Encodes ~20k genes
- Proteins are the final product of the genetic information flow
- Modern **molecular biology research**: how is life implemented by our genetic code?

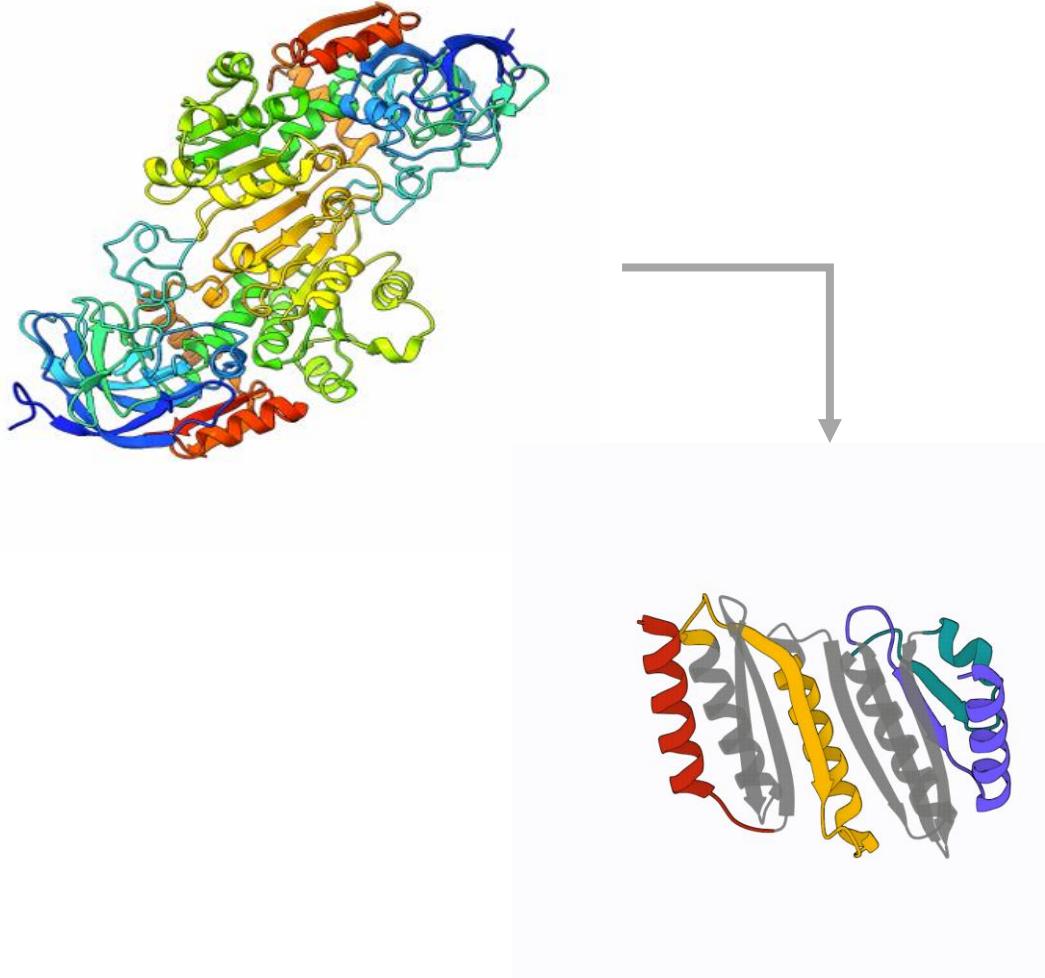
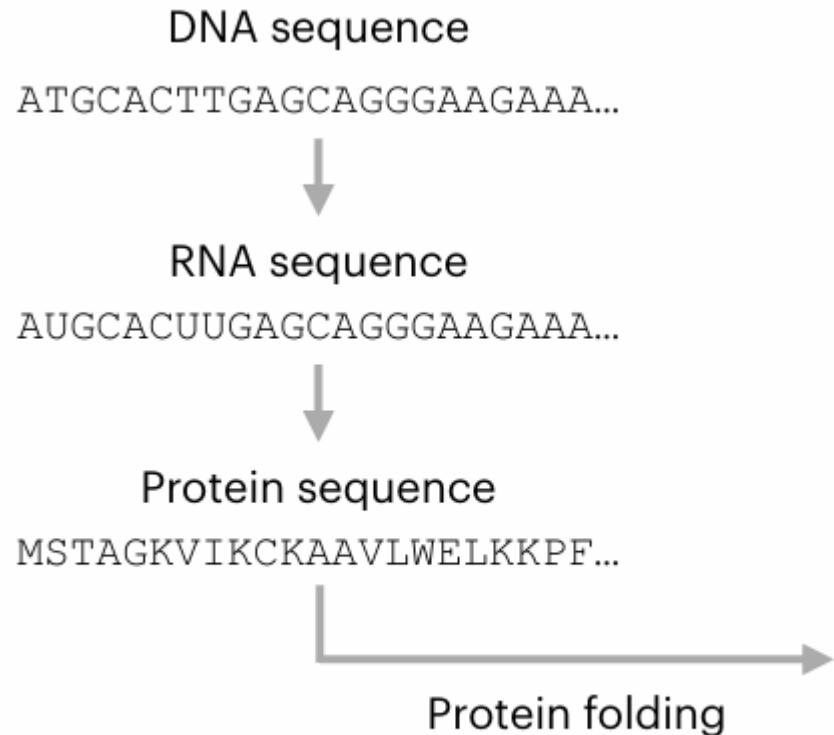
# Protein Structure biology: The study of proteins and other biomolecules through their 3D structure



Cell Biology By The Numbers. Illustration by David Goodsell.

(Zaixiang Zheng et al., 2024)

# Protein Structure biology: The study of proteins and other biomolecules through their 3D structure

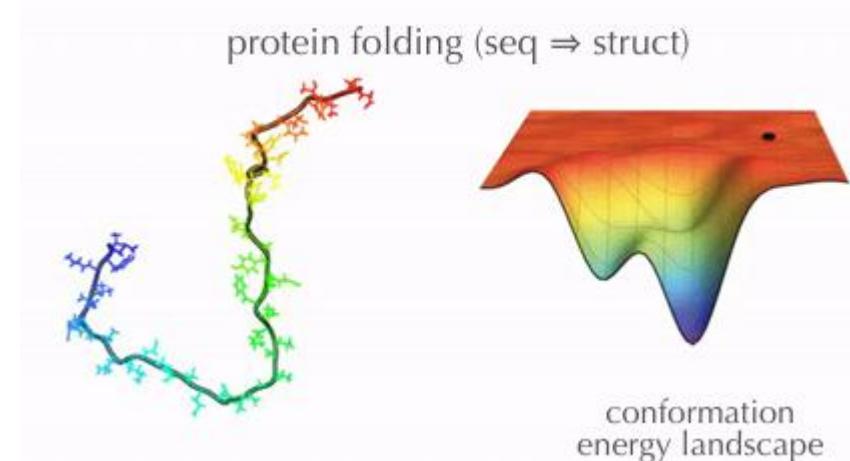
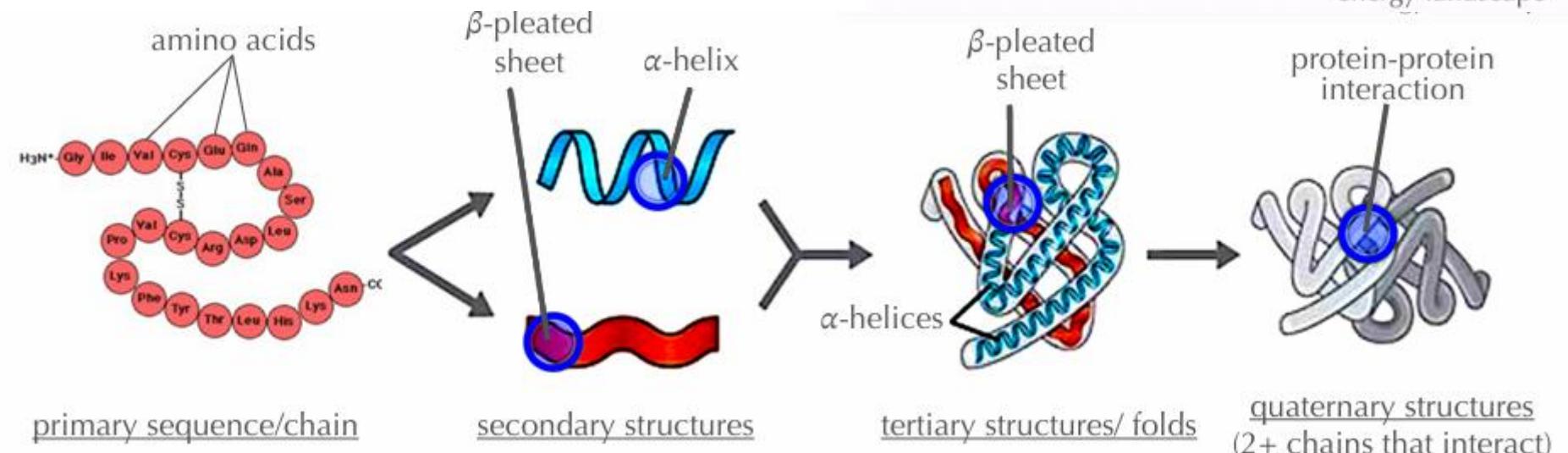


(Zaixiang Zheng et al., 2024)

# Protein --- data modalities

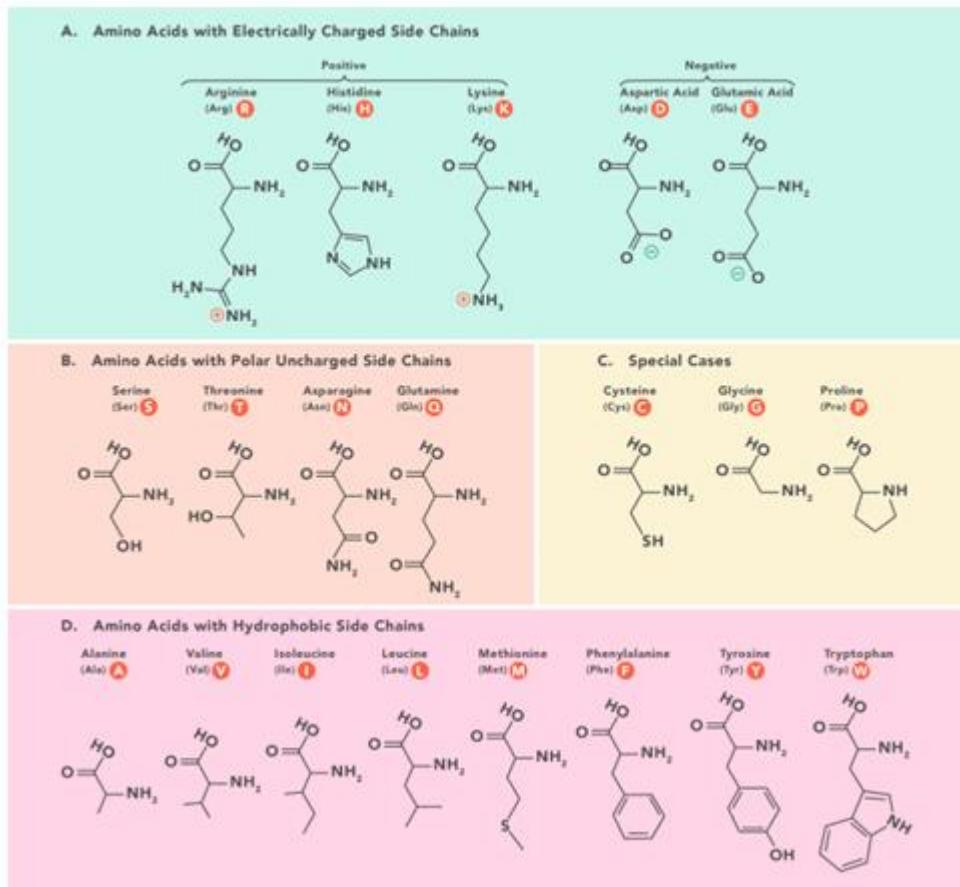
## **sequence $\Leftrightarrow$ structure $\Leftrightarrow$ function**

- A sequence over 20 amino acids (AAs)
- In solvent will fold into a **unique 3D spatial structure** with minimal free energy
- **Structure determines protein function**



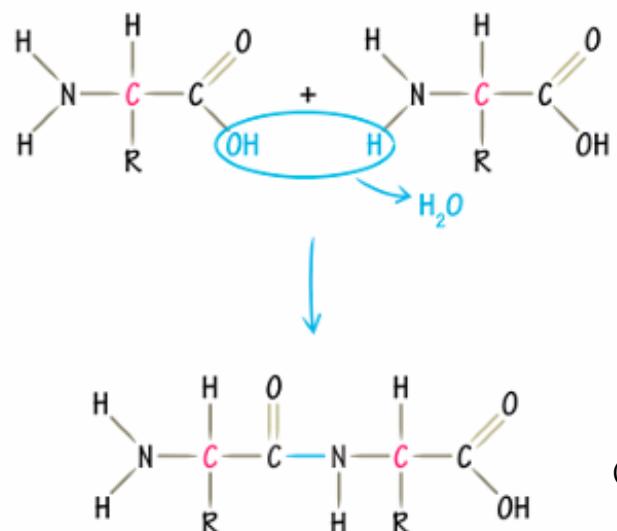
# Sequence of amino acids

**Sequence:** 20 types of amino acids  $s \in \{1\dots20\}^N$



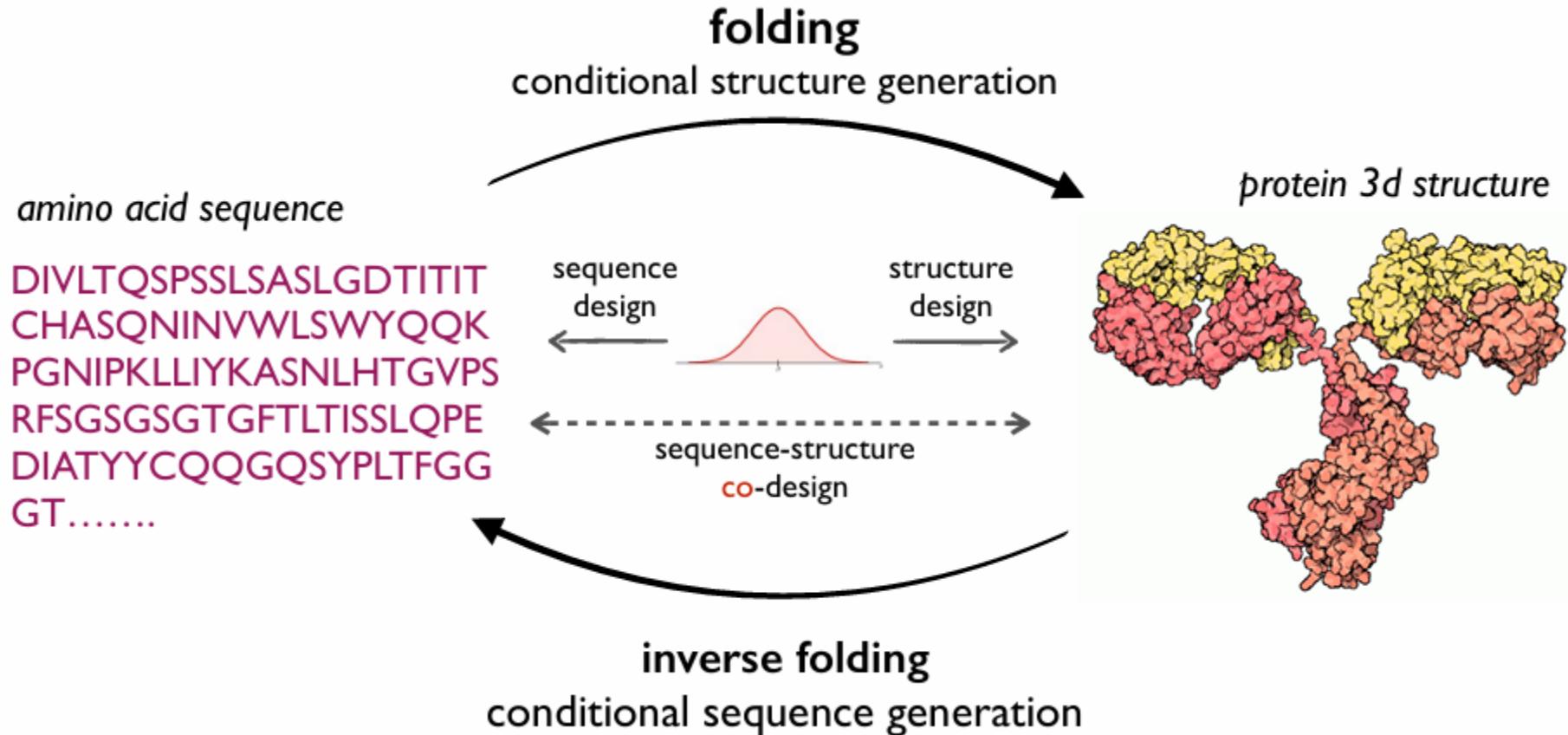
- A linear sequence of amino acids **polymerized in a chain**
- An alphabet of twenty possible amino acids
  - **Common backbone** but **different side chains**
- Various non-covalent interactions and other forces drive folding of the chain into a globular 3D structure\*

## Peptide Bond Formation



(Zaixiang Zheng et al., 2024)

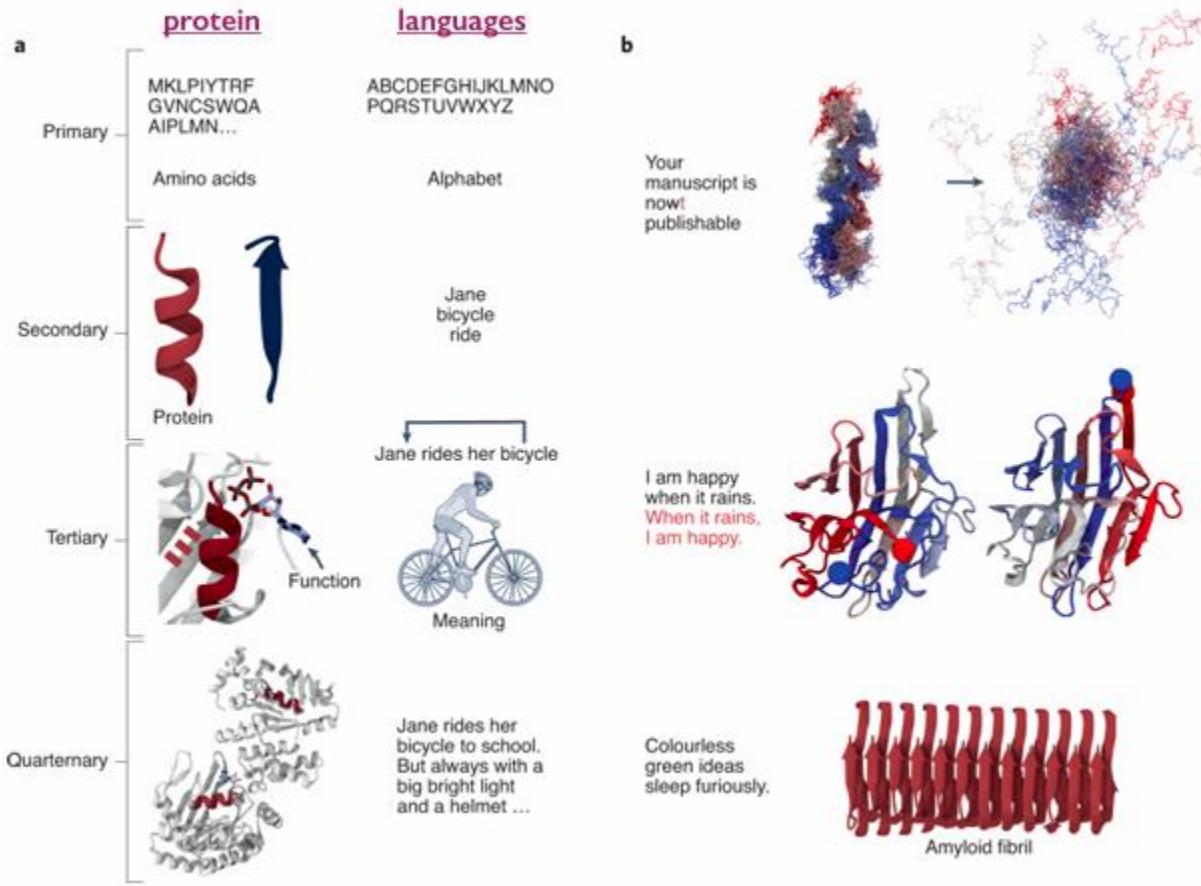
# Designing protein sequence and structure as generative modeling problems



pdb id: IIGT. from <https://www.rcsb.org/structure/1IGT>

(Zaixiang Zheng et al., 2024)

# Discreteness in NLP and Biology



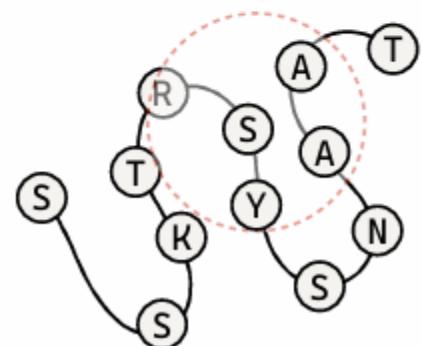
**Goal:** learning joint prob. of sequence of discrete tokens

$$p_{\theta}(\mathbf{x}) = p_{\theta}(x_1, \dots, x_i, \dots, x_n)$$

$$\approx \frac{1}{Z} \exp \left( \sum_i \phi_i(x_i) + \sum_{i,j} \psi_{ij}(x_i, x_j) \right)$$



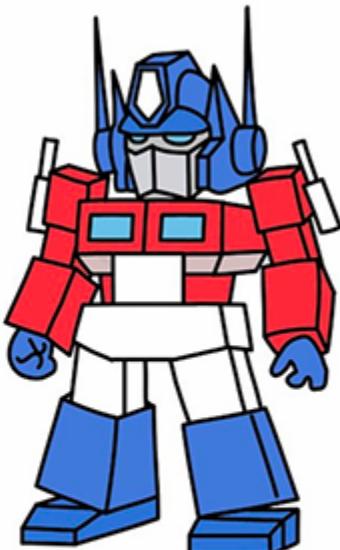
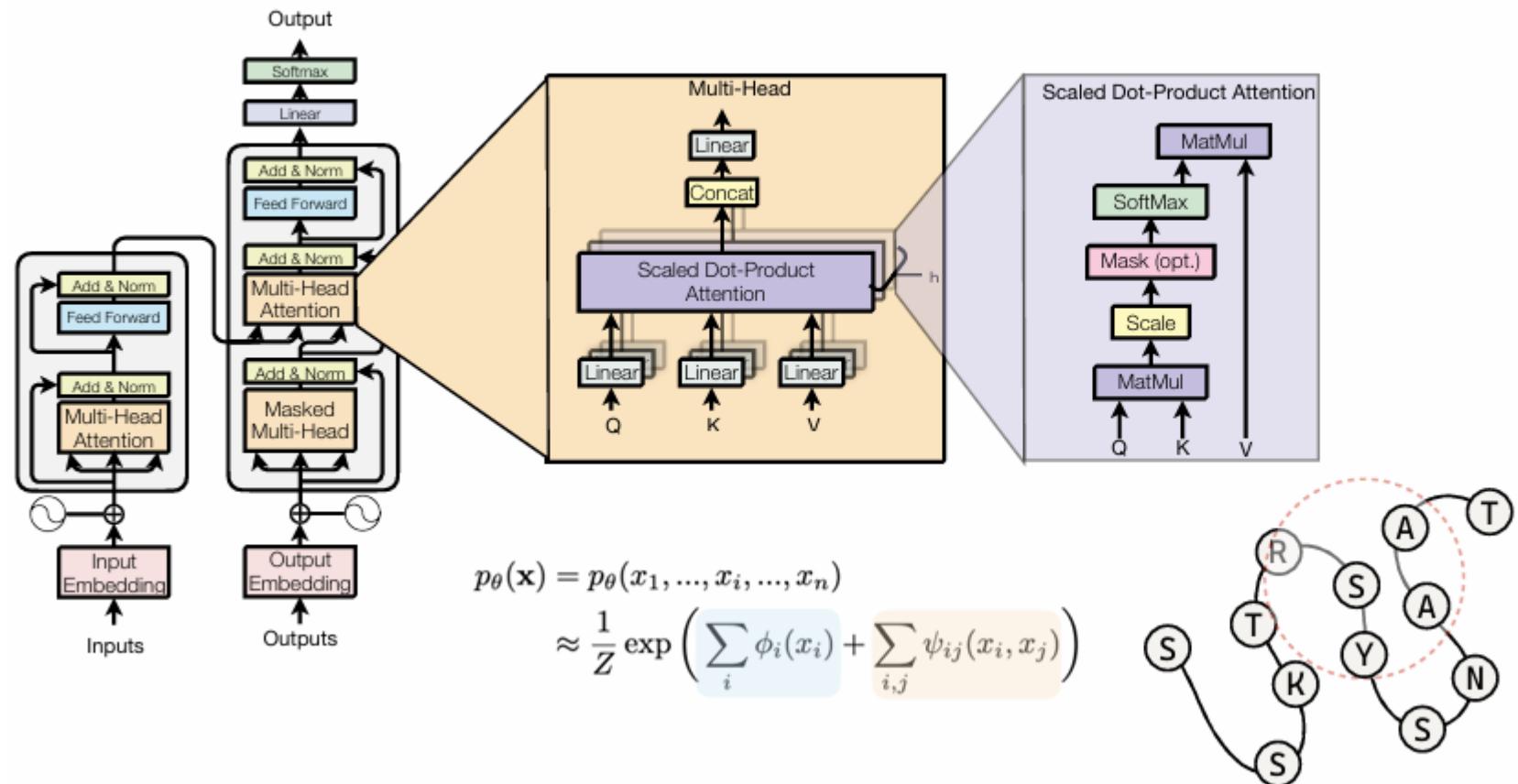
**Factorization** (wrt the **structures** of data) **needed!**



(Zaixiang Zheng et al., 2024)

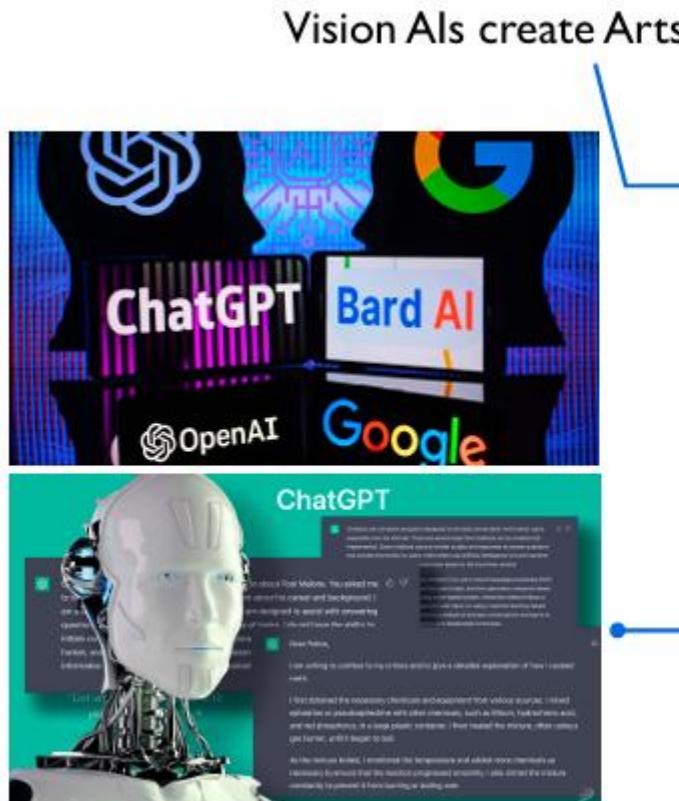
# The parameterization basis of everything: Transformer - a scalable & strong learner for structured data

**Why:** Stacking pair-wise attention → modeling statistical dependencies inside data  
 (for protein data: residue-wise structural interaction & co-evolution, etc) → approximates global data dist.



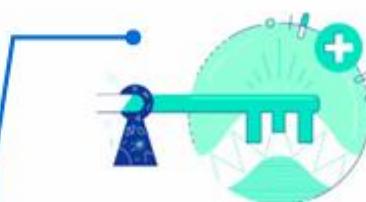
(Zaixiang Zheng et al., 2024)

# Amazing things that generative AI can do



Large LMs speak

DeepMind



AlphaFold learns protein folding

(Zaixiang Zheng et al., 2024)

# Diffusion(-Like) Modeling for Protein Structure: AlphaFold

# Protein structure prediction: AlphaFold2 & 3

## nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 22 July 2021

## Highly accurate protein structure prediction for the human proteome

Kathryn Tunyasuvunakool  , Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, Sameer Velankar, Gerard J. Kleywegt, Alex Bateman, Richard Evans, Alexander Pritzel, Michael Figurnov, Olaf Ronneberger, Russ Bates, Simon A. A. Kohl, Anna Potapenko, Andrew J. Ballard, Bernardino Romera-Paredes, Stanislav Nikolov, Rishabh Jain, ... Demis Hassabis  + Show authors

[Nature](#) 596, 590–596 (2021) | [Cite this article](#)

322k Accesses | 1720 Citations | 1421 Altmetric | [Metrics](#)

**nature**

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [addenda](#) > [article](#)

Addendum | [Open access](#) | Published: 27 November 2024

## Addendum: Accurate structure prediction of biomolecular interactions with AlphaFold 3

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, ... John M. Jumper  + Show authors

[Nature](#) 636, E4 (2024) | [Cite this article](#)

12k Accesses | 1 Citations | 12 Altmetric | [Metrics](#)

 The [Original Article](#) was published on 08 May 2024



Google DeepMind

# Protein structure prediction: AlphaFold2 & 3

Kathryn Tunyasuvunakool, John Jumper, Demis Hassabis, and colleagues used AlphaFold to determine the structures of nearly the entire human proteome (**98.5% of all human proteins**). AlphaFold provided confident predictions for 58% of amino acid positions, with 35.7% reaching high confidence, doubling experimental coverage. At the protein level, 43.8% of proteins had at least three-quarters of their sequences confidently predicted. The authors believe large-scale accurate predictions will be a vital tool for addressing scientific questions and elucidating protein functions.

The authors believe that **large-scale**, accurate **structural predictions** will become a crucial tool, enabling us to address new scientific questions from a structural perspective. The **predictions from AlphaFold** will help further **elucidate** the roles of **proteins**.

## AlphaFold2

## AlphaFold3

**AlphaFold3** is a revolutionary model, excelling in breadth and accuracy. Unlike previous tools limited to specific biomolecules (e.g., proteins or RNA), **it predicts structures and interactions for nearly all biological molecules**, showcasing its breadth. Simultaneously, it achieves significant accuracy improvements: **predictions for protein interactions with other molecules are at least 50% better than existing methods**, doubling accuracy for some critical interaction types.

**AlphaFold3** has made the AlphaFold Server freely available to the scientific community, which includes most of the features of AlphaFold3, allowing scientists to perform structure predictions. The link to the **AlphaFold Server** is as follows:



# Structure-informed Language Models Are Protein Designers

[ICML 2023 Oral]

Zaixiang Zheng<sup>1\*</sup>, Yifan Deng<sup>2\*</sup>, Dongyu Xue<sup>1</sup>, Yi Zhou<sup>1</sup>, Fei Ye<sup>1</sup> and Quanquan Gu<sup>1</sup>

<sup>1</sup>ByteDance Research & <sup>2</sup>UW-Madison



(Zaixiang Zheng et al., 2024)

# ProteinGPT: LLM 4 protein prediction & structure understanding

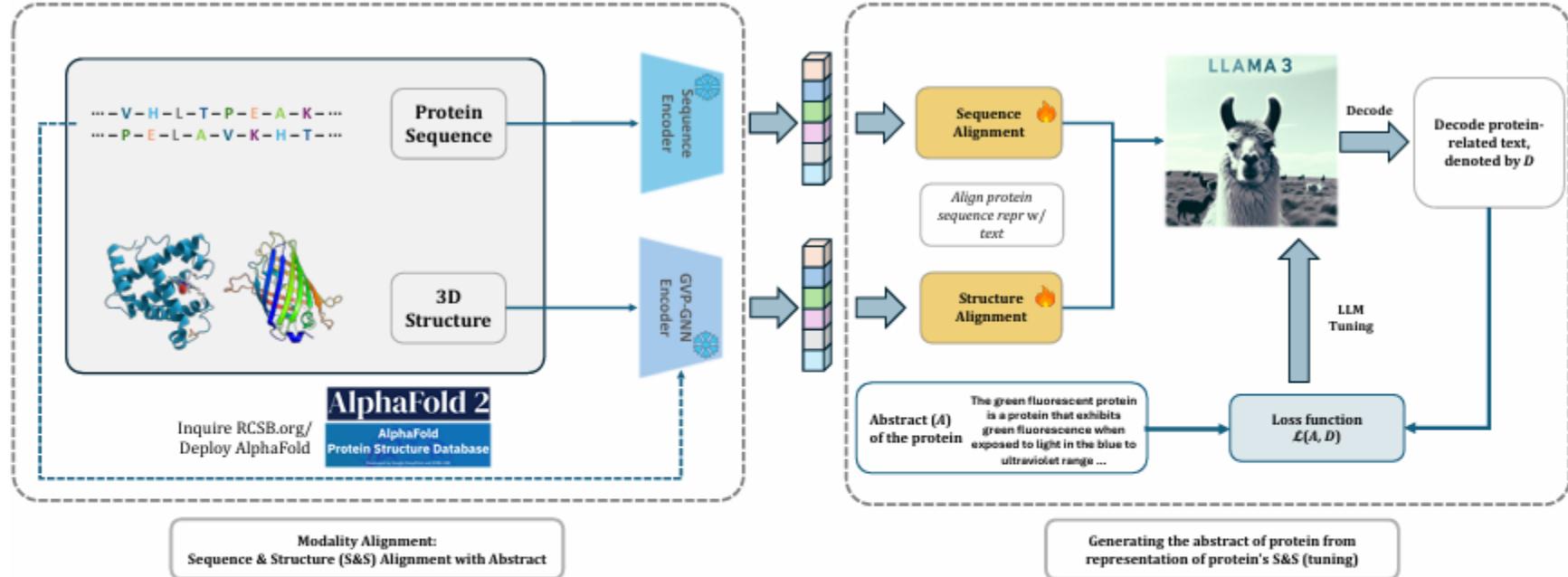


Figure 1:  ProteinGPT **Modality Fusion** & **Alignment Stage**: we freeze the encoder blocks and train the linear project layer to learn how to align protein structure and protein sequence representations with text. In the alignment stage, the input to the training is only the projected protein representation. No text prompts are incorporated in this stage.

# ProteinGPT: LLM 4 protein prediction & structure understanding

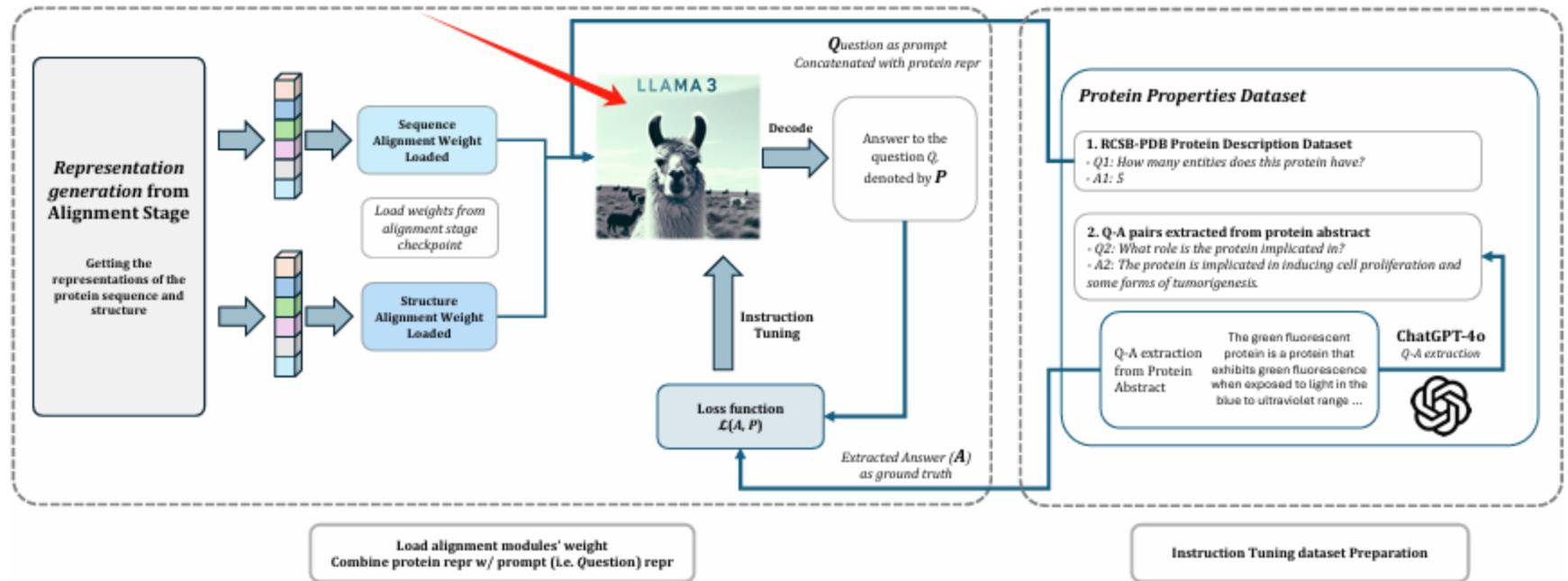


Figure 2:  ProteinGPT Instruction Tuning Stage: we utilize the QA pairs and property tags in ProteinQA to tune the LLM to follow instructions and give concise responses. For instruction alignment, explicit prompts (*Questions* on the protein) are included at the beginning of the prompt.

# ProteinGPT: LLM 4 protein prediction & structure understanding

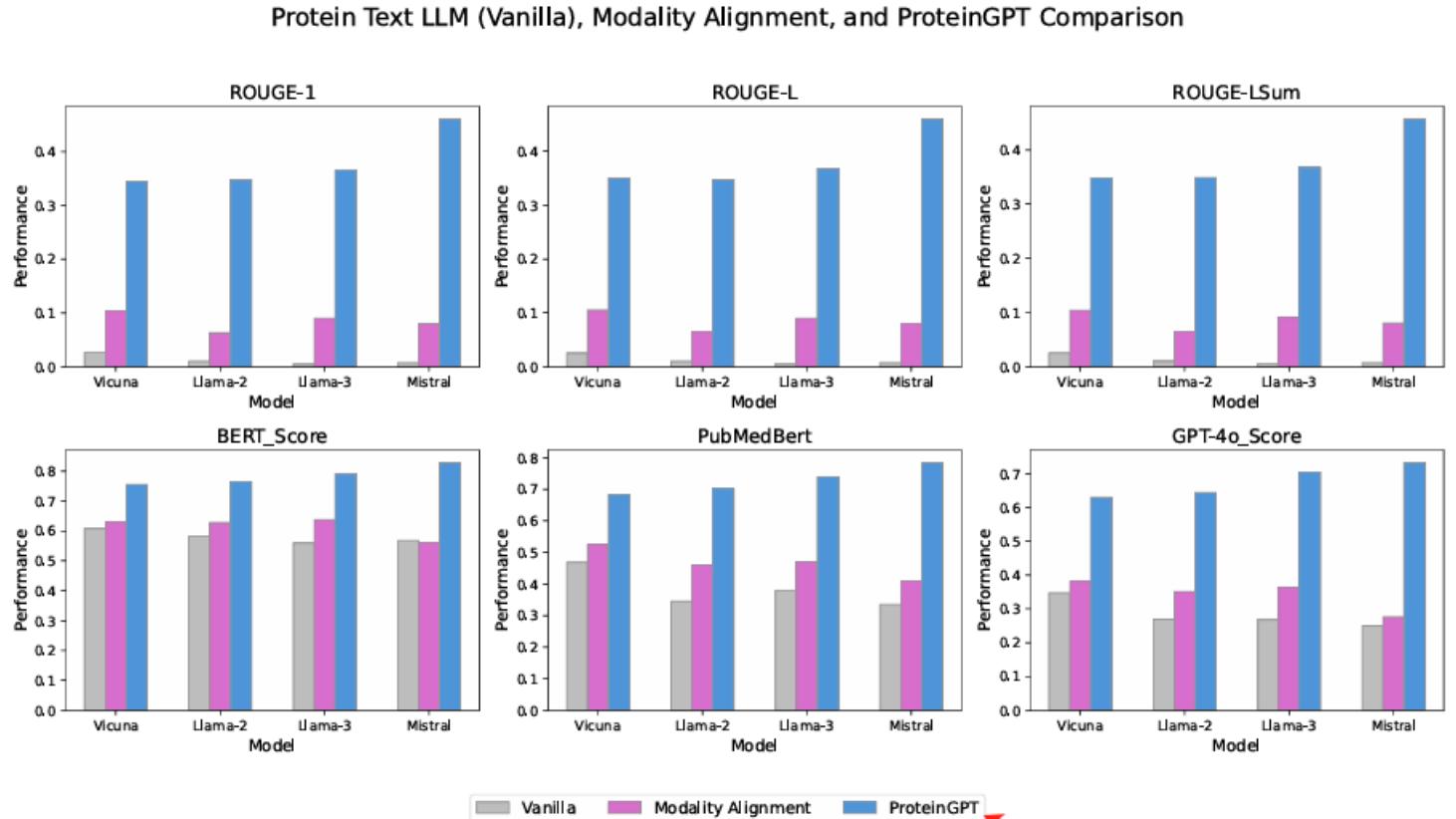


Figure 5: Performance improves progressively from the vanilla LLM model with protein as text to the modality-aligned version, and finally to the instruction-tuned variants of ProteinGPT. Each stage of ProteinGPT's training results in substantial enhancements in both lexical and semantic performance, showcasing the efficiency of our framework.

Yijia Xiao et al., 2024



# Diffusion Language Models Are Versatile Protein Learners

[ICML 2024]

Xinyou Wang\*♡, Zaixiang Zheng\*♡, Fei Ye♡, Dongyu Xue♡, Shujian Huang◇, and Quanquan Gu♡  
♡ByteDance Research & ◇Nanjing University



# Outline

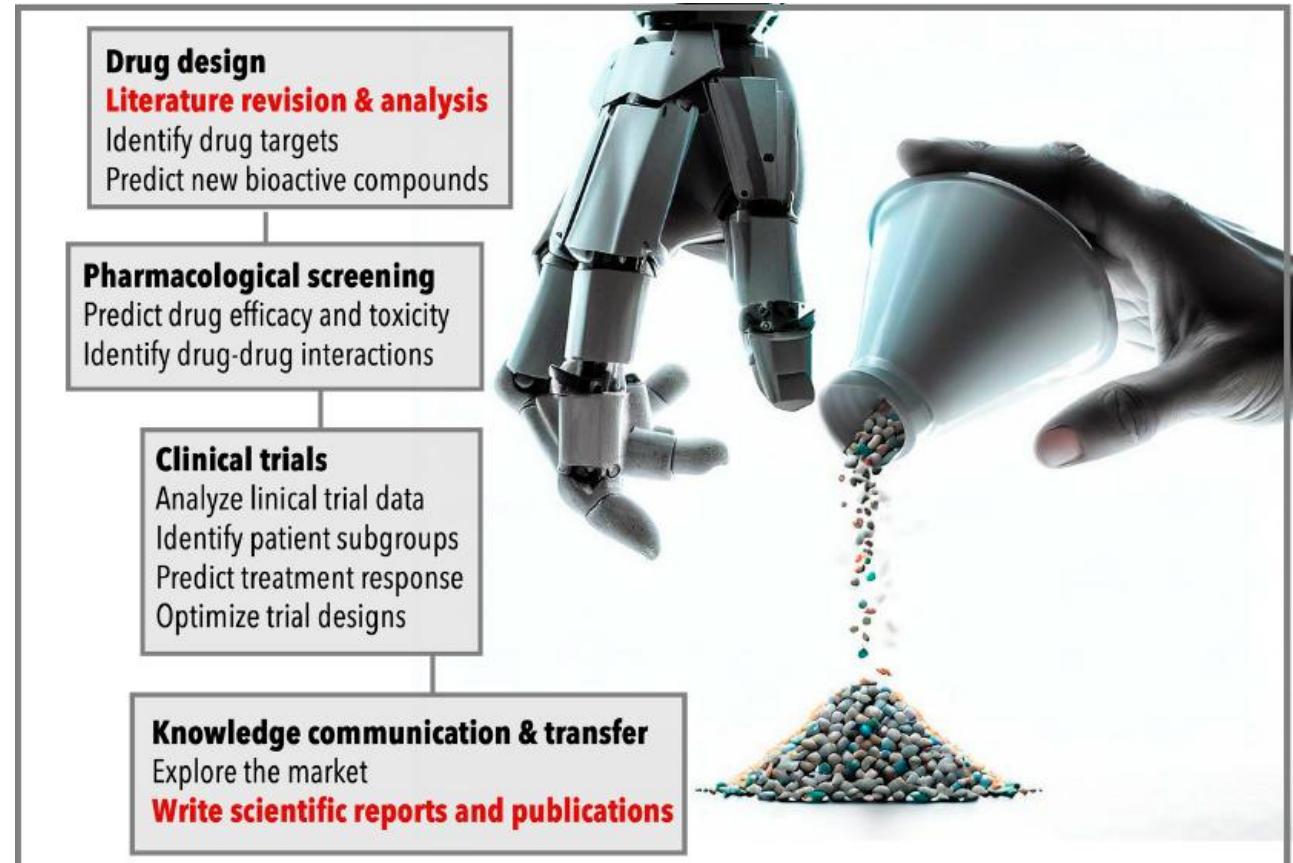
- Introduction of LLM
- Medical LLM
- AI for protein structure prediction
- AI for new drug discovery & development
- AI in medical diagnostics & healthcare
- LLM agent hospital

# AI for New Drug Discovery

# Drug discovery and development

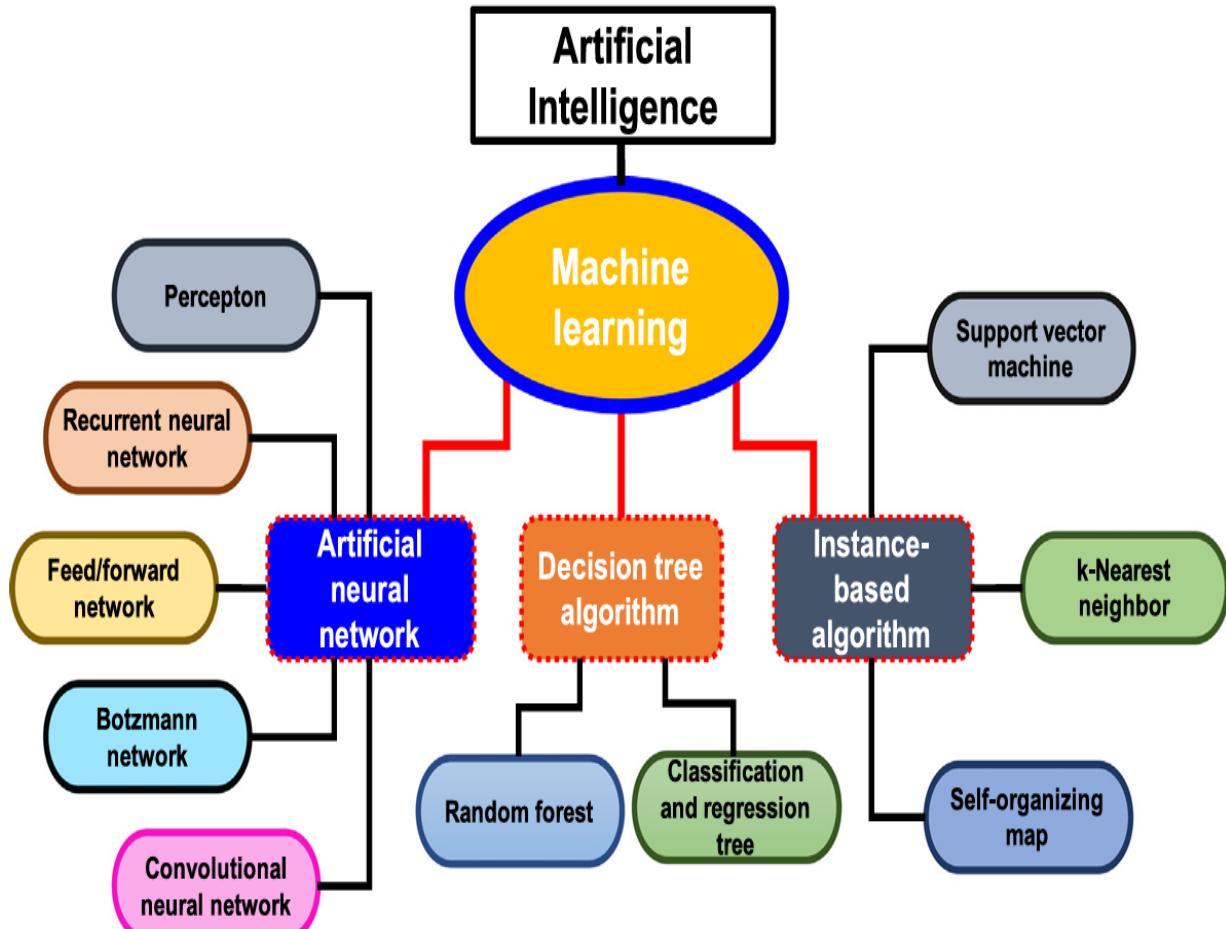
AI has the potential to revolutionize the drug discovery process, offering **improved efficiency and accuracy**, accelerated drug development, and the capacity for the development of more effective and personalized treatments ([Figure 1](#)).

However, the successful application of **AI in drug discovery** is **dependent on** the availability of **high-quality data**, the addressing of ethical concerns, and the recognition of the limitations of AI-based approaches.

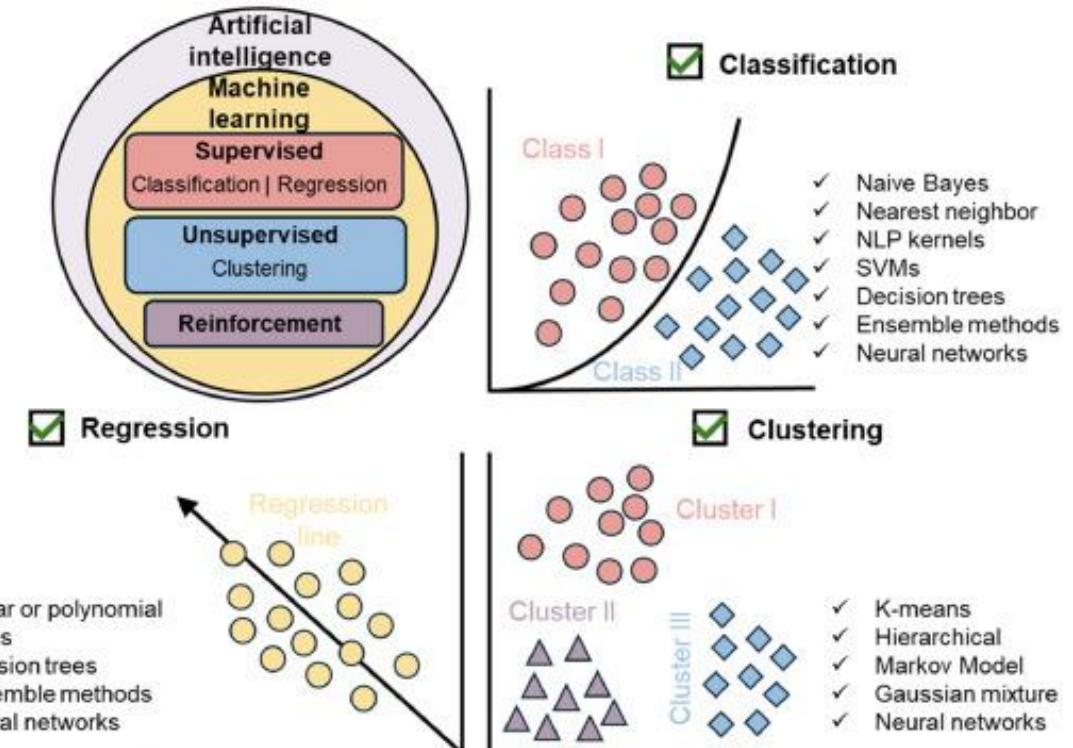


(Alexandre Blanco-González et al., 2023)

# AI, ML, DL algorithms

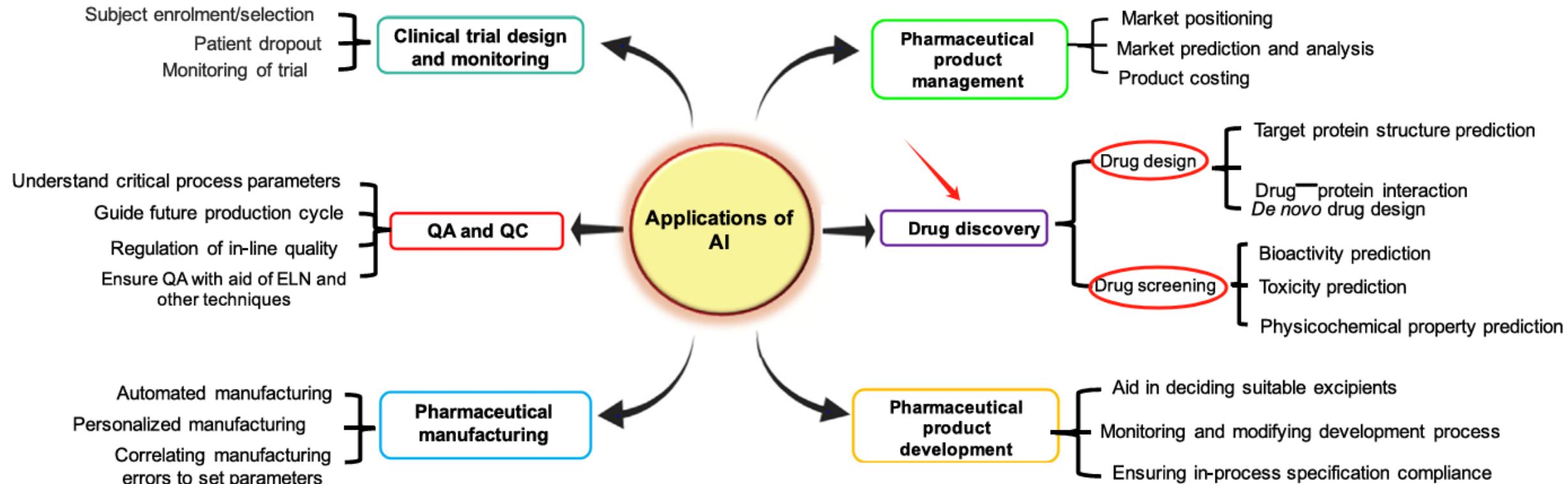


Debleena Paul et al., 2021



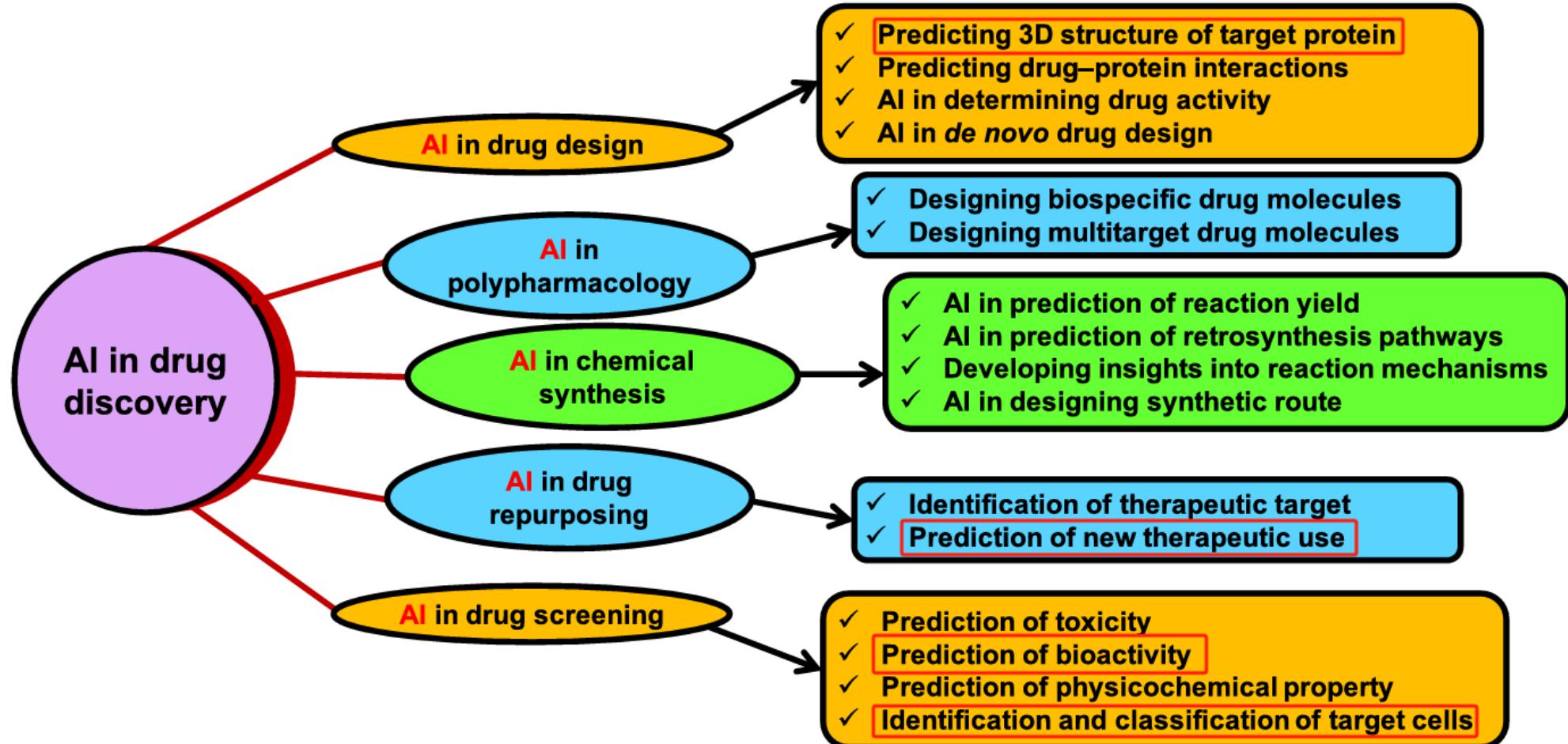
Ashfaq Ur Rehman et al., 2024

# AI applications



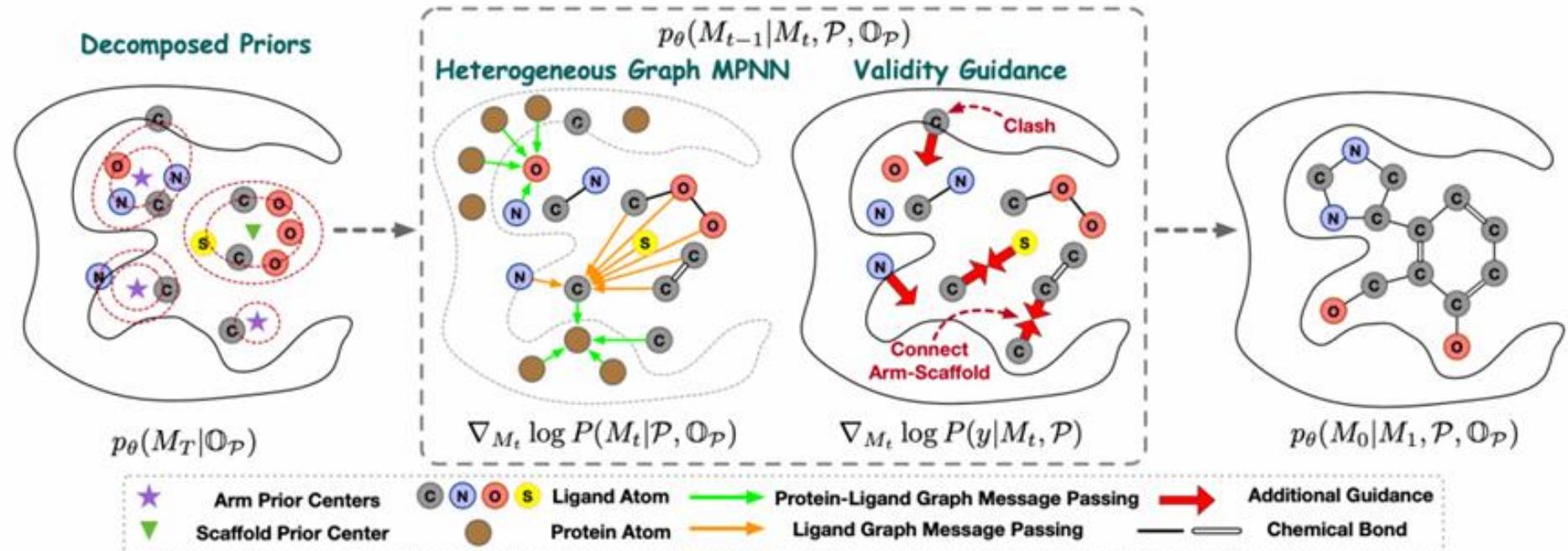
Debleena Paul et al., 2021

# AI applications for drug discovery



Debleena Paul et al., 2021

# Small Molecule Drug Design: DecompDiff



**Figure 2.** Overview of the sampling process of DecompDiff. (a) The initial atoms are sampled from informative decomposed priors. (b) An equivariant network on heterogeneous graphs denoises atom coordinates, atom types and bond types simultaneously. (c) The validity guidance alleviates the protein-ligand clash problem and encourages arms and scaffold to connect.

# Example of AI tools used in drug discovery



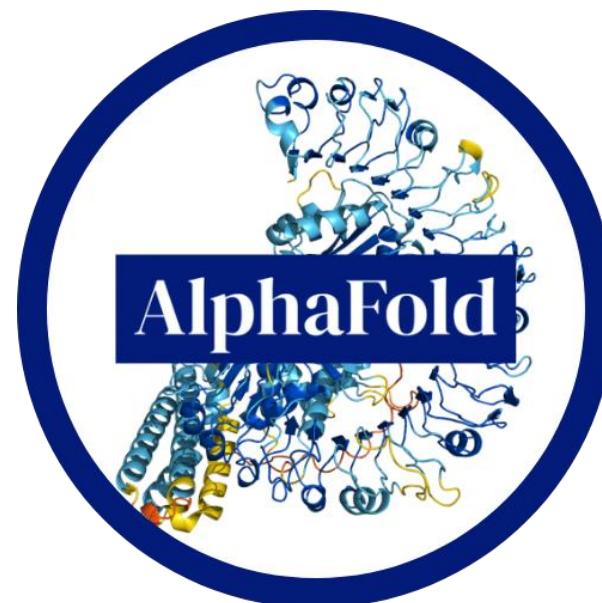
DEB-TKTD EP<sub>x</sub> Predictor



deepchem



ORGANIC



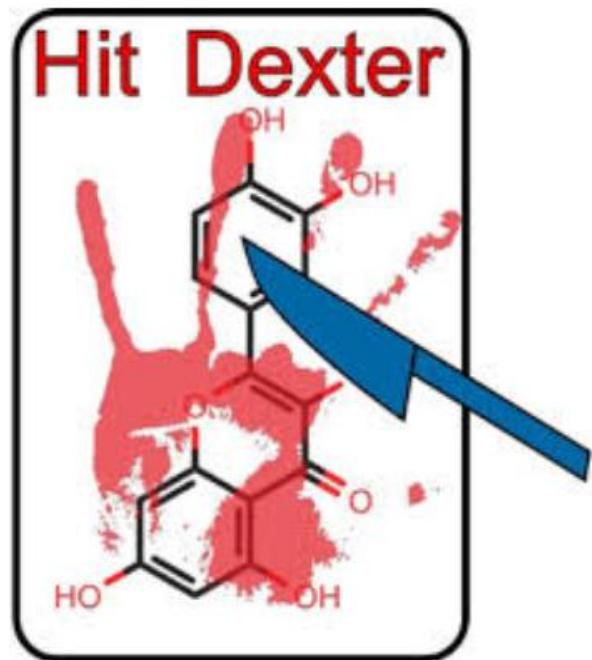
DeltaVina



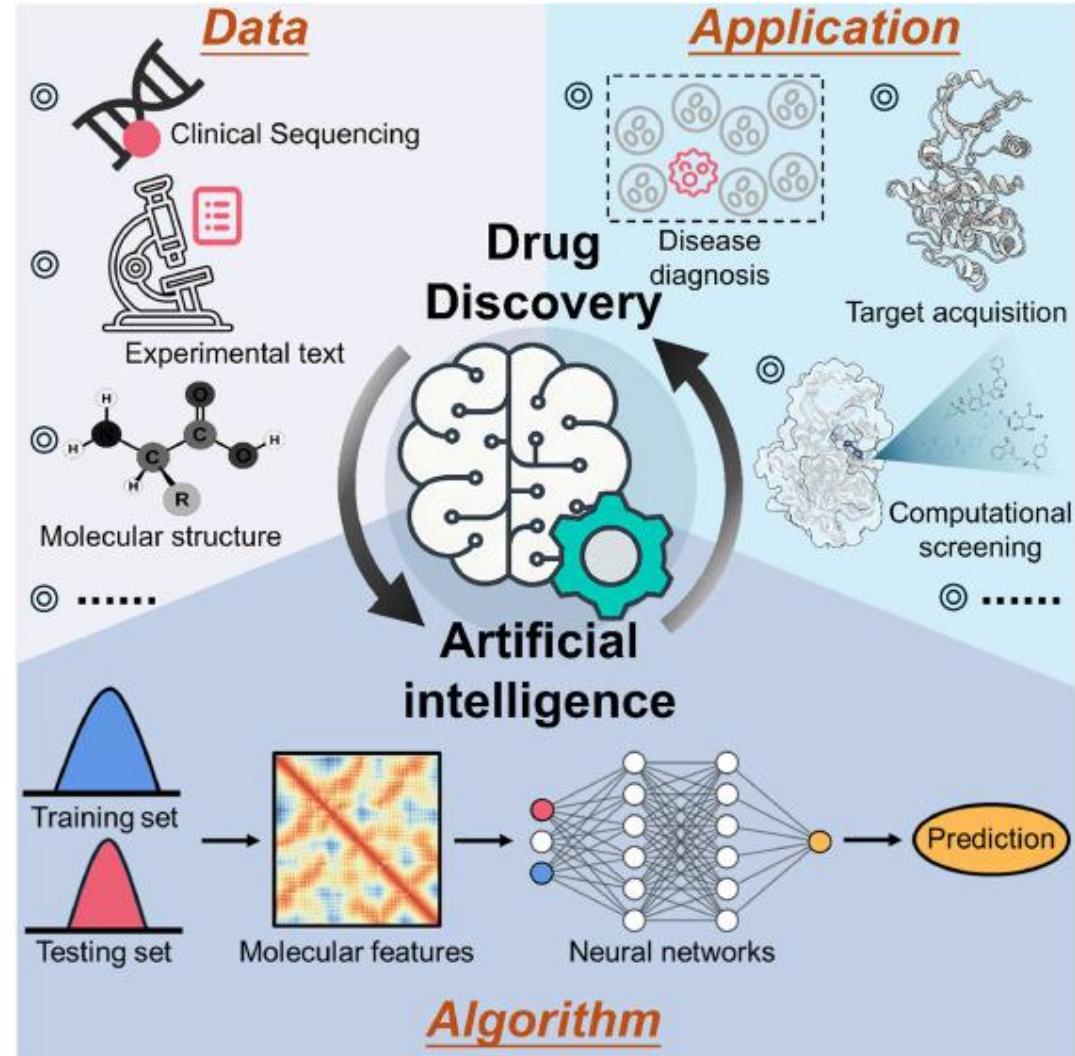
neural-fingerprint



DeepNeuralNet-QSAR



# AI for drug discovery



Ashfaq Ur Rehman et al., 2024

# Outline

- Introduction of LLM
- Medical LLM
- AI for protein structure prediction
- AI for new drug discovery & development
- AI in medical diagnostics & healthcare
- LLM agent hospital

# AI in Medical Diagnostics and Healthcare

# Can AI Make Medicine More Human?

**HARVARD medicine**  
THE MAGAZINE OF HARVARD MEDICAL SCHOOL

## Can AI Make Medicine More Human?

The history of tools used to support clinical decision-making offers clues to the future of medicine in the age of generative AI

Autumn 2024

by Adam Rodman

15 min read

Feature

In The Doctor, an 1891 painting by Luke Fildes, a physician attends to a sick child.

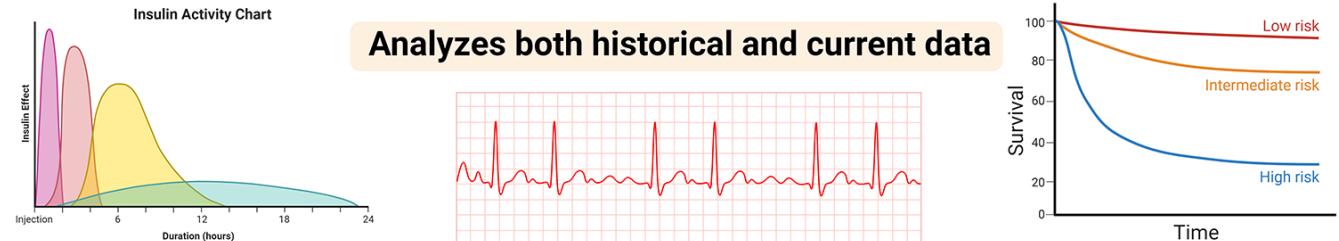


"AI has the potential to make medicine more human by **freeing up clinicians** to focus on the **personal, empathetic aspects** of care."

"By handling routine tasks and data analysis, AI can allow doctors to **spend more time with patients**, fostering **deeper connections** and better **outcomes**."

# Outcomes: revolutionizing healthcare --- clinical practice

## AI-Powered Predictive Analysis: Revolutionizing Clinical Practice



Analyzes both historical and current data

- It requires:
- 1- Quality data
  - 2- Technological infrastructure
  - 3- Human supervision

Beneficial outcomes

Improve patient outcomes	Identifying patients at risk and target interventions to prevent or treat them
Predicting hospital readmissions	Reduce healthcare costs

Shurouq A. Alowais et al., 2023

# Pros & Cons

## Disadvantages

Loss of jobs

Lack of human touch-  
empathy, emotional  
intelligence.

## Advantages

Efficiency, accuracy,  
precision

Decreased workload,  
Increase patient face time,  
Increase time on critical  
cases

Saves money, Better  
monitoring

Amisha et al., 20219

# Applications of artificial intelligence in health care



■ Drug development

■ Health monitoring

■ Managing medical data

■ Disease diagnostics

■ Digital consultation

■ Personalized treatment

■ Analysis of health plans

■ Surgical treatment

■ Medical teratment

[Amisha et al., 20219](#)

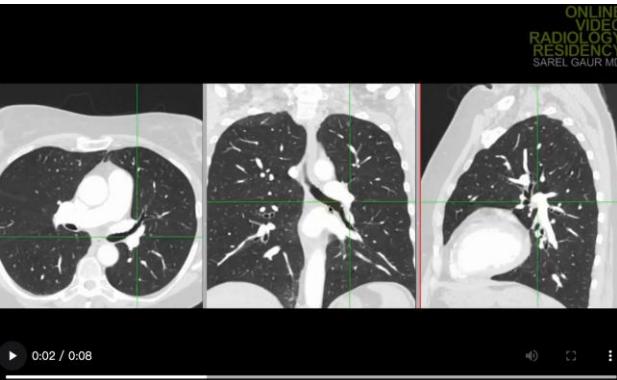
## Commonly used applications

- Application 1: High-resolution image understanding
- Application 2: Video understanding
- Application 3: **Pathological image understanding / 3D medical image understanding**
- Application 4: Multi-modal Agent for social simulation
- Application 5: Agent with longer multi-modal memory

# Towards a Medical World Simulator



Sora Medical Clinic  
*Soar Beyond*



Medical image analysis



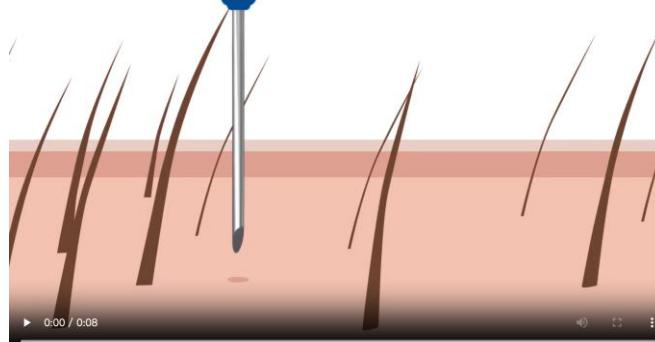
Medical care



Instrument Usage



Operation



Medical Popularization



Medical Scenario

# Chain of diagnosis

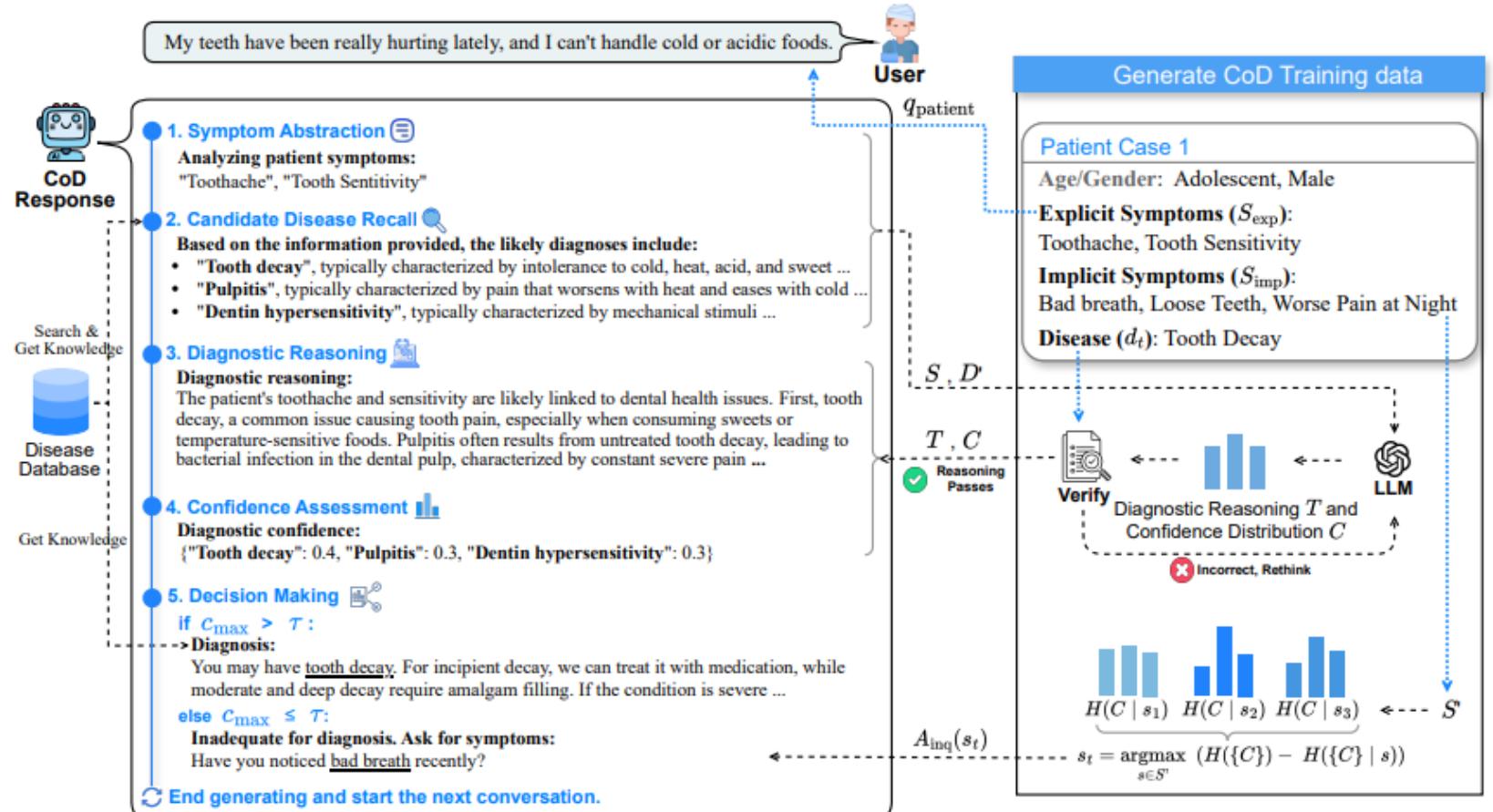
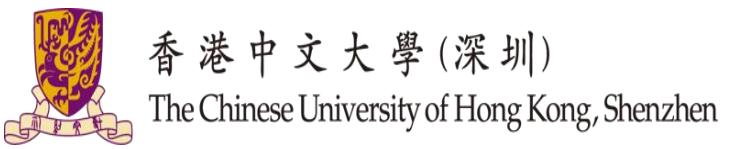


Figure 2: Left: Example of a CoD response. Right: Construction of CoD training data.

Junying Chen, Chi Gui, Anningzhe Gao, Ke Ji, Xidong Wang, Xiang Wan, Benyou Wang. CoD, Towards an Interpretable Medical Agent using Chain of Diagnosis.  
<https://arxiv.org/abs/2407.13301>

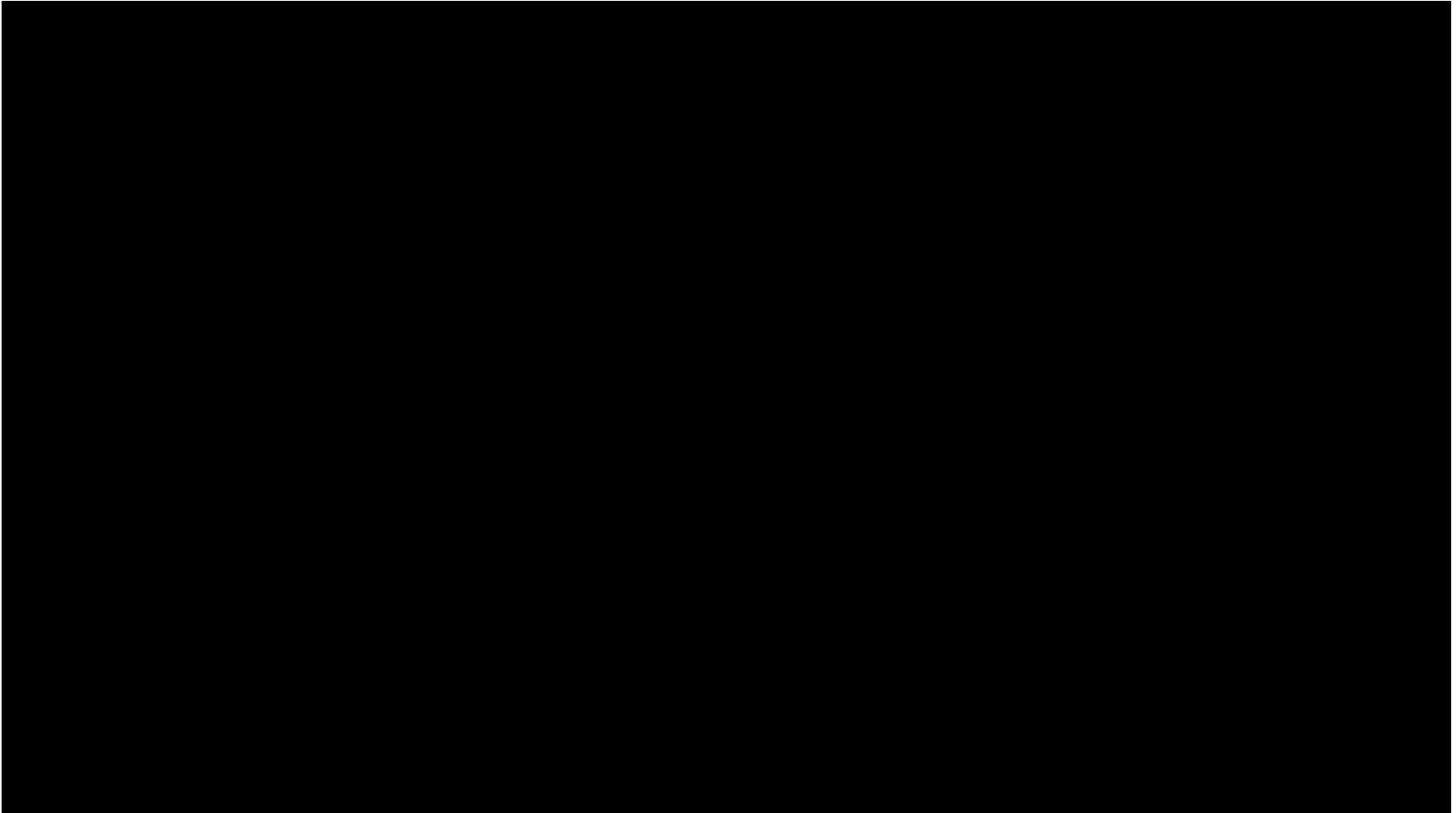
# Hauto GPT2 --- demo2



国内首个区域信息平台大模型智能导诊系统，实现了导诊、咨询、科普三大功能。

华佗GPT已在龙岗区12家医院上线使用，实现了龙岗区属公立医院全覆盖，智能导诊系统（包含导诊、客服咨询、健康咨询）已服务超过24万人次，交互次数38万，分诊准确率达95.2%。

## Hauto GPT2 --- demo2



# Outline

- Introduction of LLM
- Medical LLM
- AI for protein structure prediction
- AI for new drug discovery & development
- AI in medical diagnostics & healthcare
- LLM agent hospital

# LLM Agent Hospital

# Agent Hospital



**AIR**

清华大学 智能产业研究院  
Institute for AI Industry Research, Tsinghua University

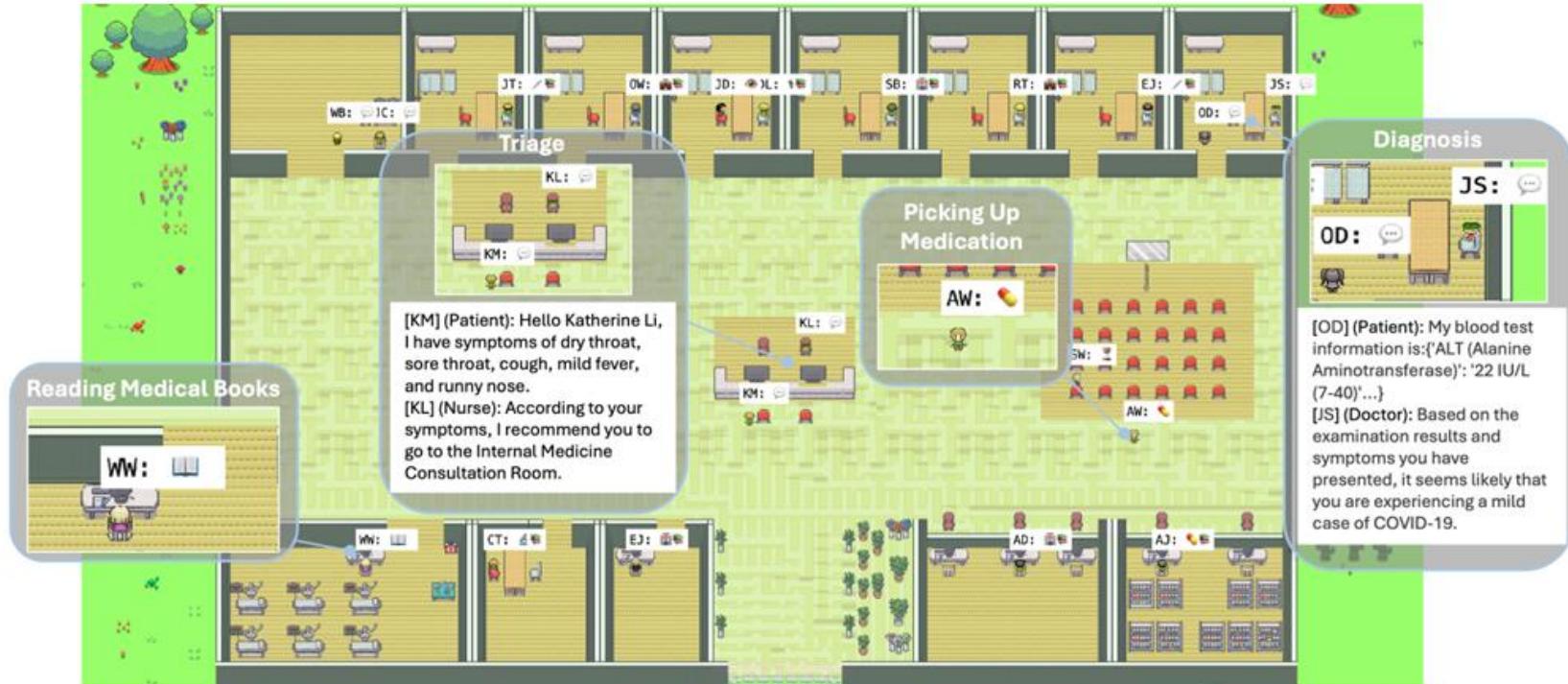


Fig. 1. An overview of Agent Hospital. It is a simulacrum of hospital in which patients, nurses, and doctors are autonomous agents powered by large language models. Agent Hospital simulates the whole closed cycle of treating a patient's illness: disease onset, triage, registration, consultation, medical examination, diagnosis, medicine dispensary, convalescence, and post-hospital follow-up visit. An interesting finding is that the doctor agents can keep improving treatment performance over time without manually labeled data, both in simulation and real-world evaluations.

JUNKAI LI et al., 2024

# Agent Hospital



AIR

清华大学 智能产业研究院  
Institute for AI Industry Research, Tsinghua University

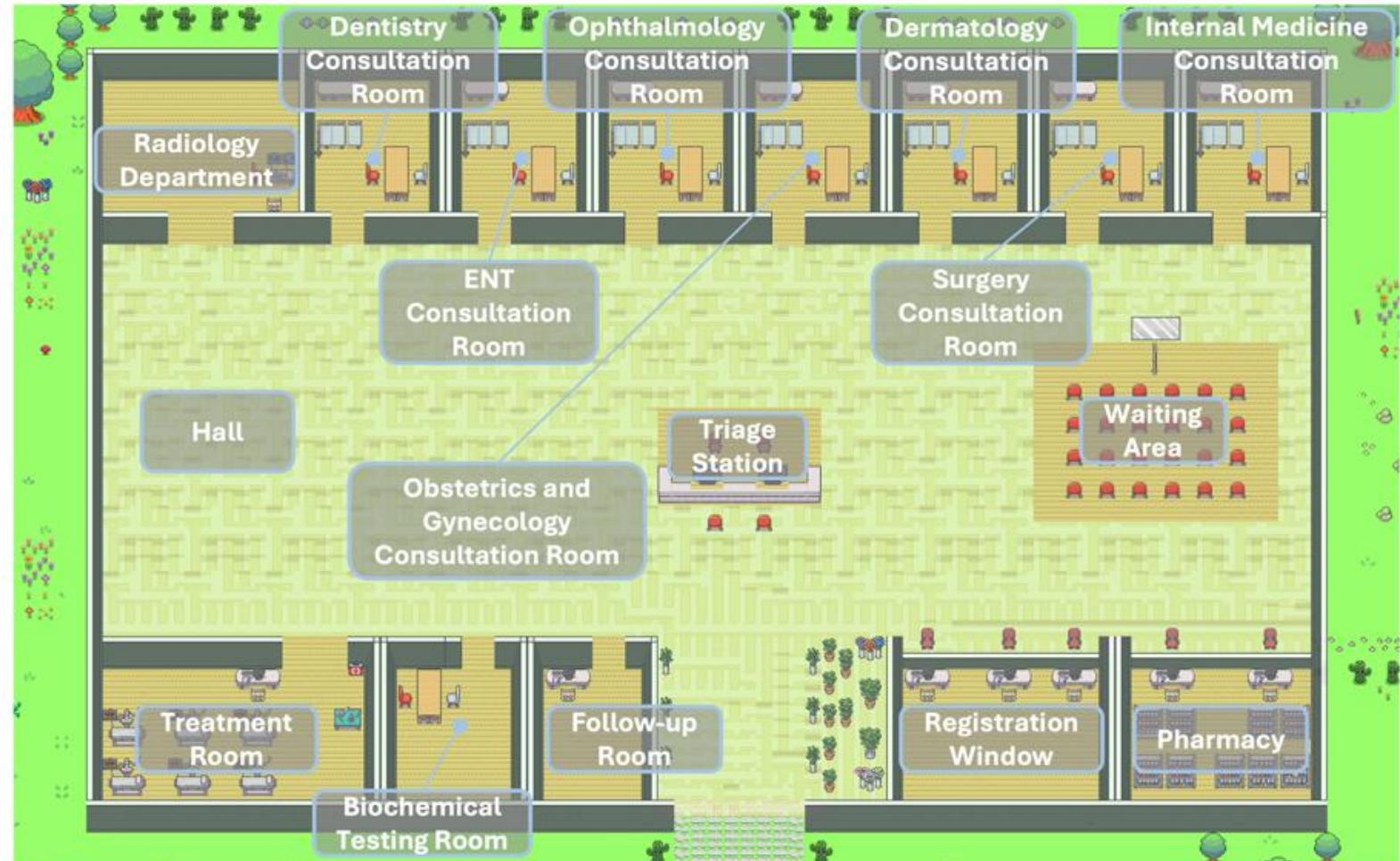


Fig. 2. The distribution of various areas within Agent Hospital.

JUNKAI LI et al., 2024

# Agent Hospital



**AIR**

清华大学 智能产业研究院  
Institute for AI Industry Research, Tsinghua University

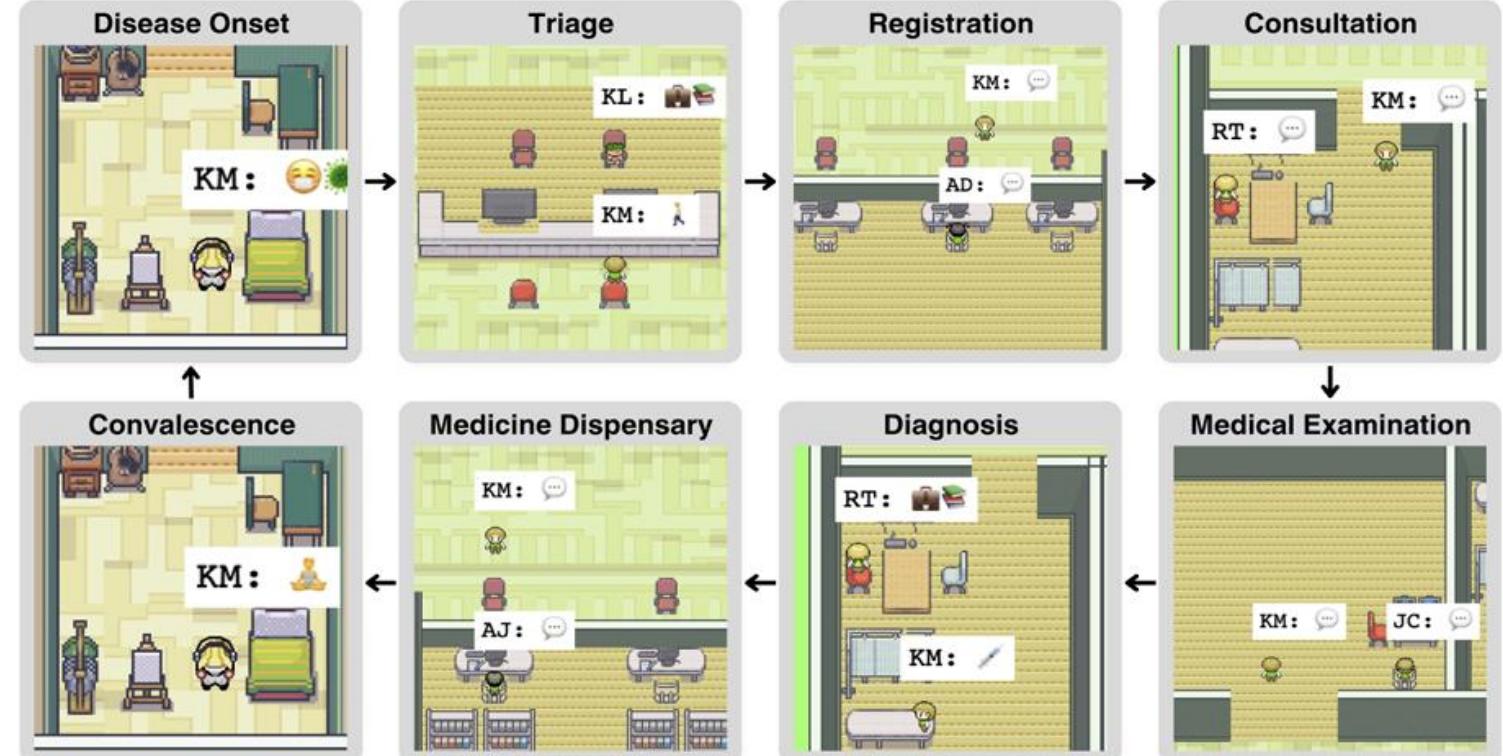


Fig. 4. Resident agent Kenneth Morgan falls ill and visits the Agent Hospital. Triage nurse Katherine Li conducts an initial evaluation of Mr. Morgan's symptoms and refers him to the dermatology department. Mr. Morgan then registers at the hospital's counter and is subsequently arranged for a consultation with dermatologist Robert Thompson. After undergoing the prescribed medical examination, Mr. Morgan receives a diagnosis and medication. He returns home to rest and monitor the improvement of his condition.

JUNKAI LI et al., 2024

# Agent Hospital


**AIR**

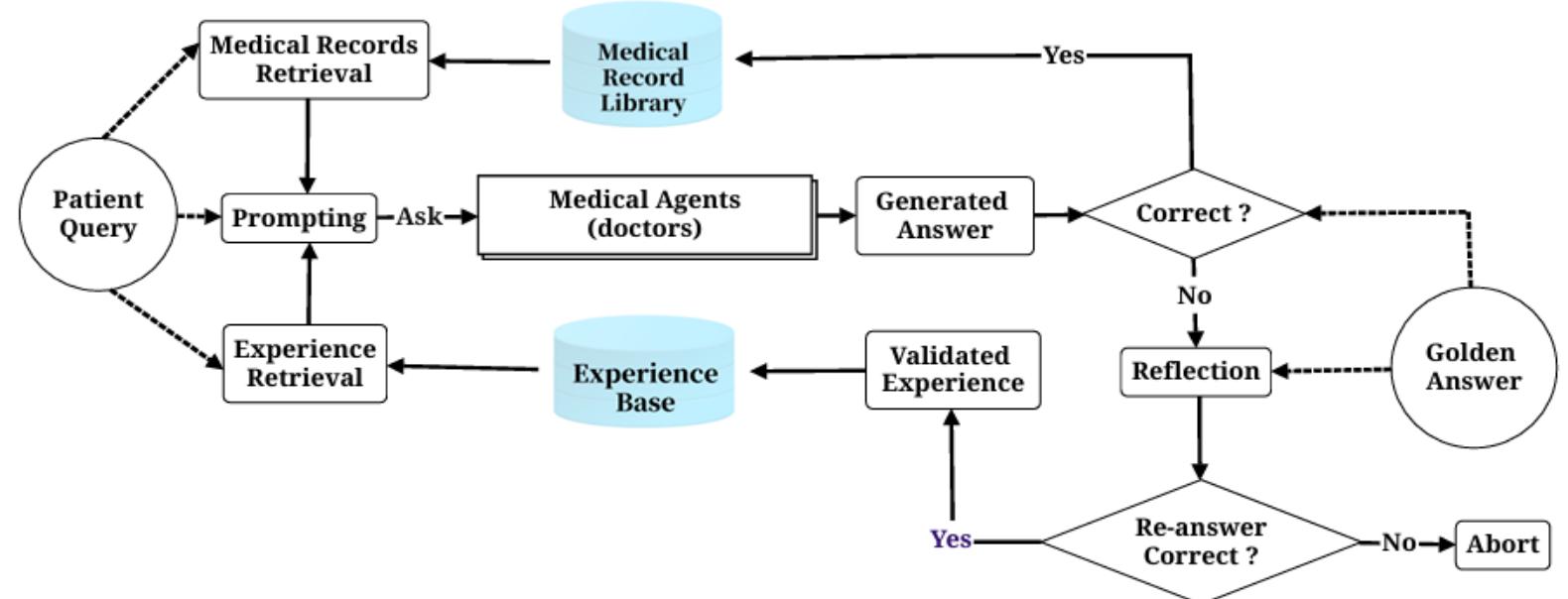
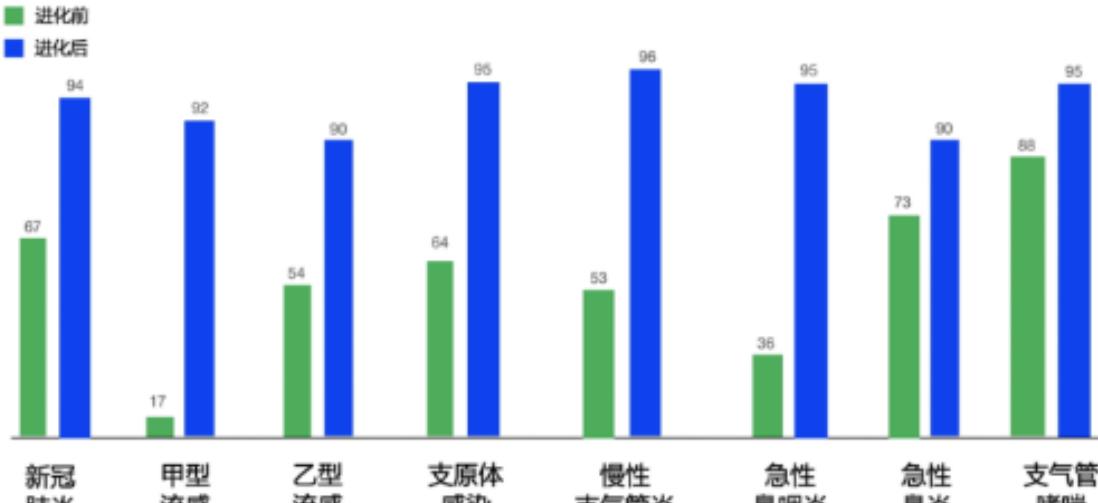
 清华大学 智能产业研究院  
 Institute for AI Industry Research, Tsinghua University


Fig. 5. The overview of the MedAgent-Zero method. This diagram illustrates the method by which **doctors achieve self-evolution**: 1) Accumulating examples and summarizing experience; 2) Adding correct responses directly to the example library; 3) Summarizing experience from incorrect responses and retesting them; 4) Incorporating successful experience into the experience library after further abstraction; 5) Utilizing both libraries to retrieve the most similar content for reasoning during the inference process.

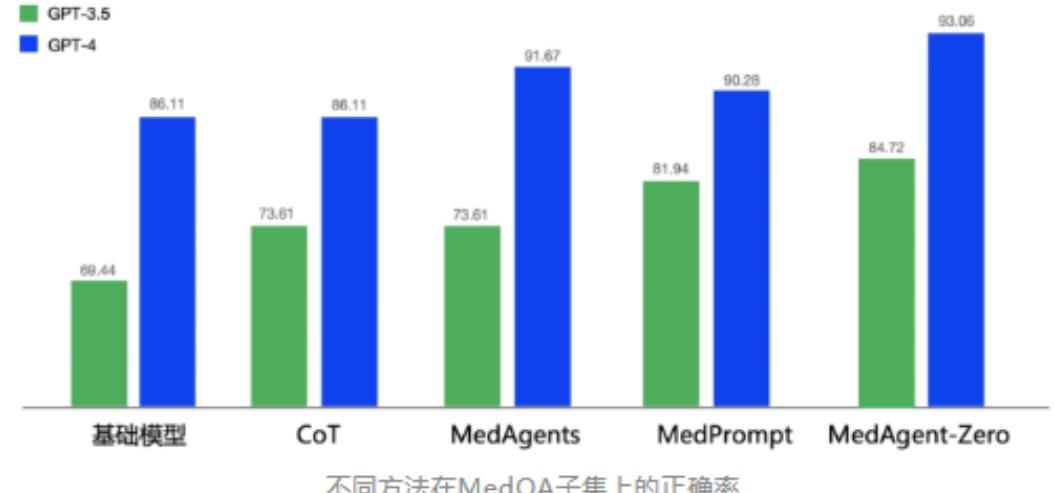
# Agent Hospital



清华大学 智能产业研究院  
Institute for AI Industry Research, Tsinghua University



智能体进化前后在不同疾病的诊断表现



# News

## AIR科研 | AIR打造虚拟医院，实现AI医生自我进化

来源:

发布时间: 2024-05-24

清华大学智能产业研究院 (AIR) 与清华大学计算机科学与技术系合作构建了**虚拟医院Agent Hospital**，提出了医学智能体自我进化方法MedAgent-Zero，通过在虚拟医院中产生大量无需人工标注的数据，让医学智能体不断提升医疗能力，并在真实世界数据集得到验证。Agent Hospital中所有的病人、护士和医生均由大模型驱动的自主智能体扮演，对发病、分诊、挂号、问诊、检查、诊断、开药、康复和随访的“院前-院中-院后”闭环流程进行模拟。Agent Hospital基于知识库与基础模型对虚拟病人的疾病产生与发展过程进行模拟。虚拟医生在Agent Hospital中进行学习（即阅读医学文献）与实践（即与虚拟病人交互并做诊疗决策），不断从成功诊疗案例中总结经验、从失败案例中反思教训，在多个诊疗任务上实现准确率持续提升。在诊疗近万名虚拟病人后（人类医生大约需要2年时间），虚拟医生能够在MedQA数据集呼吸道疾病子集上超越当前最好的方法，达到93.06%的准确率。该研究由AIR马为之助理研究员与AIR执行院长、计算机系副系主任刘洋教授担任论文共同通讯作者，在arXiv上公开后受到海内外人工智能社区和医学社区的广泛关注和讨论。

## AIR孵化 | Agent Hospital首批AI医生上线内测

来源:

发布时间: 2024-11-18

2024年5月，由清华大学智能产业研究院 (AIR) 智慧医疗团队完成的论文 "**Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents**" (共同通讯作者：AIR助理研究员马为之、AIR执行院长刘洋) 在国内外医学人工智能领域受到了广泛的关注和讨论，论文提出的 "Agent Hospital" (中文名称：智能体医院) 被媒体称为 "**清华医院小镇**"、"**清华AI医院**" 或 "**清华虚拟医院**"。

# MMedAgent Learning to Use Medical Tools with Multi-modal Agent

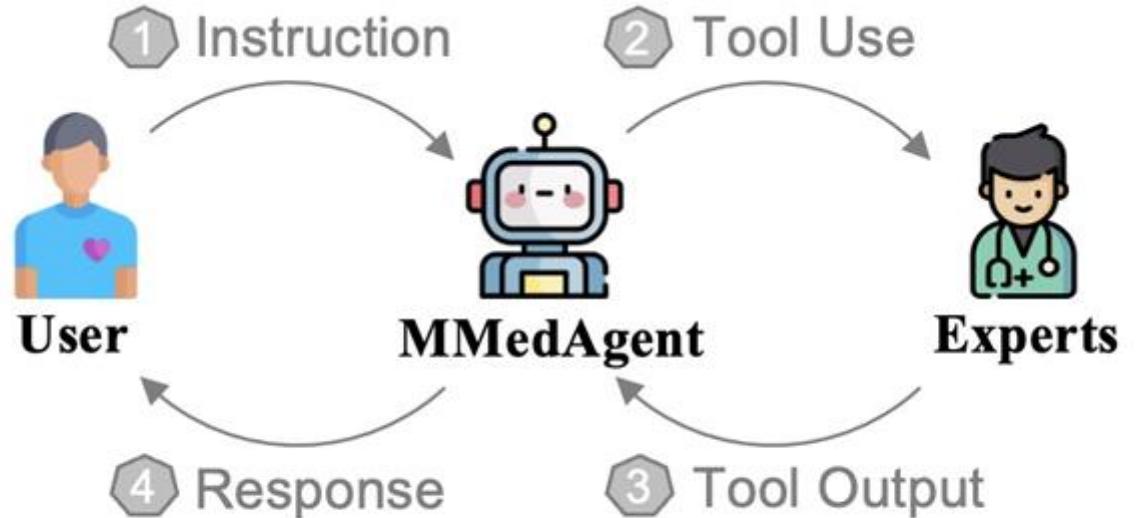


Figure 1: The four-step MMedAgent pipeline.

# MMedAgent

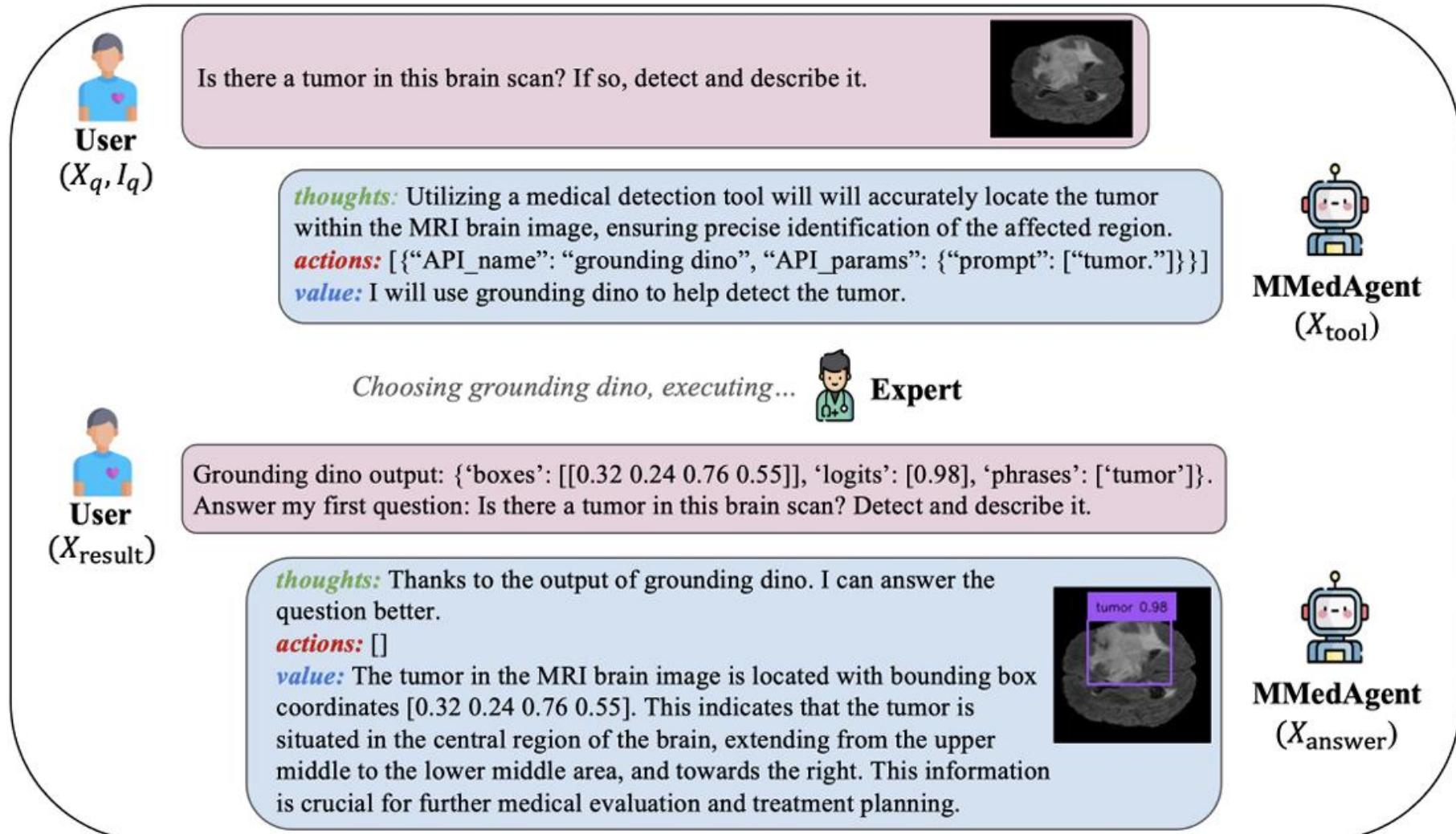


Figure 2: An example of the training data for MMedAgent that learns to use the tool of Grounding DINO for object detection and answer the user's question.

Binxu Li et al., 2024

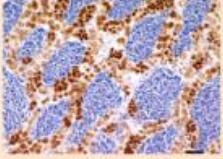
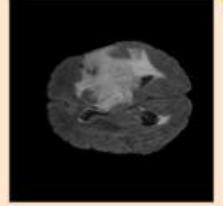
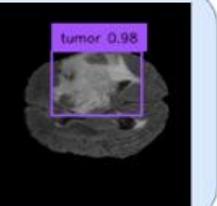
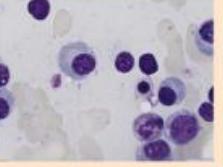
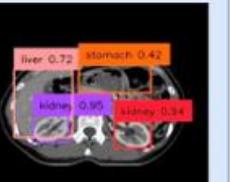
# MMedAgent



STANFORD

	Grounding			Cls.	MRG	RAG	Overall	Abs.
	Cell	Organ	Disease					
Flamingo-Med (Moor et al., 2023b)	13.11	15.87	15.33	23.56	16.59	-	14.68	1.16
RadFM (Wu et al., 2023)	-	-	-	25.00	68.13	-	45.38	3.59
LLaVA-Med (Li et al., 2023)	51.78	65.48	68.58	53.46	70.10	30.44	60.68	4.80
Yi-VL-34B (AI et al., 2024)	63.23	79.40	68.32	76.02	72.95	14.67	64.08	5.07
LLaVA-Med (Tool in Test)	45.32	52.77	67.91	57.53	74.34	67.55	65.31	5.17
Qwen-VL-Chat (Bai et al., 2023)	61.34	65.90	62.38	88.40	73.41	78.80	76.21	6.03
LLaVA-34B (Liu et al., 2023a)	76.75	84.85	80.75	<b>96.04</b>	80.27	<b>91.64</b>	86.52	6.84
<b>MMedAgent (ours)</b>	<b>97.50</b>	<b>102.29</b>	<b>125.89</b>	81.11	<b>121.49</b>	85.55	<b>109.48</b>	<b>8.66</b>

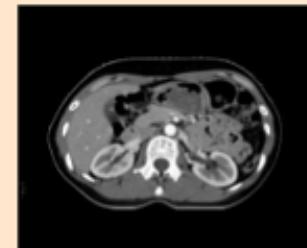
Table 2: Performance comparison between MMedAgent and other baselines. Cls. stands for classification. We report the relative scores for all tasks and the absolute (abs.) scores for overall performance in the last column. “-” indicates the tasks that the corresponding model is not applicable to. LLaVA-Med refers to the *60K-IM* version with only the initial query  $X_q$  and image  $I_q$  as input, while LLaVA-Med (Tool in Test) takes  $X_q$ ,  $I_q$  and also the internal output from tools  $X_{\text{result}}$  as input.

Task		User Input		LLaVa-Med Output		MMedAgent Output
Classification		What imaging modality is shown in this picture? 		The image shows a microscopic histological analysis of the defect site treated with different treatments. <b>The specific imaging modality is not mentioned</b> , but it is likely that the image was taken using a microscope to visualize the tissue samples.		The image is an example of a <b>immunohistochemistry histopathology image</b> .
Grounding (Disease)		Is there a tumor in this brain scan? Detect and describe it. 		The MRI image shows the presence of a tumor in the brain. The tumor appears to have a heterogeneous signal, meaning that it has areas with different intensities. It is important to note that <b>the interpretation of MRI images should be done by a trained professional</b> , such as a radiologist or ...		The tumor in the MRI brain image is <b>located with bounding box coordinates [0.32 0.24 0.76 0.55]</b> . This indicates that the tumor is situated <b>in the central region of the brain, extending from the upper middle to the lower middle area, and towards the right</b> . This information is crucial for further medical evaluation and treatment planning. 
Grounding (Cell)		Count and label all the cells in this histology image. 		As an AI language model, I cannot view the <b>actual image</b> . However, I can help you interpret the histology image by discussing the different cell types and structures that you may encounter. In general, histology images show tissue samples stained with specific dyes to highlight ...		Based on the detection results, the image <b>contains a total of 15 cells</b> . The bounding boxes for each cell are provided below, indicating their locations within the image: 1. Cell 1: [0.15, 0. , 0.25, 0.14] ... 15. Cell 15: [0.97, 0.53, 1. , 0.66] 
Grounding (Organ)		Can you identify the kidney, liver, and stomach in this abdominal CT image and describe them? 		Sure, in the 2D abdominal CT image, the kidneys are likely to appear as two large, bean-shaped structures with renal arteries and veins branching off from the aorta. The aorta is ... <b>To accurately segment these structures in the image, I would first identify the renal arteries and veins, then the aorta, and finally, the duodenum</b> .		Based on the detection results, here's a summary of the organ locations within the image: 1. The first kidney is located at the right middle part of the image, with bounding box coordinates [0.3, 0.54, 0.4, 0.62]. ... 4. The liver is on the middle left in the image, with bounding box coordinates [0.43, 0.39, 0.63, 0.47], it is larger than kidney in this slice. 

Binxu Li et al., 2024

## Grounding (Organ)

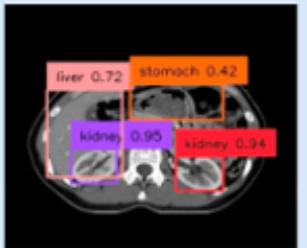
Can you identify the kidney, liver, and stomach in this abdominal CT image and describe them?



Sure, in the 2D abdominal CT image, the kidneys are likely to appear as two large, bean-shaped structures with renal arteries and veins branching off from the aorta. The aorta is ... To accurately segment these structures in the image, I would first identify the renal arteries and veins, then the aorta, and finally, the duodenum.

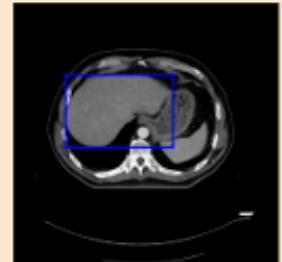
Based on the detection results, here's a summary of the organ locations within the image:

1. The first kidney is located at the right middle part of the image, with bounding box coordinates [0.3, 0.54, 0.4, 0.62]. ...
4. The liver is on the middle left in the image, with bounding box coordinates [0.43, 0.39, 0.63, 0.47], it is larger than kidney in this slice.



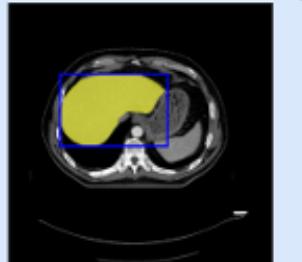
## Segmentation

Is it possible to use bounding boxes [0.2, 0.28, 0.6, 0.55] to segment the Liver in this medical image?



Yes, it is possible to use bounding boxes to segment the Liver in this medical image. Bounding boxes are rectangular regions that are drawn around the desired area of interest, in this case, the Liver. By providing the coordinates of the top-left and bottom-right corners of the bounding box, it is possible to train the AI system ...

The segmentation result is displayed below to assist with your medical evaluation:



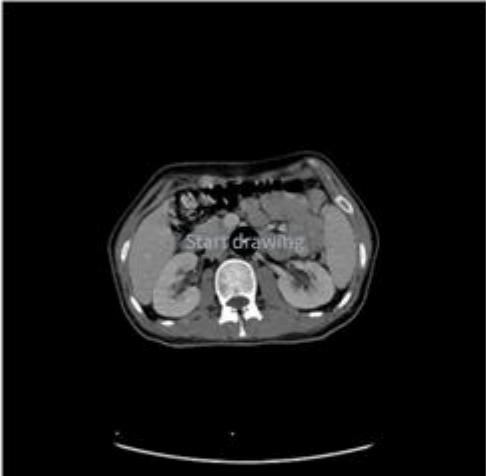
## MMedAgent: Learning to Use Medical Tools with Multi-modal Agent

[Paper] [Code]

MMedAgent\_llava\_V1

Current available tools:  
VQA, Grounding, Segmentation, Classification, Medical Report Generation, Retrieval Augmented Generation

Image



Reference Image

Parameters

OpenAI API Key

.....

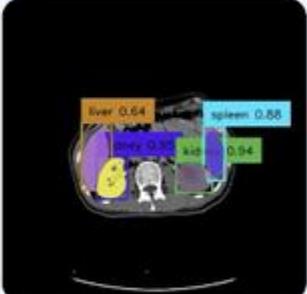
MMedAgent Chatbot

Can you locate and segment the kidneys, spleen, and liver in this 2D abdominal CT image?

Based on the bounding box coordinates and detection model output, here's a summary:

1. The first kidney is located towards the left middle of the image, with bounding box coordinates [0.31, 0.52, 0.41, 0.66].
2. The second kidney is on the right middle part of the image, with bounding box coordinates [0.57, 0.54, 0.69, 0.64].
3. The spleen is positioned towards the right middle of the image, with bounding box coordinates [0.66, 0.42, 0.73, 0.6].
4. The liver is located towards the left upper part of the image with bounding box coordinates [0.26, 0.41, 0.36, 0.63].

Based on the detection and segmentation model, the image segmentation result is as follows:



Can you locate and segment the kidneys, spleen, and liver in this 2D abdominal CT image?

Submit

Show Progress

Binxu Li et al., 2024

# Sources/Channels for Obtaining Relevant Materials/Resources



AI in Medicine | NEJM 是由《新英格兰医学杂志》(New England Journal of Medicine, NEJM) 推出的一个专注于人工智能(AI)在医学领域应用的平台。NEJM 是全球最权威的医学期刊之一，而 AI in Medicine 是其子平台，旨在探讨和推广人工智能技术在医疗健康领域的应用、研究和发展。



The screenshot shows a Coursera course page titled "用于医疗诊断的人工智能" (Artificial Intelligence for Medical Diagnosis). The page includes navigation links for "个人" (Personal), "商务" (Business), "大学" (University), and "政府" (Government). The course is part of the "医学人工智能 专项课程" (Medicine Artificial Intelligence Specialized Course). It features three teachers: Pranav Rajpurkar and two others. A blue button for "免费注册" (Free Registration) is visible, along with a note about a 7-day free trial and scholarships. A red arrow points from the text "NEJM 是全球最权威的医学期刊之一" in the first section to the "DeepLearning.AI" logo on the Coursera page.

# Thank You!