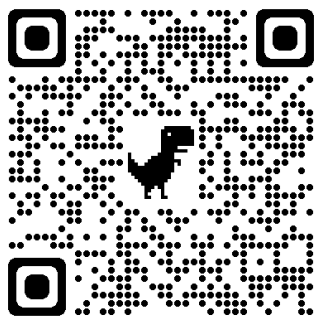


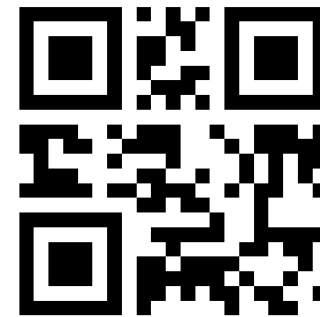
The ChatGLM's Road to AGI

Jie Tang

KEG, Tsinghua University



<https://github.com/THUDM/>



A Very Big Team



What is ChatGLM

- ChatGPT and GPT4 has gained enormous popularity
 - However, techniques behind GPT become a **secret** to all
- ChatGLM, an open-source ChatGPT alternative, toward unclosing the secrets
 - **GLM-130B**: an open-source LLM base model
 - **ChatGLM-6B**: a lightweight open-source ChatGPT alternative
 - **ChatGLM-130B&GLM-4**: not open-sourced, but available through API



GLM-130B

An Open Bilingual Pre-Trained Model



ChatGLM
Alpha

Open LLM Research

#star

39,369

15,516

12,186

7,788

7,618

7,203

5,110

ChatGLM-6B

Public

ChatGLM-6B: An Open Bilingual Dialogue Language Model | 开源双语对话语言模型

Python 39,025 stars Apache-2.0 license 5,068 forks 535 issues 42 discussions Updated on Feb 18



ChatGLM2-6B

Public

ChatGLM2-6B: An Open Bilingual Chat LLM | 开源双语对话语言模型

Python 15,424 stars Apache-2.0 license 1,824 forks 421 issues 27 discussions Updated yesterday



ChatGLM3

Public

ChatGLM3 series: Open Bilingual Chat LLMs | 开源双语对话语言模型

Python 11,687 stars Apache-2.0 license 1,336 forks 11 (1 issue needs help) issues 1 discussion Updated 2 days ago



CodeGeeX

Public

CodeGeeX: An Open Multilingual Code Generation Model (KDD 2023)

Python 7,729 stars Apache-2.0 license 551 forks 147 issues 7 discussions Updated last week



GLM-130B

Public

GLM-130B: An Open Bilingual Pre-Trained Model (ICLR 2023)

Python 7,600 stars Apache-2.0 license 609 forks 115 issues 5 discussions Updated on Jul 25, 2023



CodeGeeX2

Public

CodeGeeX2: A More Powerful Multilingual Code Generation Model

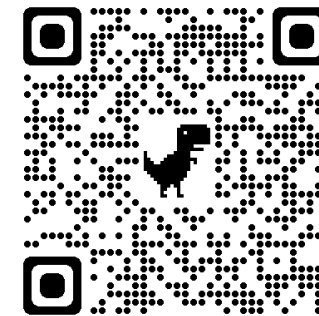
Python 7,008 stars Apache-2.0 license 486 forks 180 issues 2 discussions Updated on Feb 6



CogVLM

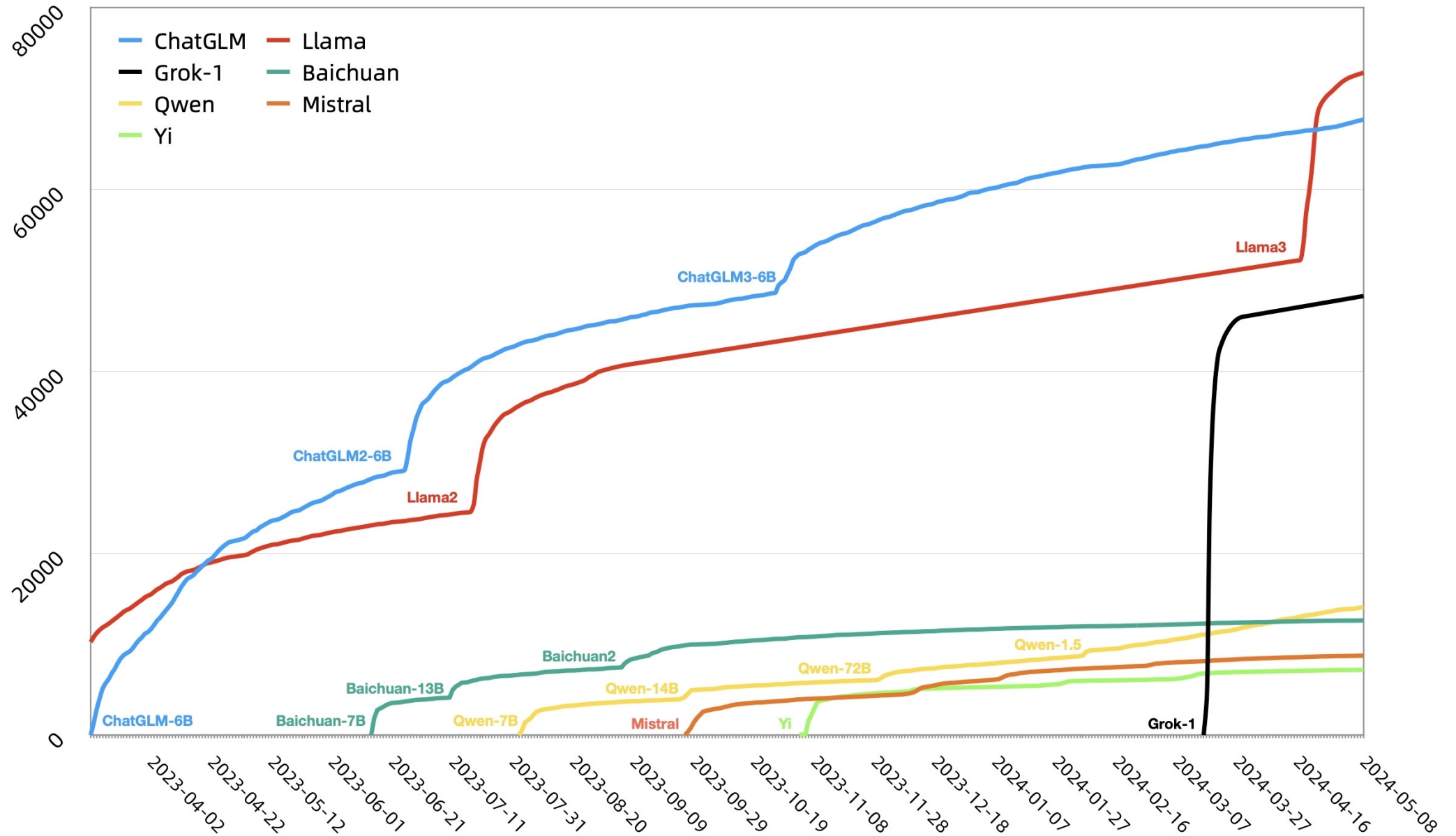
Public

a state-of-the-art-level open visual language model | 多模态预训练模型

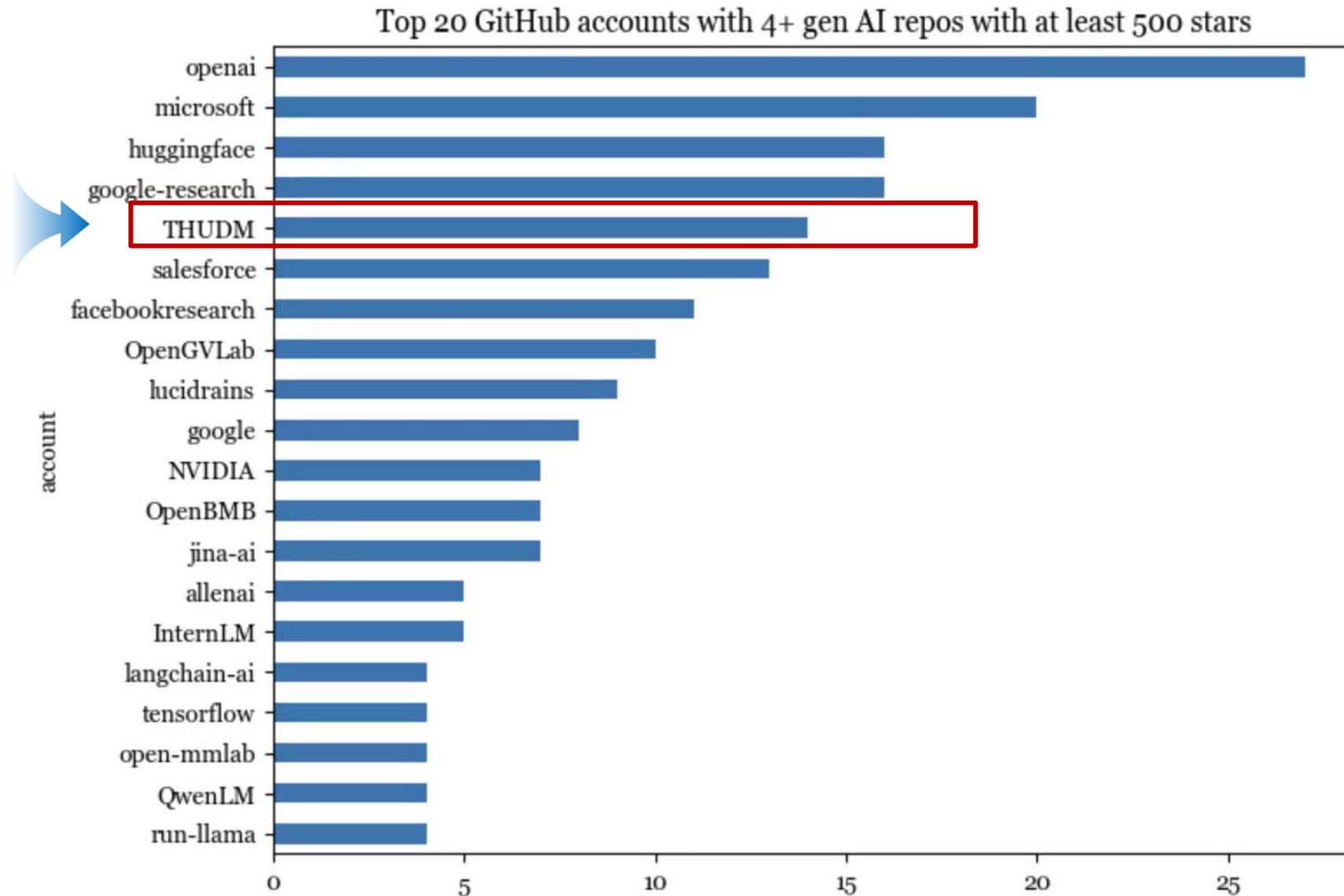


<https://github.com/THUDM>

ChatGLM vs. LLaMA



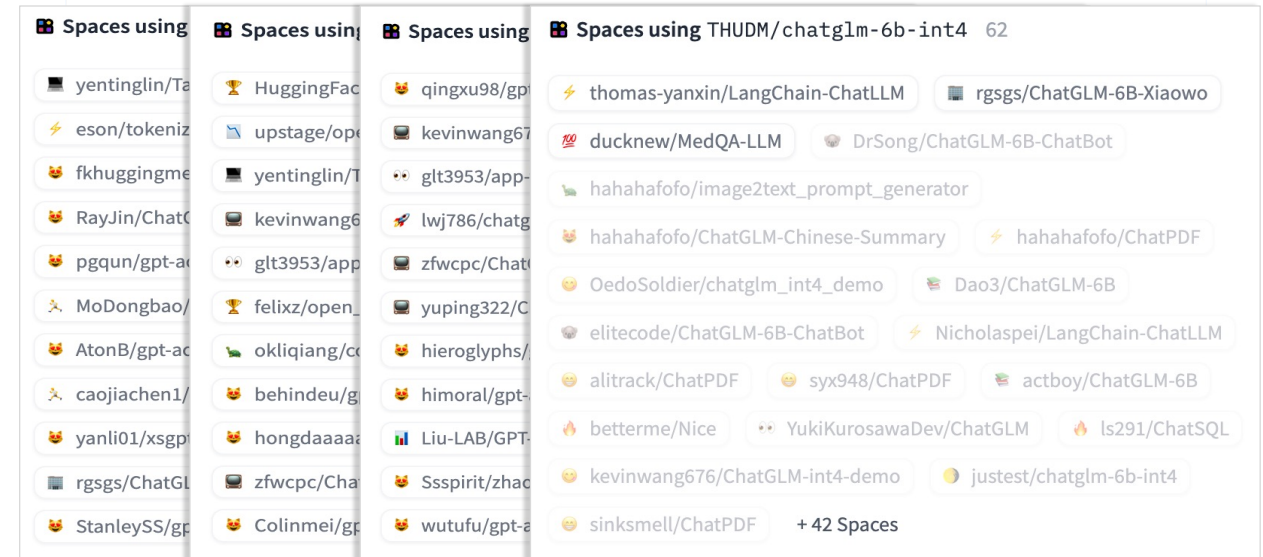
Number of Projects with 500 stars



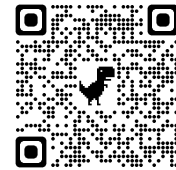
ChatGLM-6B: An Open-Source Alternative

- ▶ ChatGLM-6B: **6.2B** parameters, **INT4** quantization (only need 6G memory)
- ▶ **>60,000 stars** on github
- ▶ **>13,000,000** downloads on Huggingface
- ▶ **No. 1** on Github Trending (2 week)
- ▶ **No. 1** on Huggingface Trending (2 weeks)

>700 Open-Sourced Apps developed based on ChatGLM



GLM-130B
An Open Bilingual Pre-Trained Model



ChatGLM
Alpha

The top 15 most-liked organizations on @huggingface

01. [@StabilityAI](#) 20k likes

02. [@AlatMeta](#) 20k

03. [@runwayml](#) 11k

04. [Compvis](#) 10k

05. [@thukeyg](#) 7k

06. [@BigscienceW](#) 7k

07. [@Tlluae](#) 7k

08. [@Microsoft](#) 6.5k

09. [@GoogleAI](#) 6k

10. [@OpenAI](#) 4k

11. [@BigCodeProject](#) 4k

12. [@MosaicML](#) 4k

13. [@UKPLab](#) 3k

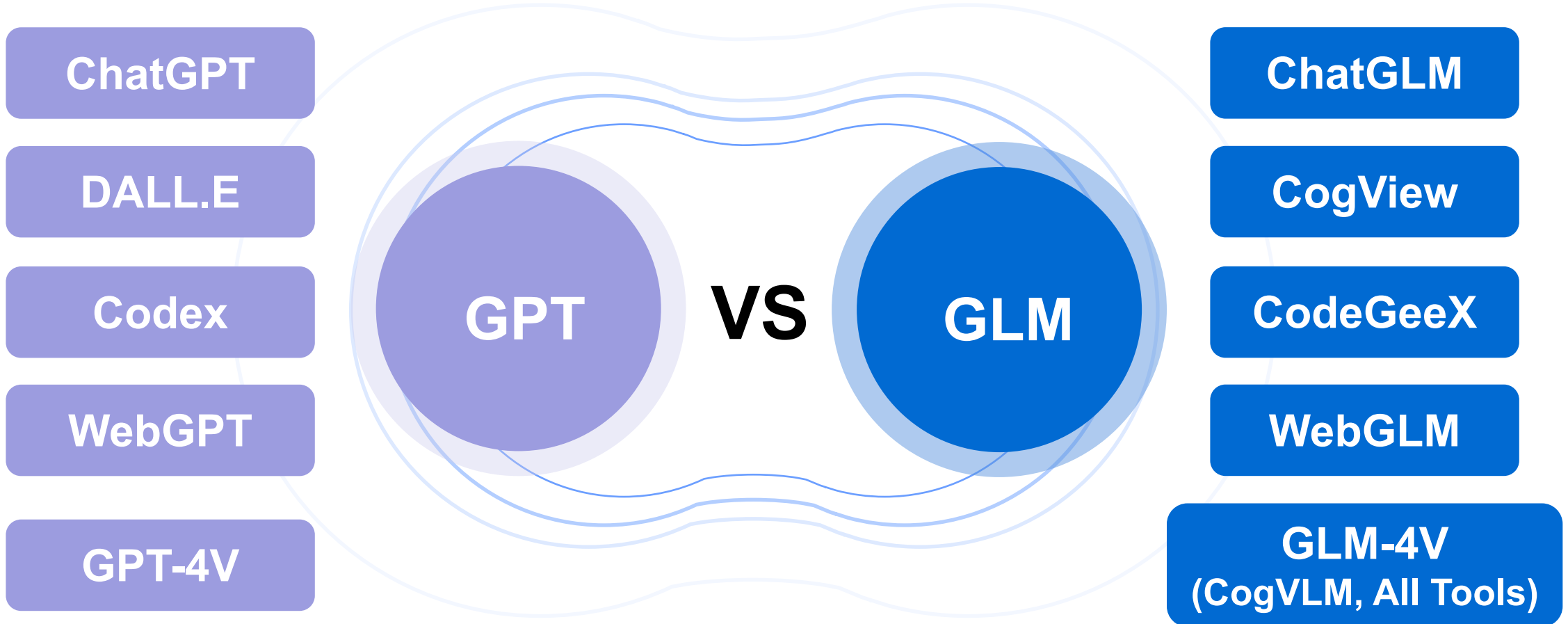
14. [@AiEleuther](#) 3k

15. [@salesforce](#) 3k

huggingface.co/spaces/PulsarA...

5:40 PM - Nov 22,2023 - **6,094** views

ChatGPT vs. ChatGLM



chatglm.ai

GLM

XDAI

GLM-130B

CodeGeeX

QAGLM

ChatGLM

who are you?

I am an AI assistant named ChatGLM, which is developed based on the language model jointly trained by Tsinghua University KEG Lab and Zhipu AI Company in 2023. My job is to provide appropriate answers and support to users' questions and requests.

👍 🗨️ 📄 Copy

help me to write a blog outline which introduces ChatGLM

Sure, here's a blog outline that introduces ChatGLM:

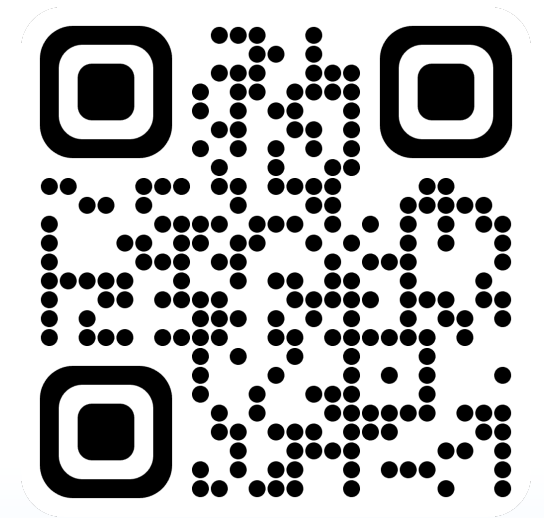
I. Introduction

A. Brief overview of ChatGLM

B. Importance of AI assistants in today's world

II. How ChatGLM works

A. Basics of AI language models



Welcome to try

Math Exercise



The apple orchard produces 75 tons of apples every year, with a fixed income of 8 yuan per kilogram and a cost of 3 yuan per kilogram. In order to improve land utilization, it has been decided to plant purple sweet potatoes under apple trees next year (the maximum yield is 50,000 kilograms) with a fixed income of 4 yuan per kilogram and a cost of 2 yuan per kilogram. When the yield of purple sweet potato is greater than 30,000 kilograms, every n kilogram increase in its yield will cause the apple yield to decrease by $0.3n$ kilograms. What is the maximum profit of the apple orchard in the next year?



To calculate the maximum profit of the apple orchard in the next year, we need to find the combination of apple and purple sweet potato yields that maximizes the total profit for the orchard. Let's define the variables:

$$R = (69,000) * (8 - 3) + 4P - 2P$$

445,000 yuan

This is the maximum profit under the constraints. Therefore, the maximum profit of the apple orchard in the next year is 445,000 yuan, which is achieved by producing 69,000 kg of apples and 50,000 kg of purple sweet potatoes.



您

The apple orchard produces 75 tons of apples every year, with a fixed income of 8 yuan per kilogram and a cost of 3 yuan per kilogram. In order to improve land utilization, it has been decided to plant purple sweet potatoes under apple trees next year (the maximum yield is 50,000 kilograms) with a fixed income of 4 yuan per kilogram and a cost of 2 yuan per kilogram. When the yield of purple sweet potato is greater than 30,000 kilograms, every n kilogram increase in its yield will cause the apple yield to decrease by $0.3n$ kilograms. What is the maximum profit of the apple orchard in the next year?



ChatGPT

To find the maximum profit of the apple orchard in the next year, we need to find the combination of apple and purple sweet potato yields that maximizes the total profit for the orchard. Let's define the variables:

Since this equation is not linear, we need to find the maximum profit under the constraints: $P = 0$ or $P = 50,000$.

1. When $P = 0$:

$$\text{Profit} = 45,375 + 0.5(0) = 45,375$$

2. When $P = 50,000$:

$$\text{Profit} = 45,375 + 0.5(50,000) = 70,375$$

PB

The apple orchard produces 75 tons of apples every year, with a fixed income of 8 yuan per kilogram and a cost of 3 yuan per kilogram. In order to improve land utilization, it has been decided to plant purple sweet potatoes under apple trees next year (the maximum yield is 50,000 kilograms) with a fixed income of 4 yuan per kilogram and a cost of 2 yuan per kilogram. When the yield of purple sweet potato is greater than 30,000 kilograms, every n kilogram increase in its yield will cause the apple yield to decrease by $0.3n$ kilograms. What is the maximum profit of the apple orchard in the next year?

To solve this problem, we need to find the combination of apple and purple sweet potato yields that maximizes the total profit for the orchard. Let's define the variables:

.....

Therefore, the maximum profit for the apple orchard in the next year is 475,000 yuan, which is achieved by producing 75,000 kg of apples and 50,000 kg of purple sweet potatoes.

Copy

Applied Math



Can you please prove $e^x = x^e$ has only and only has one solution?



The equation $e^x = x^e$ doesn't have a
understand the number of solutions

First, let's consider the function $f(x)$
analyze this function, we can use ca
the function.

1. Differentiate $f(x)$ with respect to
2. Find the critical points by setting
3. Analyze the behavior of the func
solutions.

Let's perform these steps.

Generating codes...

python

```
1 from sympy import symbols, diff, exp
2
3 # Define the variable and the function
4 x = symbols('x')
5 e = exp(1) # Define e as the base of the natural logarithm
6 f = exp(x) - x**e
7
8 # Calculate the derivative
9 f_prime = diff(f, x)
10 f_prime_simplified = f_prime.simplify()
11 f_prime_simplified
```

Results

$[(1 - e) \cdot \text{LambertW}(-\exp(-\log(e))/(e - 1))/(e - 1)]$

The critical point of the function $f(x) = e^x - x^e$ is given by $x = (1 - e) \cdot \text{LambertW}\left(-\frac{e^{-\frac{\log(e)}{e-1}}}{e-1}\right)$, where

Text-to-Image Generation (2048x2048)



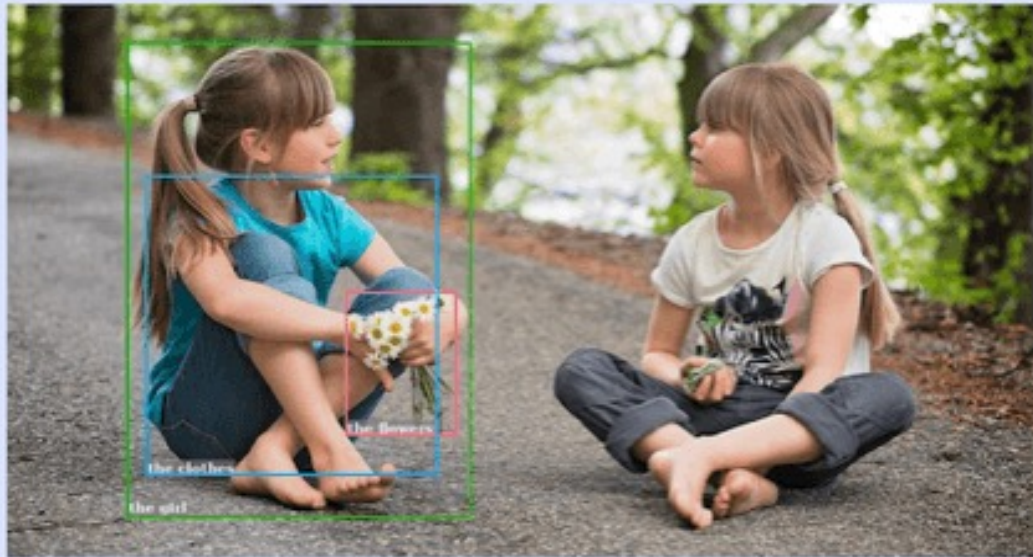
Fig. 1: Showcases of CogView3 generation of resolution 2048×2048 (top) and 1024×1024 (bottom). All prompts are sampled from Partiprompts [31].

Image Understanding

Grounding Visual Question Answering

Q1: What color are the clothes of the girl whose hands are holding flowers? Let's think step by step.

CogVLM: Find the flowers [[328,519,436,780]]. Check the girl holding it, got the girl [[116,069,450,927]]. Check the clothes of it, got the clothes [[134,309,418,852]]. The clothes [[134,309,418,852]] are blue. So the answer is blue.



Q2: What color are the jeans of the girl whose hands are not holding flowers? Let's think step by step.

CogVLM: Find the not(flowers) hands [[654,637,710,728]]. Check the girl with it, got the girl [[526,109,880,946]]. Check the jeans of it, got the jeans [[530,613,880,871]]. The jeans [[530,613,880,871]] are gray. So the answer is gray.



GUI Agents

“Change my Phone’s display mode to light mode”

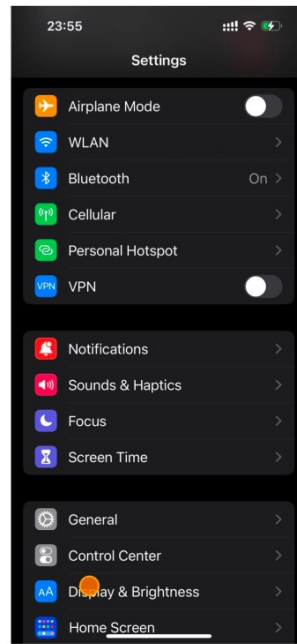


User: Could you guide me through the steps to "change my phone's display mode to light mode"?

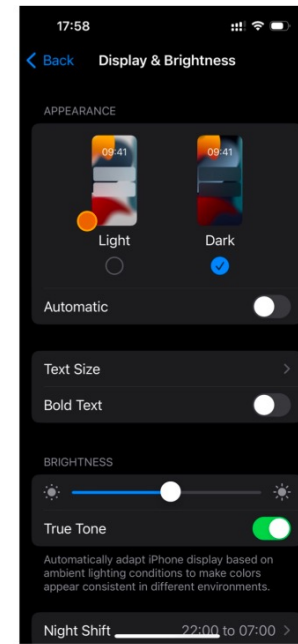
Smartphone Agent



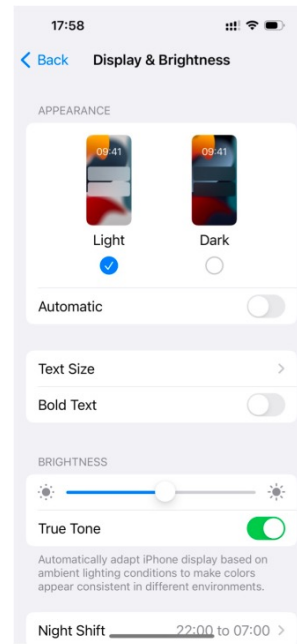
CogAgent: Touch the area on the screen where the 'Settings' icon is located. This action aims to open the settings menu where you can proceed with changing the phone's display mode to light mode.
[Operation]:
tap [x,y]=[0.387,0.625]



CogAgent: Touch the area of the screen where the 'Display & Brightness' text is located. This action aims to open the display settings where you can proceed with changing the light mode.
[Operation]:
tap [x,y]=[0.29,0.895]



CogAgent: Tap the 'Light' option for display mode, which is currently highlighted and located at the coordinates [x,y]=[0.156,0.349] on the screen. This will select 'Light' mode and activate it.
[Operation]:
tap [x,y]=[0.176,0.349]

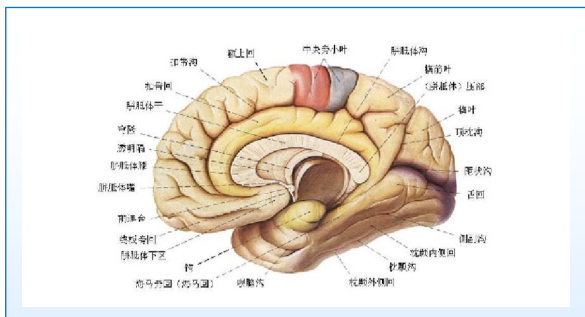




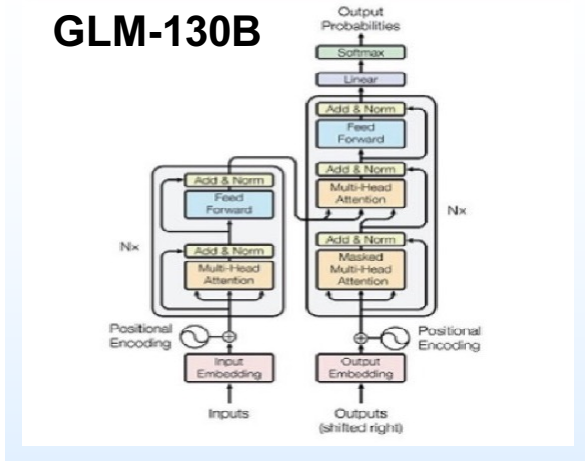
How to make this happen?

—let us open the secrets

Our Plan since 2019



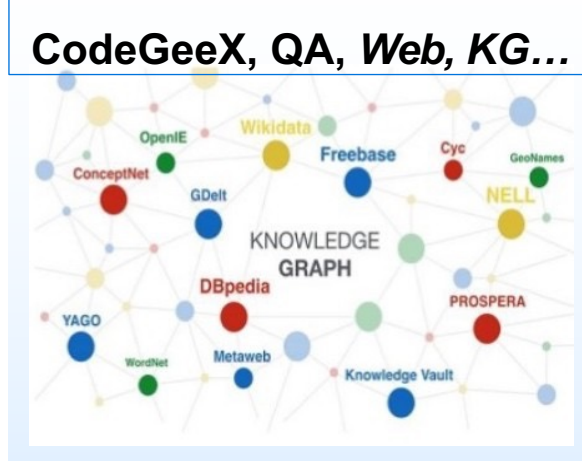
LLM



System 1

Short-term Memory
Long-term Memory

Memory (Knowledge)

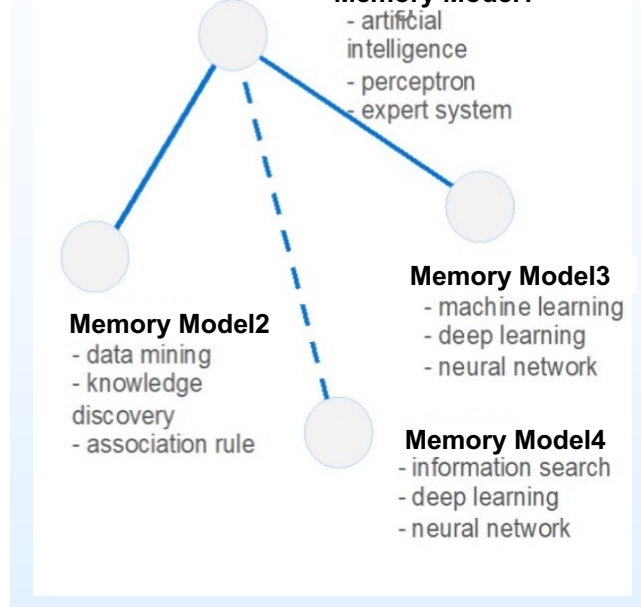


System 2

Cognition

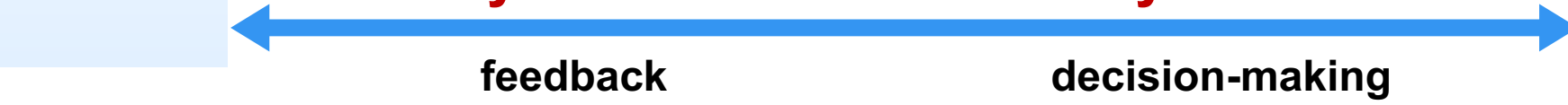
“Unconscious” learning
Reasoning

Self-instruct



ChatGLM

Real World



THU & ZhipuAI's GLM

GLM

Du et al. **GLM**: General Language Model Pretraining with Autoregressive Blank Infilling. arXiv:2103.10360. ACL'22

P-Tuning

Liu et al. **GPT understands, Too**. arXiv:2103.10385. ACL'22

CogView

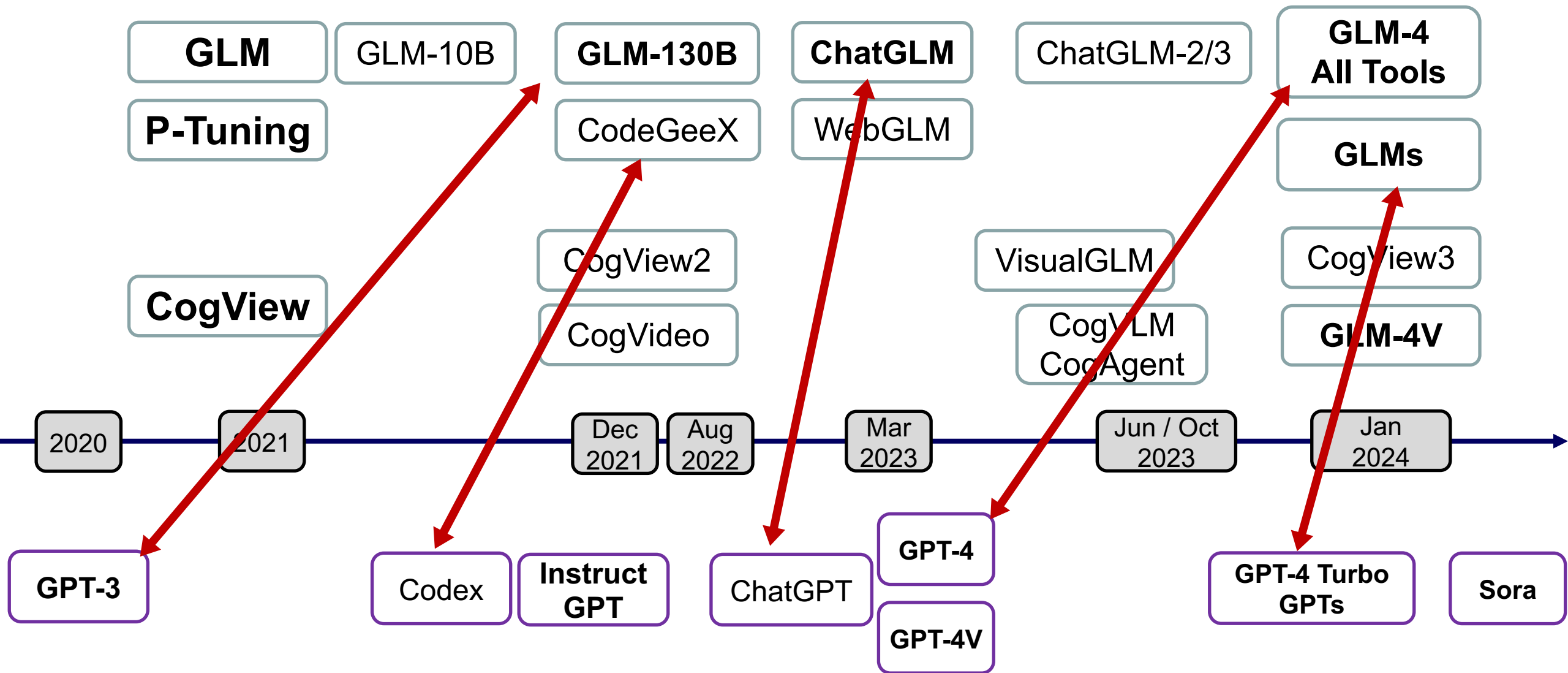
Ding et al. **CogView**: Mastering Text-to-Image Generation via Transformers. arXiv:2105.13290. NeurIPS 2021

2020

2021

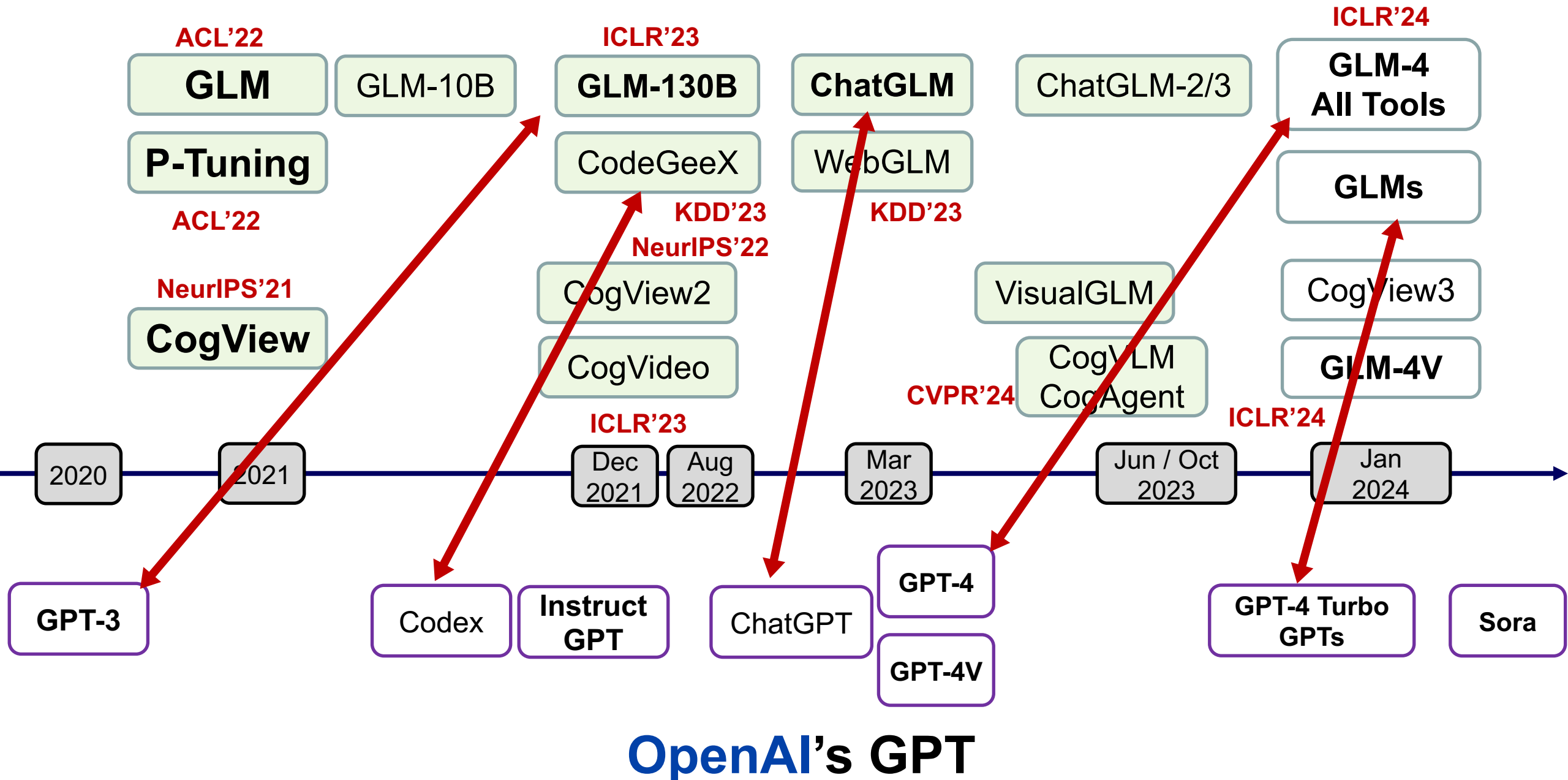


THU & ZhipuAI's GLM



OpenAI's GPT

THU & ZhipuAI's GLM



General Language Model (GLM)

Framework	NLU	Cond. Gen.	Uncond. Gen.
Autoregressive (GPT)	—	—	√
Autoencoding (BERT)	√	×	×
Encoder-Decoder (T5)	—	√	—
Autoregressive Blank-Infilling (GLM)	√	√	√

General Language Model (GLM)

x_1 x_2 x_3 x_4 x_5 x_6

(a) Sample spans from the input text

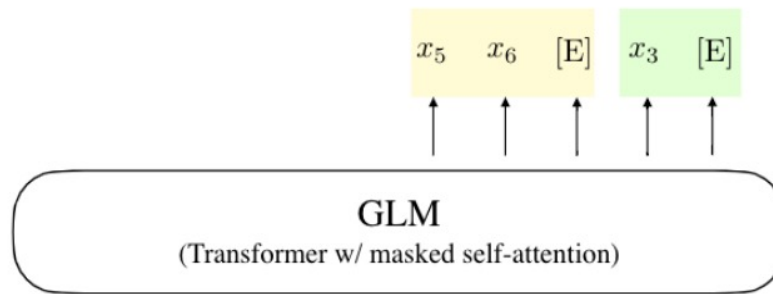
Part A:

x_1 x_2 [M] x_4 [M]

Part B:

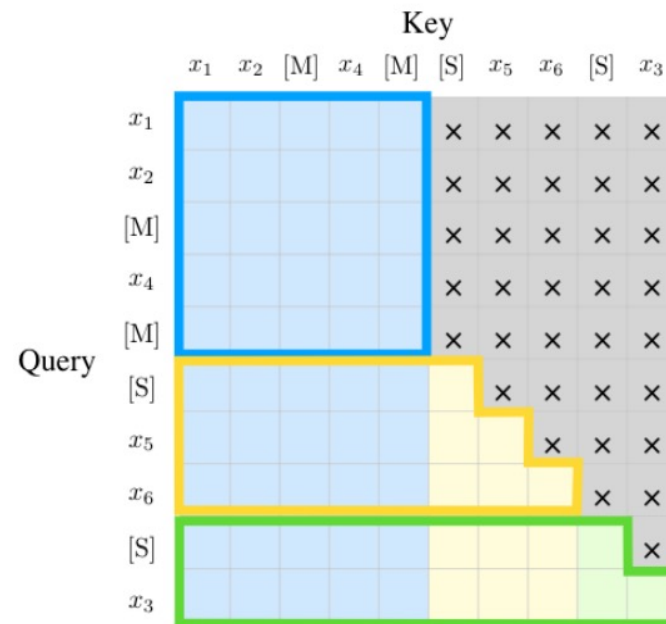
x_5 x_6 x_3

(b) Divide the input into Part A and Part B



Position 1 1 2 3 4 5 5 5 5 3 3
Position 2 0 0 0 0 0 1 2 3 1 2

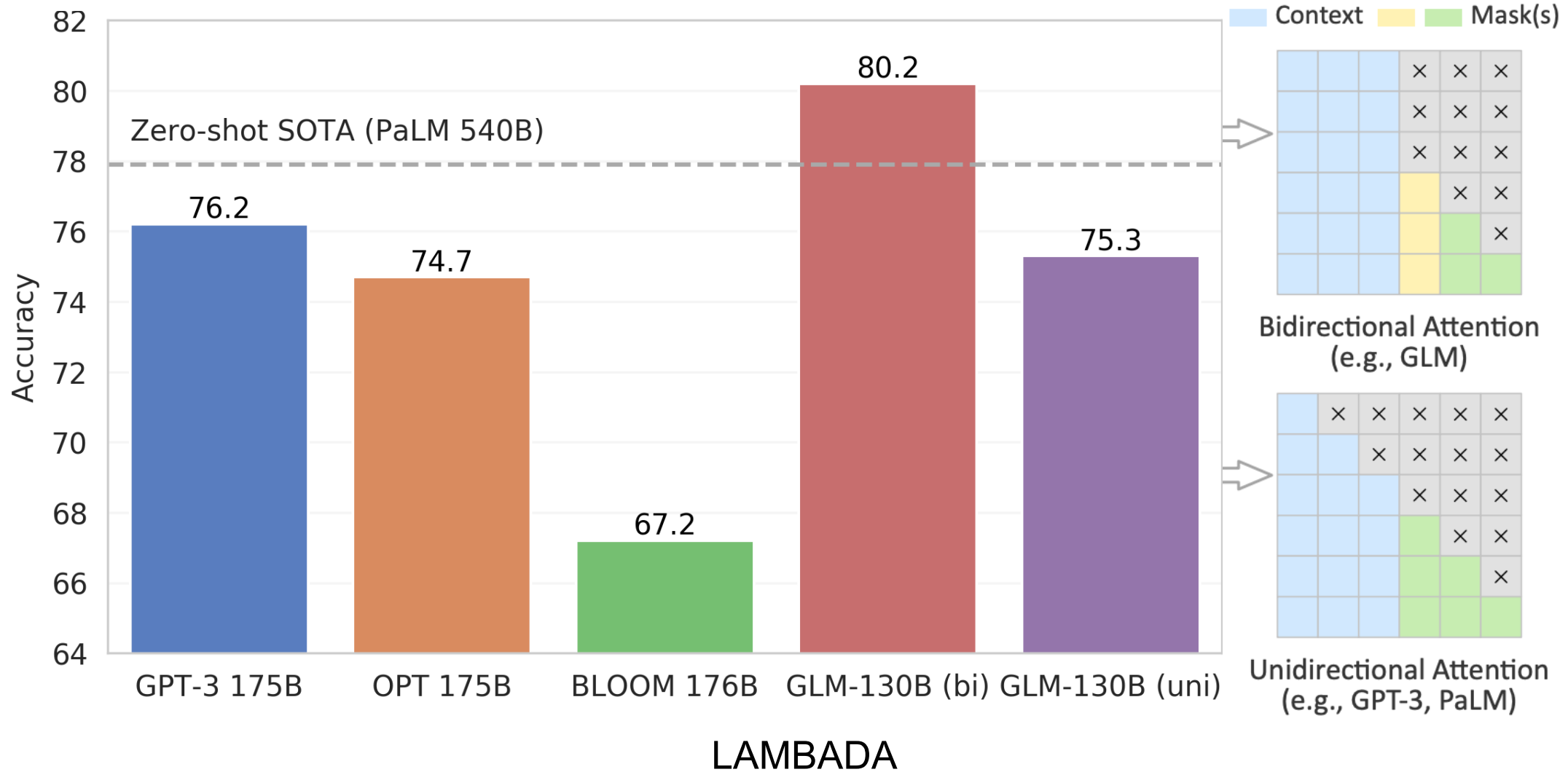
(c) Generate the Part B spans autoregressively



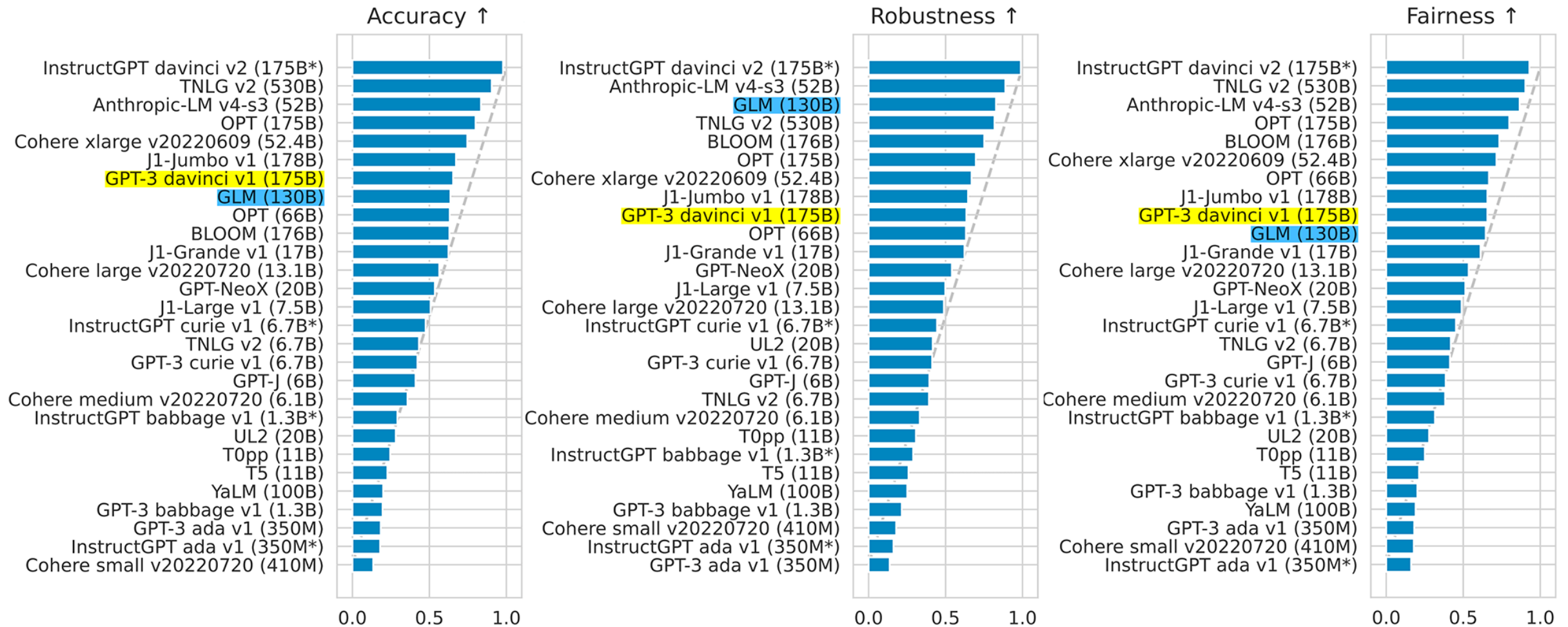
(d) Self-attention mask

$$\mathcal{L}_{\text{GLM}} = \mathbb{E}_{\mathbf{z} \sim Z_m} \left[\sum_{i=1}^m \sum_{j=1}^{l_i} -\log p(s_{z_i, j} | \mathbf{x}_{\text{corrupt}}, \mathbf{s}_{z_{<i}}, \mathbf{s}_{z_i, <j}) \right]$$

General Language Model (GLM)

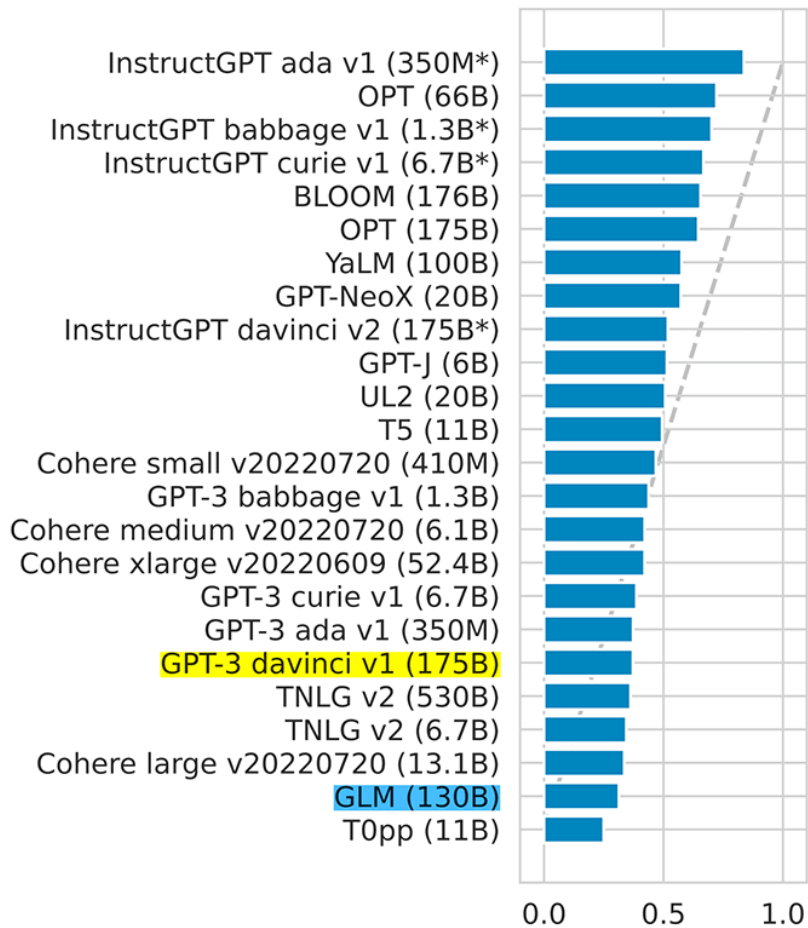


GLM-130B in HELM

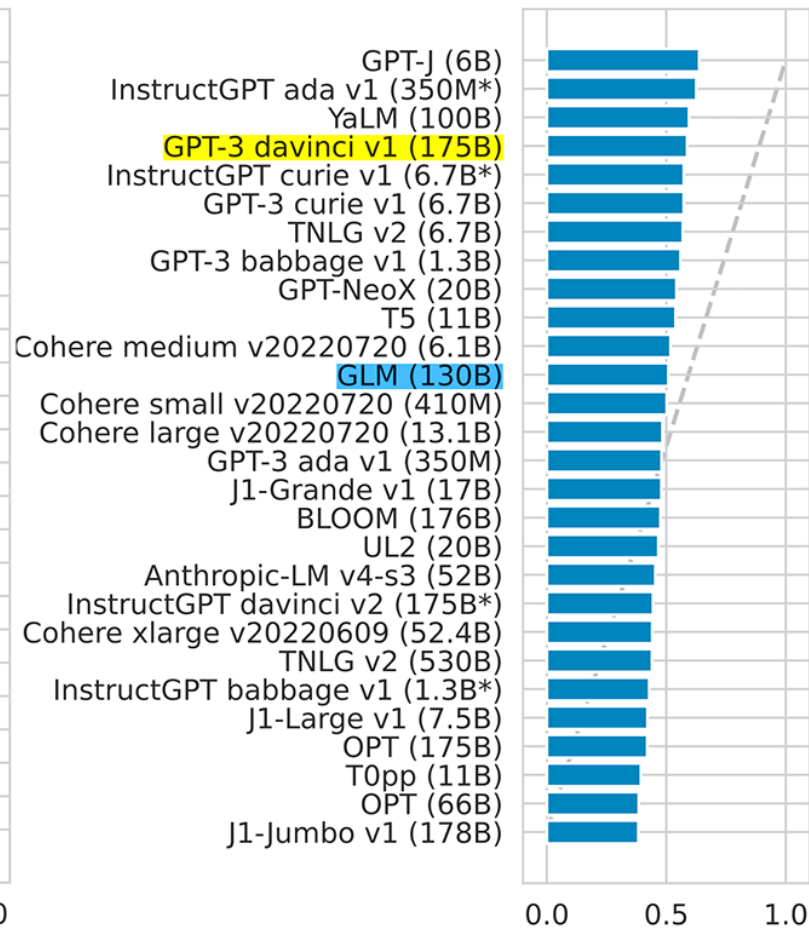


GLM-130B in HELM

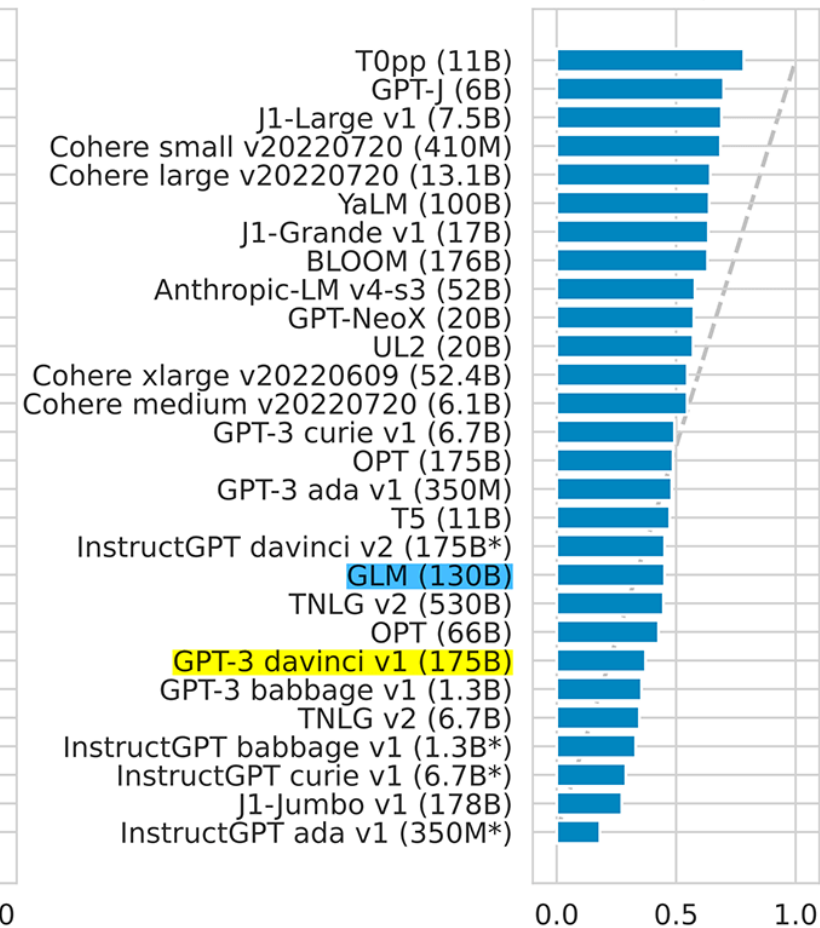
Calibration error ↓



Bias ↓

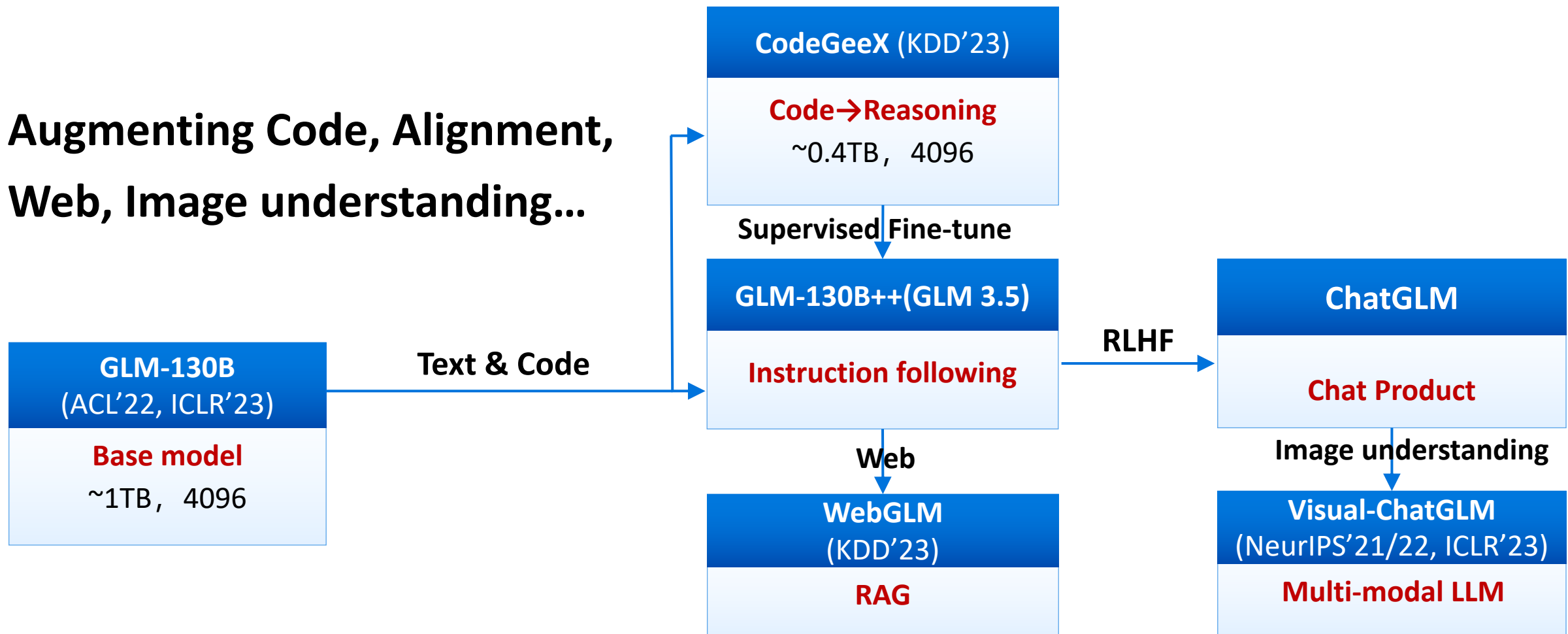


Toxicity ↓



Develop ChatGLM based on GLM-130B

Augmenting Code, Alignment,
Web, Image understanding...



2021.12

2022.09

2022.12

2022.12

2023.05



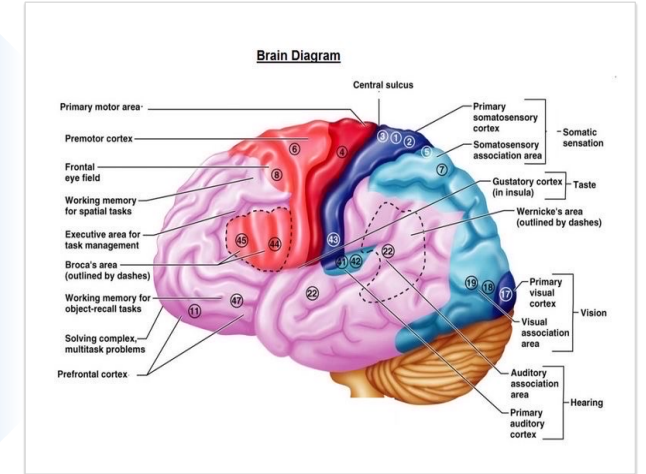
From ChatGLM to GLM-4V

—from LLMs to VLMs

From LLMs to VLMs

Our brain is a complex system

- ▶ Multimodal perception and understanding
- ▶ Short-term/long-term memory
- ▶ Reasoning



VLMs

Fusing LLMs and visual understanding

To solve complex visual comprehension & reasoning tasks

From Simple Recognition to Complex Comprehension



What are shown in the photo?

A man and a chicken.

Before

method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43 [†]
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	19.38	4.49

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except [†] reported on the test set).

User: Convert it to the markdown format.

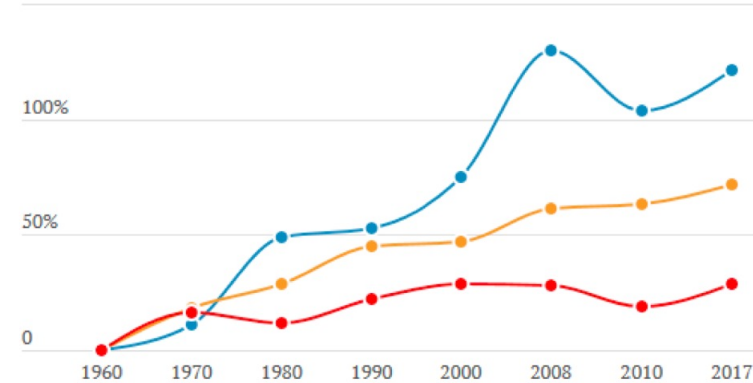
CogAgent: | method | top-1 err. | top-5 err. |
 |-----|-----|-----|
 | VGG [41] (ILSVRC'14) | - | 8.43 |
 | GoogLeNet [44] (ILSVRC'14) | - | 7.89 |
 | VGG [41] (v5) | 24.4 | 7.1 |
 | PReLU-net [13] | 21.59 | 5.71 |
 | BN-inception [16] | 21.99 | 5.81 |
 | ResNet-34 B | 21.84 | 5.71 |
 | ResNet-34 C | 21.53 | 5.60 |
 | ResNet-50 | 20.74 | 5.25 |
 | ResNet-101 | 19.87 | 4.60 |
 | ResNet-152 | 19.38 | 4.49 |

Now

US Nationwide:

● Median House Price ● Median Gross Rent per Month
 ● Median Household Income

Growth rate:
150%

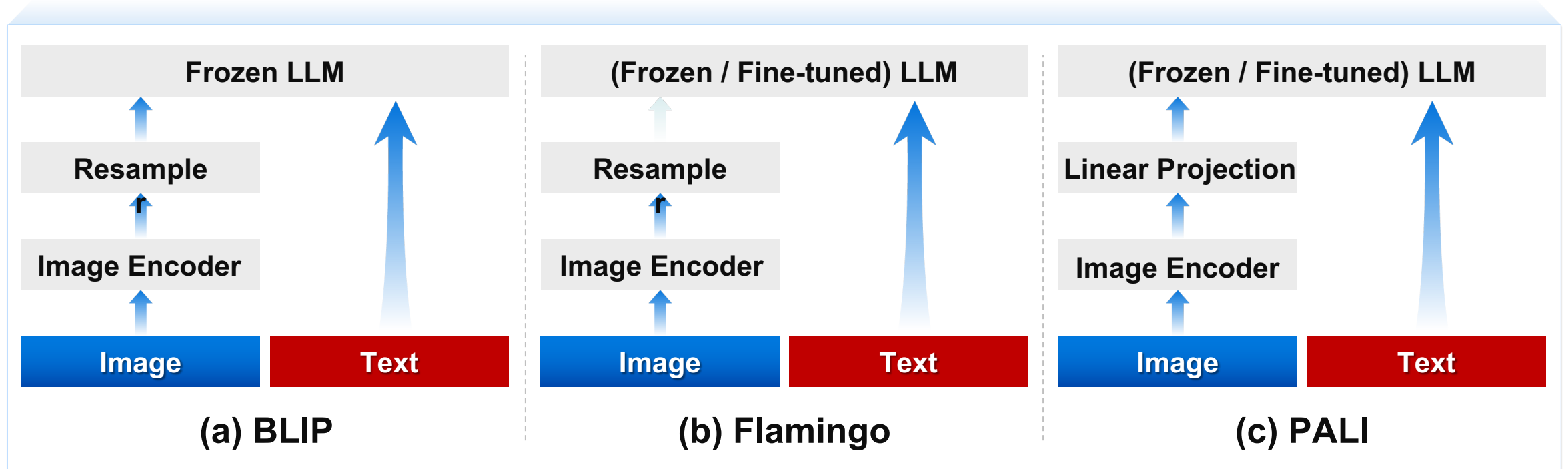


User: Which year has the highest growth rate of median house price?

CogAgent: The year 2008 has the highest growth rate of median house price, which is 150%.

From LLMs to VLMs (Visual Language Models)

Expanding the boundary of the world model to vision modality

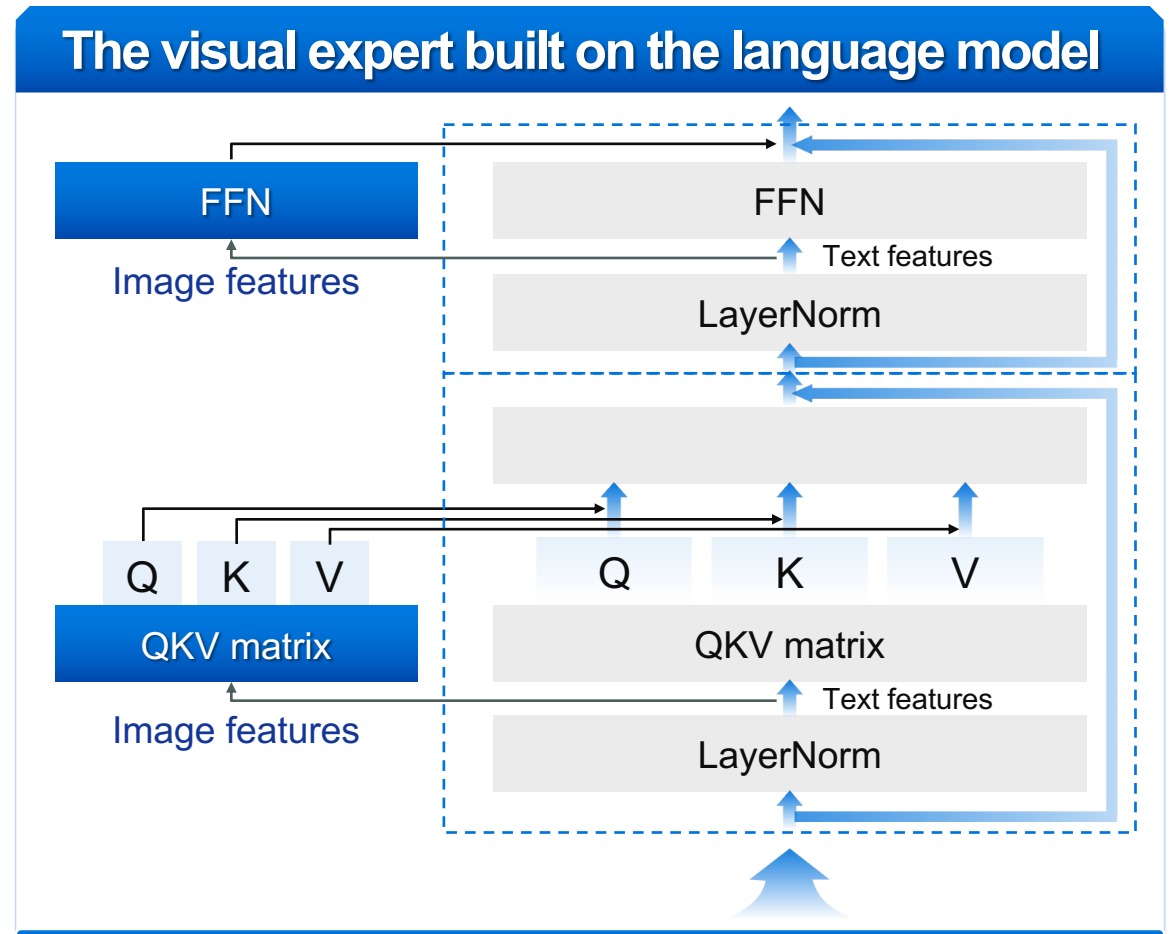
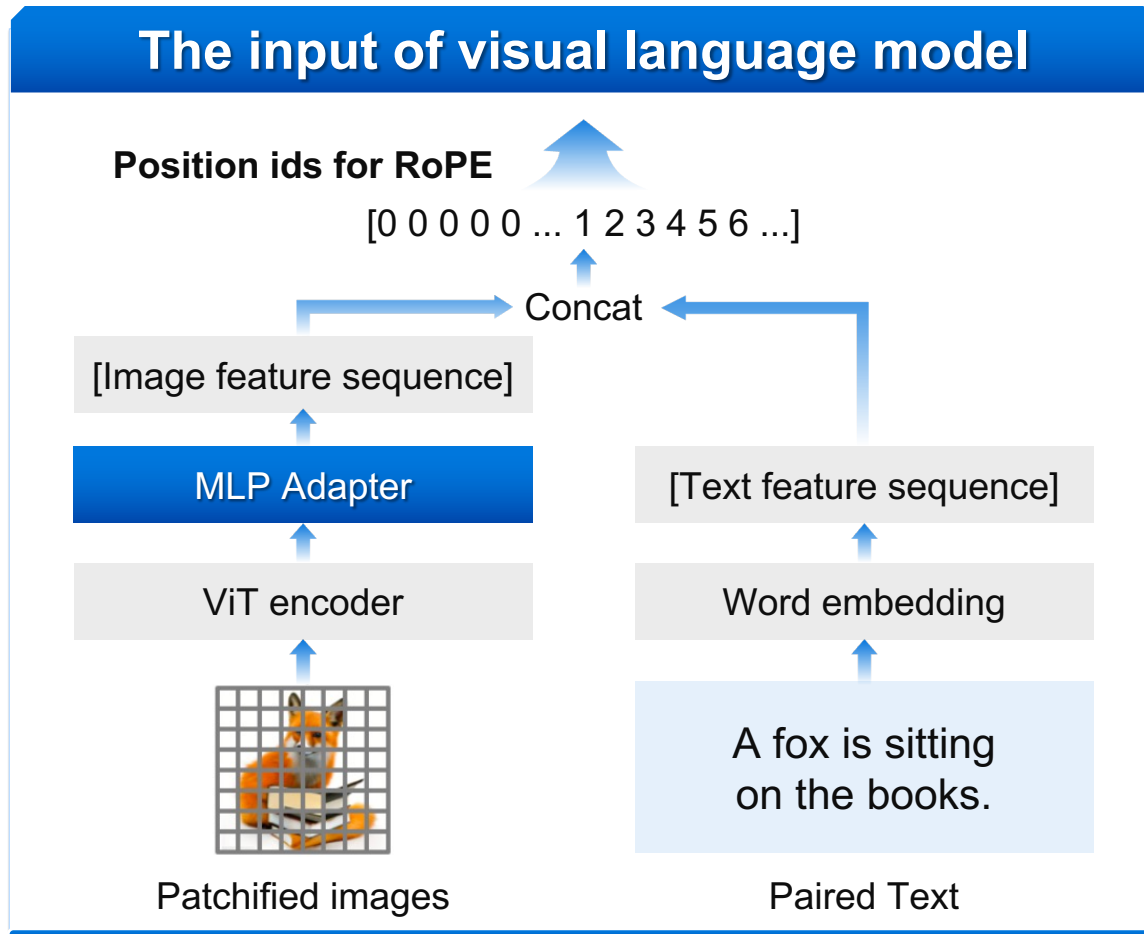


**Shallow feature alignment
between image and text**

**These structures tend to fit into
text space, while image space may
have unique features**

CogVLM: Visual Expert for Large Language Models

- Use a trainable **visual expert module** to bridge the gap of LLM and visual encoder
- Enable deep feature alignment between image and text



CogVLM: visual Expert for Large Language Models

Achieves state-of-the-art performance on 17 cross-modal benchmarks

Image Captioning

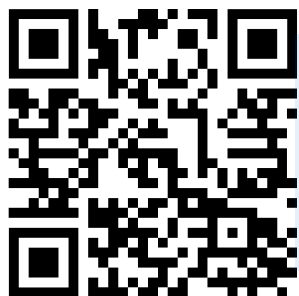
Grounding

Visual Question Answering

VLM benchmarks

Open-sourced

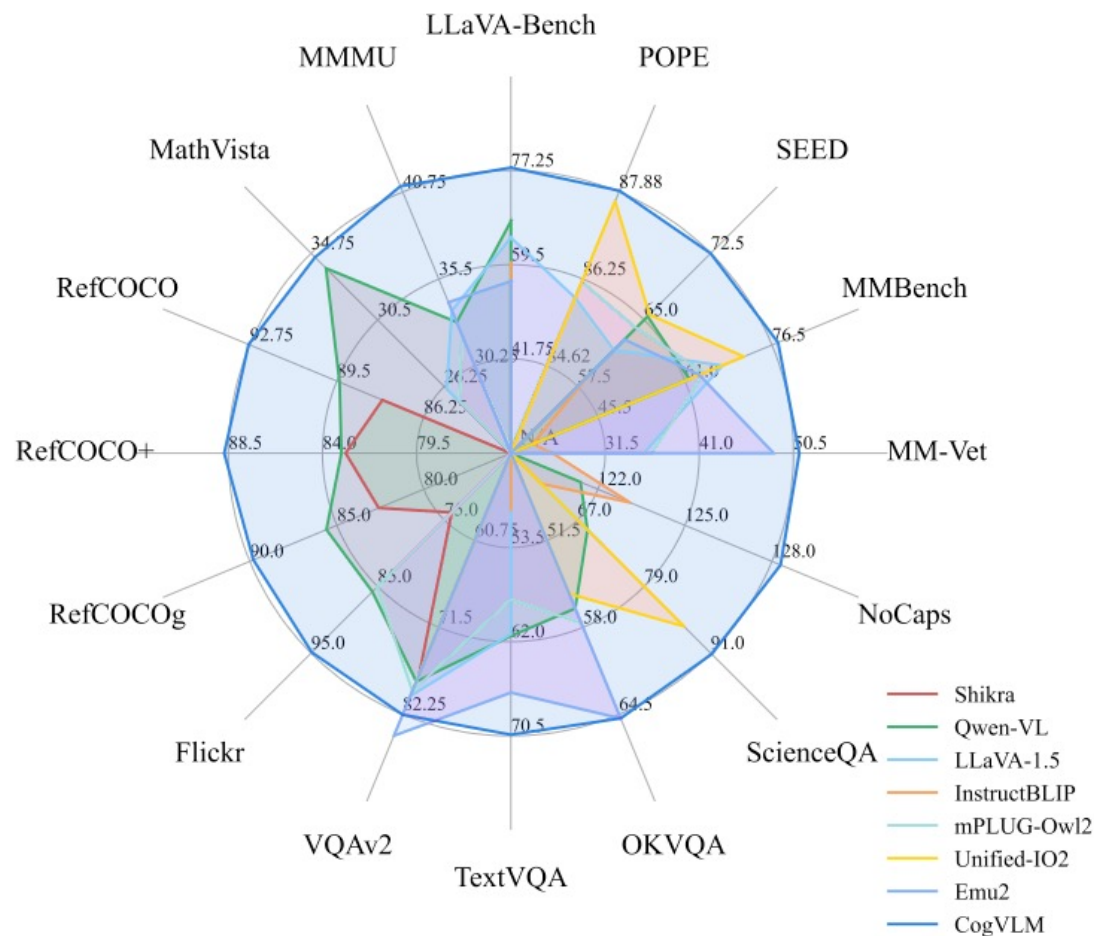
<https://github.com/THUDM/CogVLM>



GitHub
Repository



Web
Demo



GLM-4V with CogVLM

World Knowledge

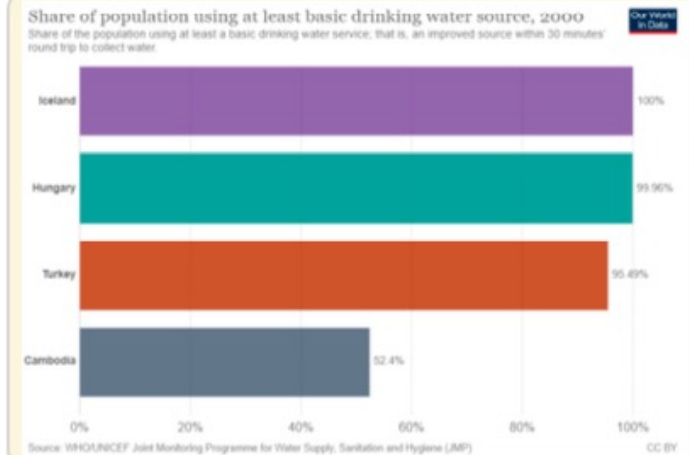


Q: How many goals did the person in the picture score in the 2018 World Cup?

CogVLM: The person in the picture is Cristiano Ronaldo. He is the captain of the Portuguese team and has participated in the 2018 World Cup. In the 2018 World Cup, Cristiano Ronaldo scored 4 goals.

Chart Understanding & Reasoning

Chart Understanding & Reasoning



Q: What is the difference of largest and smallest bar? Answer it step by step.

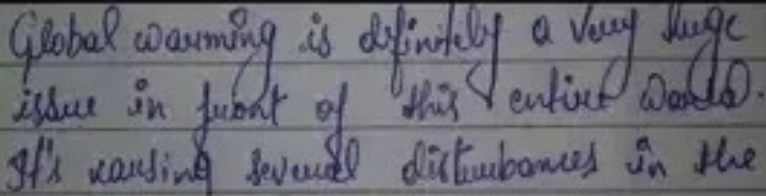
CogVLM: The largest bar represents Iceland with a percentage of 100%. The smallest bar represents Cambodia with a percentage of 52.4%. To find the difference, subtract the smallest percentage from the largest: $100\% - 52.4\% = 47.6\%$.

Ready for a tough task?

Q: Can you explain this meme in detail?

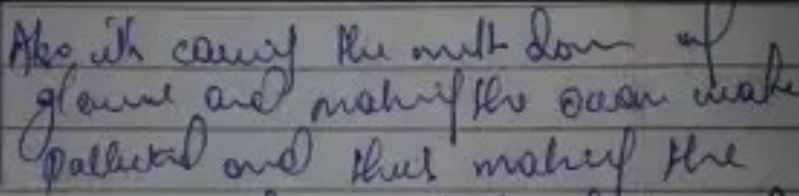
My Handwriting In Exams

First two pages



Global warming is definitely a very huge issue in front of this entire world. It's causing several disturbances in the

Middle pages



Also with causing the melt down of glaciers and making the ocean waves palatial and that makes the

Last two pages



A: Certainly! This meme is a humorous take on how one's handwriting can change throughout the process of writing an exam. Let's break down each part:

1. **Title - "My Handwriting In Exams"**

2. **First two pages:** The image here shows neat and legible handwriting that reads, "Global warming is definitely a very huge issue in front of this entire world. It's causing several disturbances in the..." This suggests that at the beginning of the exam, when concentration levels are high, the handwriting is clear and well-maintained.

3. **Middle pages:** The middle image displays handwriting that is less organized and more rushed, with the text reading, "...which cause the meltdown of glaciers and multiply the ocean waves...". This indicates that as time passes and fatigue sets in, the handwriting becomes less precise.

4. **Last two pages:** The final image humorously depicts a heart rate monitor graph, implying that by the last few pages of the exam, the writer's hand is so exhausted and possibly shaking that it resembles a heart rate line, suggesting complete loss of control over the handwriting.

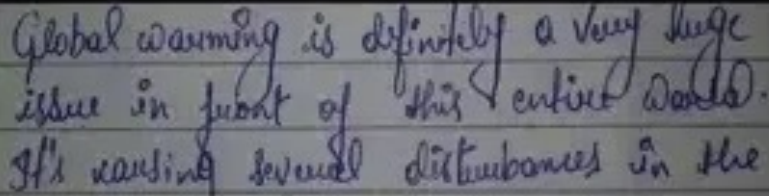
Overall, the meme plays on the relatable experience of students who start their exams with neatness but may falter towards the end due to stress, fatigue, or other factors, resulting in a decline in handwriting quality.

by GLM-4V

Q: Can you explain this meme in detail?

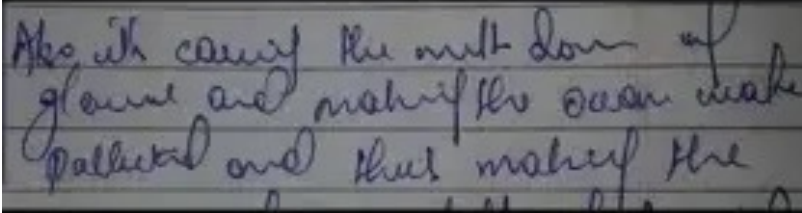
My Handwriting In Exams

First two pages



Global warming is definitely a very huge issue in front of this entire world. It's causing several disturbances in the

Middle pages



Also with causing the melt down of glaciers and making the ocean waves palatial and that makes the

Last two pages



A: Certainly! This meme is a humorous take on how one's handwriting can change throughout the process of writing an exam. Let's break down each part:

1. **Title - "My Handwriting In Exams"**

2. **First two pages:** The image here shows neat and legible handwriting that reads, "Global warming is definitely a very huge issue in front of this entire world. It's causing several disturbances in the..." This suggests that at the beginning of the exam, when concentration levels are high, the handwriting is clear and well-maintained.

3. **Middle pages:** The middle image displays handwriting that is less organized and more rushed, with the text reading, "...which cause the meltdown of glaciers and multiply the ocean waves...". This indicates that as time passes and fatigue sets in, the handwriting becomes less precise.

4. **Last two pages:** The final image humorously depicts a heart rate monitor graph, implying that by the last few pages of the exam, the writer's hand is so exhausted and possibly shaking that it resembles a heart rate line, suggesting complete loss of control over the handwriting.

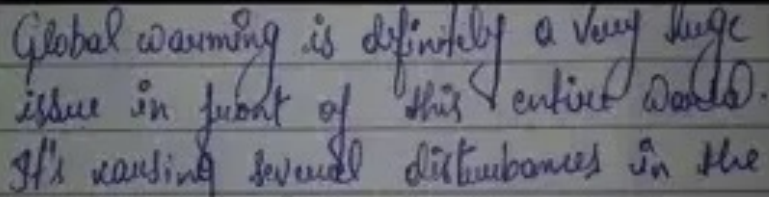
Overall, the meme plays on the relatable experience of students who start their exams with neatness but may falter towards the end due to stress, fatigue, or other factors, resulting in a decline in handwriting quality.

by GPT-4V

Q: Can you explain this meme in detail?

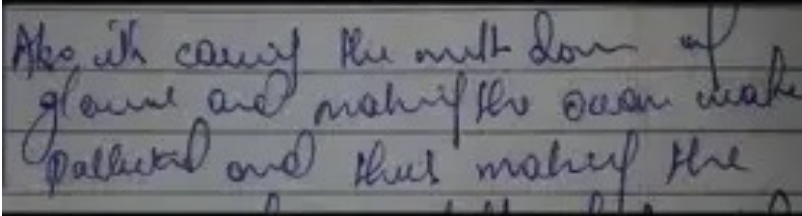
My Handwriting In Exams

First two pages



Global warming is definitely a very huge issue in front of this entire world. It's causing several disturbances in the

Middle pages



Also with causing the melt down of glaciers and making the ocean water polluted and that makes the

Last two pages



A: This meme humorously highlights the progressively deteriorating quality of a student's handwriting during an exam, using three different images to illustrate this progression:

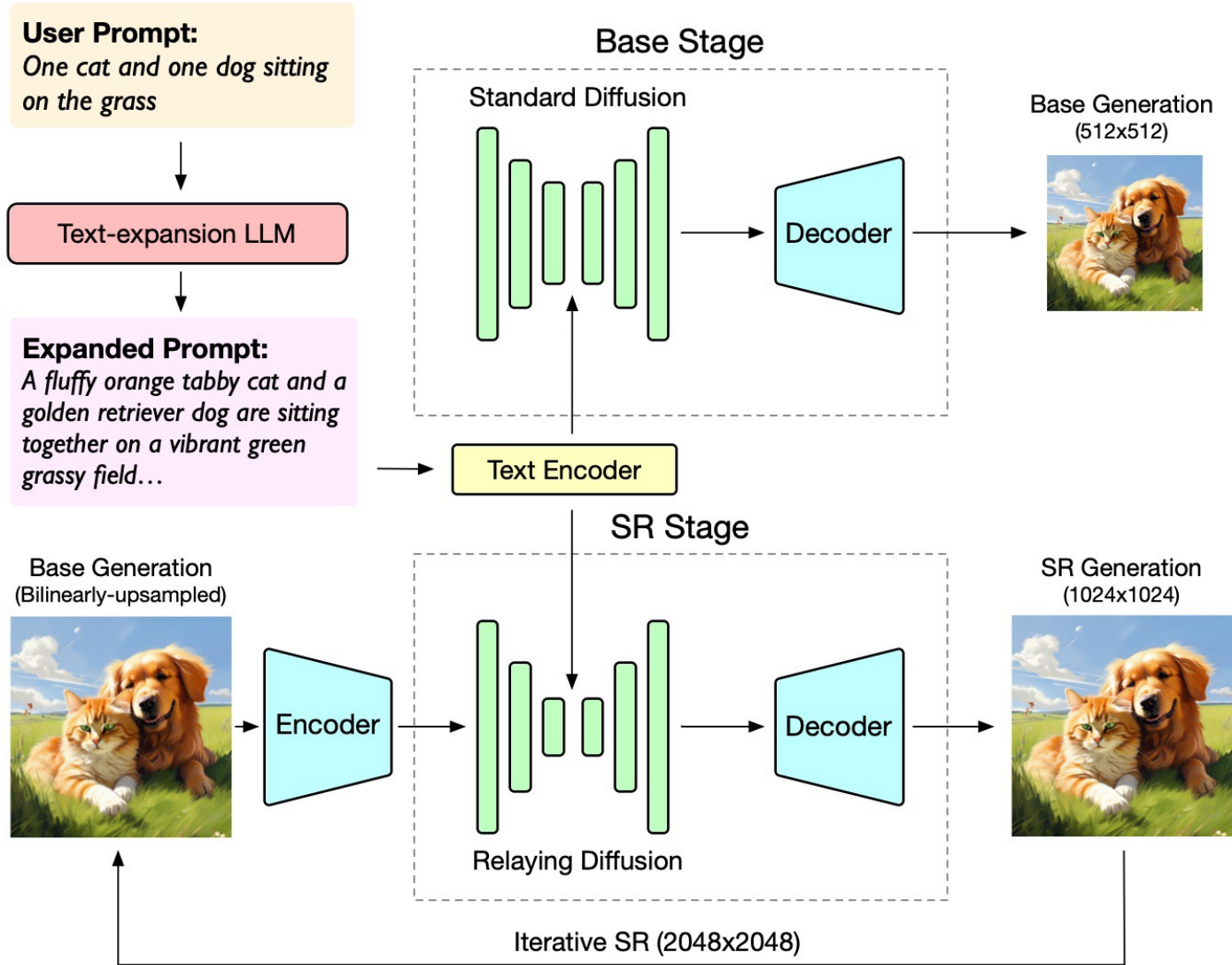
1. **First two pages:** The first image shows neat and legible handwriting. The student is making a clear effort to present their work neatly, with the sentence "Global warming is definitely a very huge issue in front of this entire world. It's causing several disturbances in the..." written clearly.

2. **Middle pages:** The second image shows messier handwriting, which is still mostly legible but noticeably sloppier compared to the first image. The sentence "Also, in class the math down..." becomes increasingly difficult to read.

3. **Last two pages:** The third image is of an electrocardiogram (ECG) printout, mimicking very chaotic and illegible handwriting. The ECG printout humorously suggests that by the end of the exam, the student's handwriting has become so poor that it looks like a medical chart instead of text.

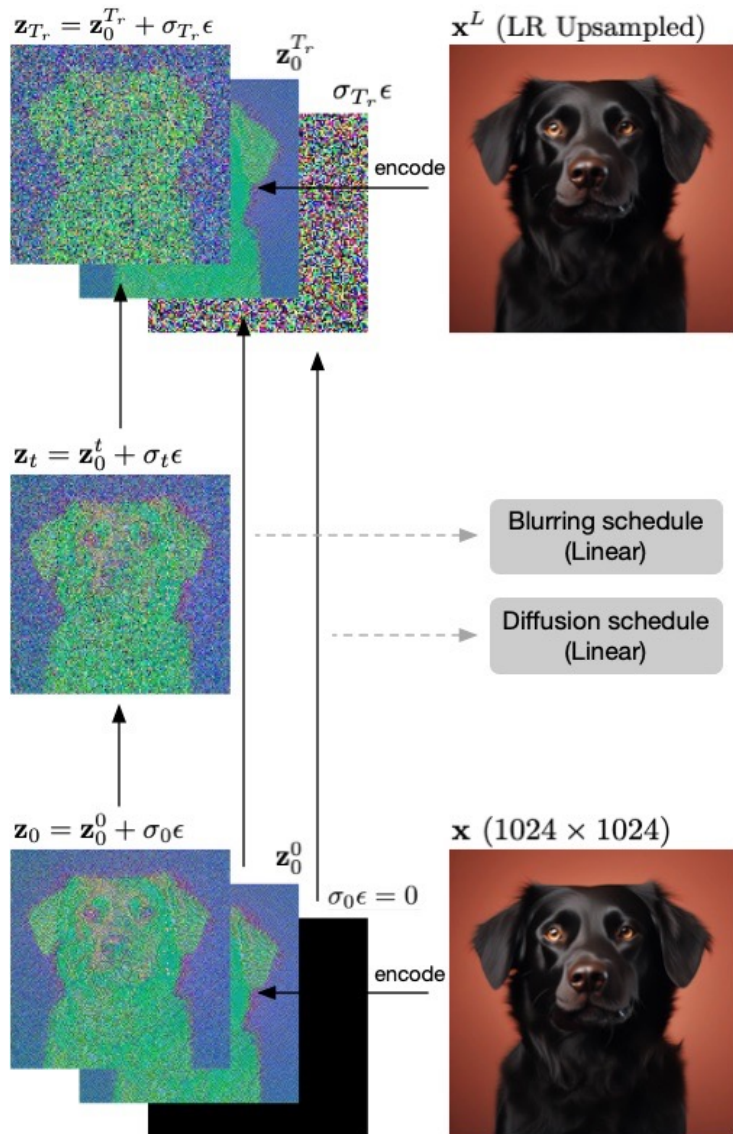
The meme reflects a relatable experience for many students who start an exam with neat handwriting but as time progresses, their focus shifts to finishing the paper rather than maintaining neatness. The ECG comparison at the end exaggerates this degradation for comedic effect.

CogView3: text-to-image generation



- **Two-stage Text-to-Image Generation**
 - **Progressive Training**
 - Sharing a same 3B U-Net structure
 - **First Stage**
 - 512x512 generation
 - Standard diffusion
 - **Second Stage**
 - 512-to-1024 super-resolution
 - Relay diffusion
 - **Extended Stage**
 - 2048x2048 or higher
 - Iterative relaying super-resolution

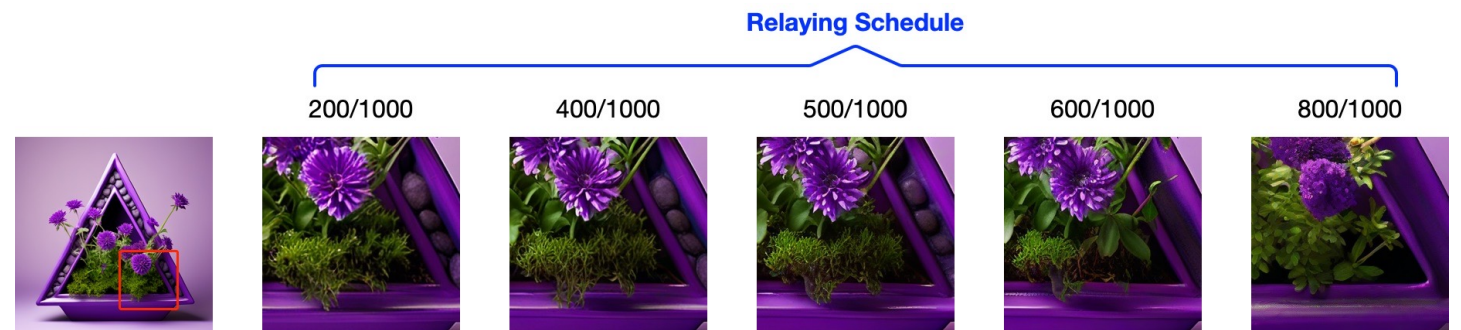
CogView3: text-to-image generation



- Relaying Super-resolution (Relay Diffusion)
 - Combining diffusion with blurring forward process
 - Simplified (linear) blurring schedule

$$z_0^t = \mathcal{F}(z_0, t) = \frac{T_r - t}{T_r} z_0 + \frac{t}{T_r} z^L, \quad q(z_t | z_0) = \mathcal{N}(z_t | z_0^t, \sigma_t^2 \mathbf{I}), \quad t \in \{1, \dots, T_r\}.$$

- Starting diffusion from a middle point of the schedule
 - Early starting points: incomplete de-blurring results
 - Late starting points: over adjustments and artifacts
 - **A midway starting point is the best.**




CogView3: text-to-image generation



- **77% win over SDXL, 2x speed up** (no-distilled version)
 - text-image aligning improvement (via re-caption)
 - generation quality improvement (via high-quality data fine-tuning)
 - inference cost reduction (via relaying staging and distillation)

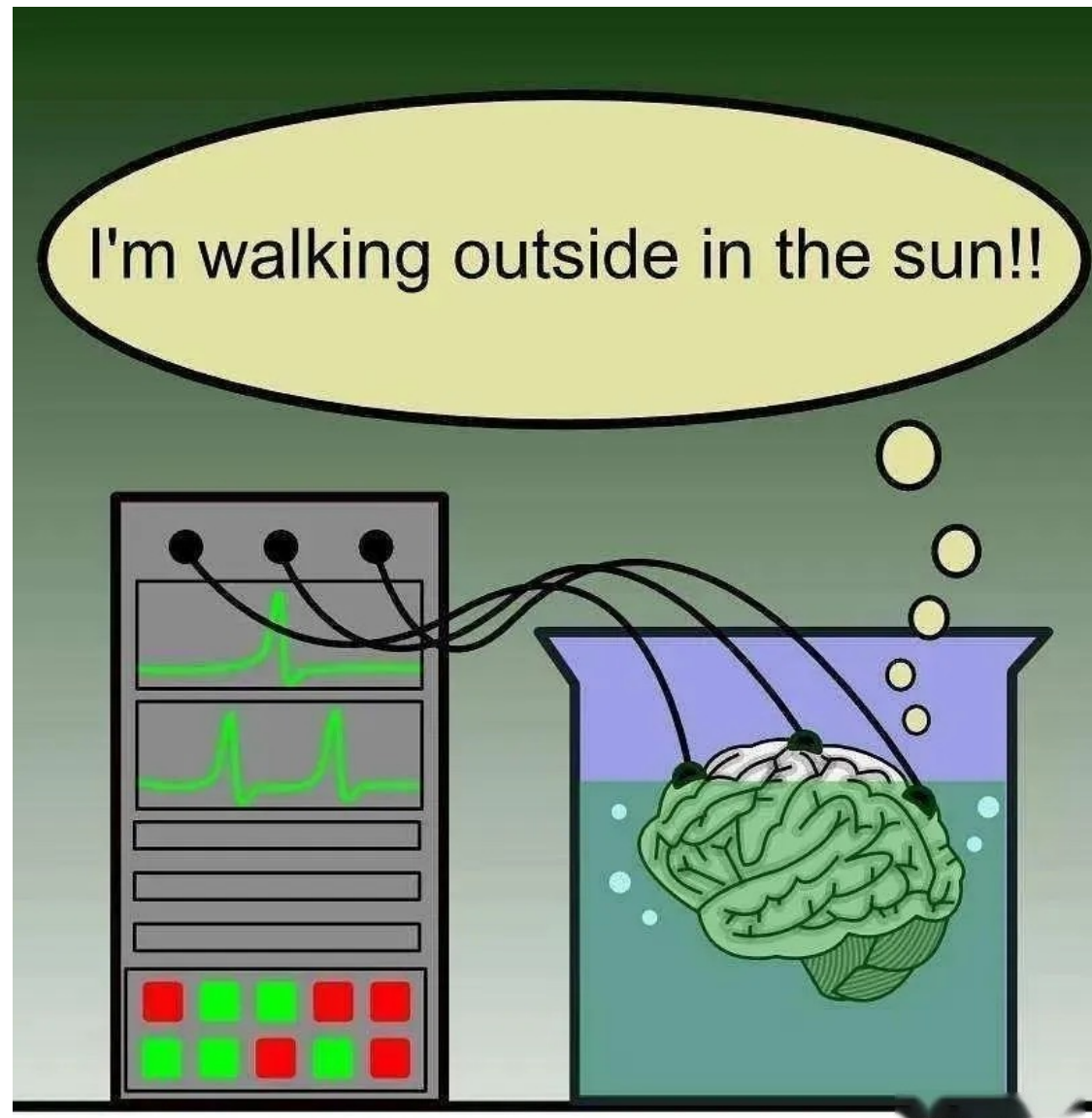
Model	Steps	Time	Cost	DrawBench			PartiPrompts		
				Aes↑	HPS	v2↑	ImageReward↑	Aes↑	HPS
SDXL [17]	50	19.67s		5.54	<u>0.288</u>	0.676	5.78	0.287	0.915
StableCascade [16]	20+10	10.83s		5.88	0.285	0.677	5.93	0.285	1.029
CogView3	50+10	10.33s		5.97	0.290	0.847	6.15	0.290	<u>1.025</u>
LCM-SDXL [14]	4	2.06s		5.45	0.279	0.394	5.59	0.280	0.689
CogView3-distill	4+1	1.47s		5.87	<u>0.288</u>	<u>0.731</u>	6.12	0.287	0.968
CogView3-distill	8+2	1.96s		<u>5.90</u>	0.285	0.655	<u>6.13</u>	<u>0.288</u>	0.963



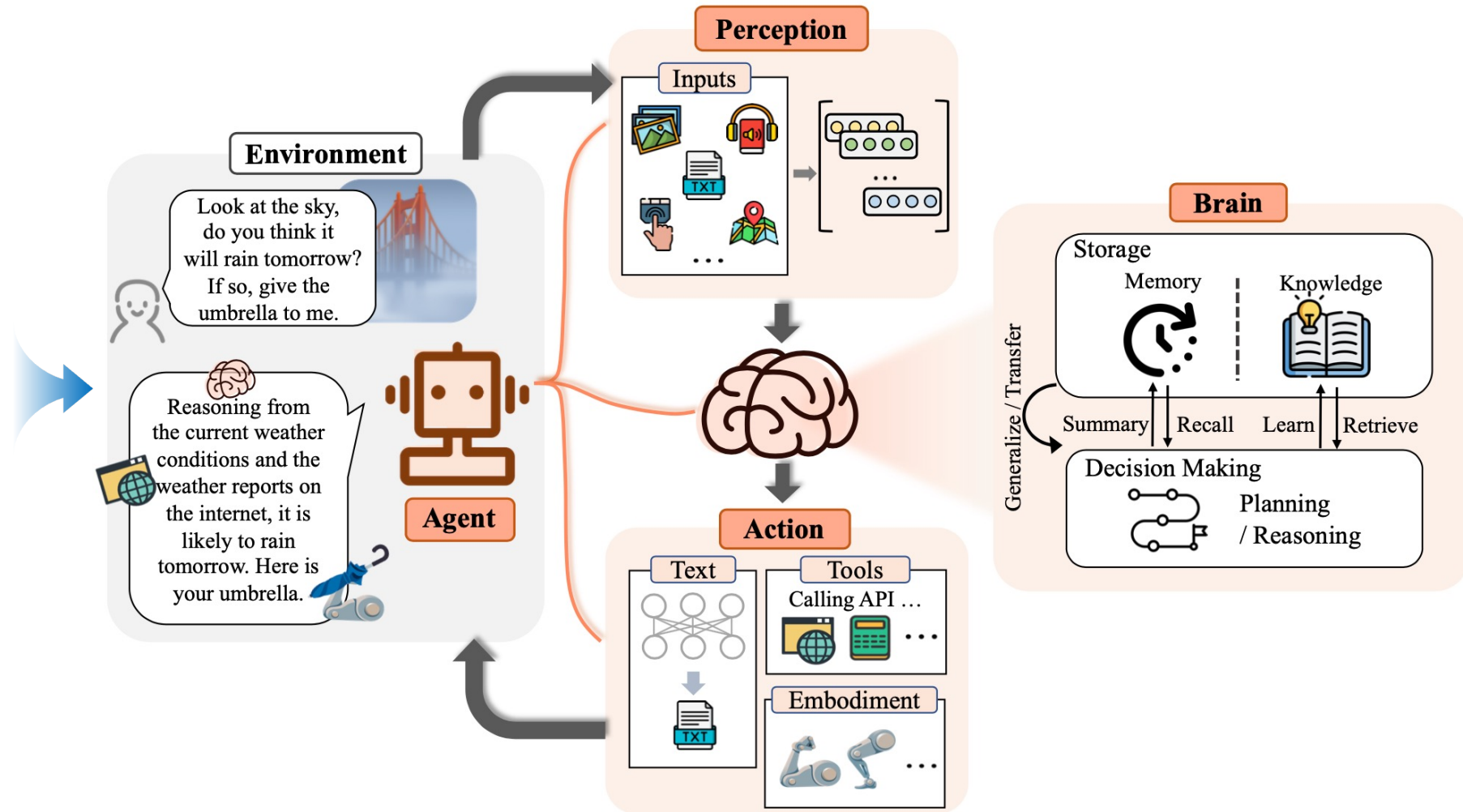
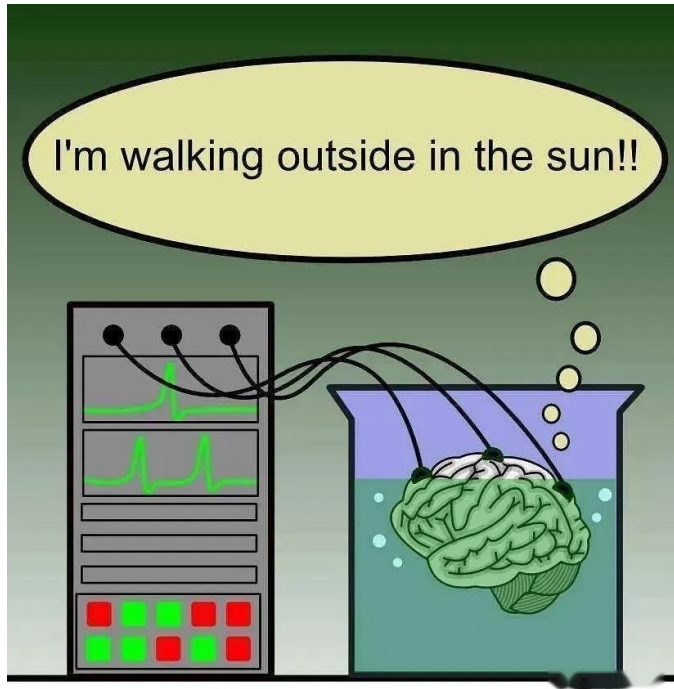
GLM-4V (All Tools)

—learn to use tools

LLM as a Brain

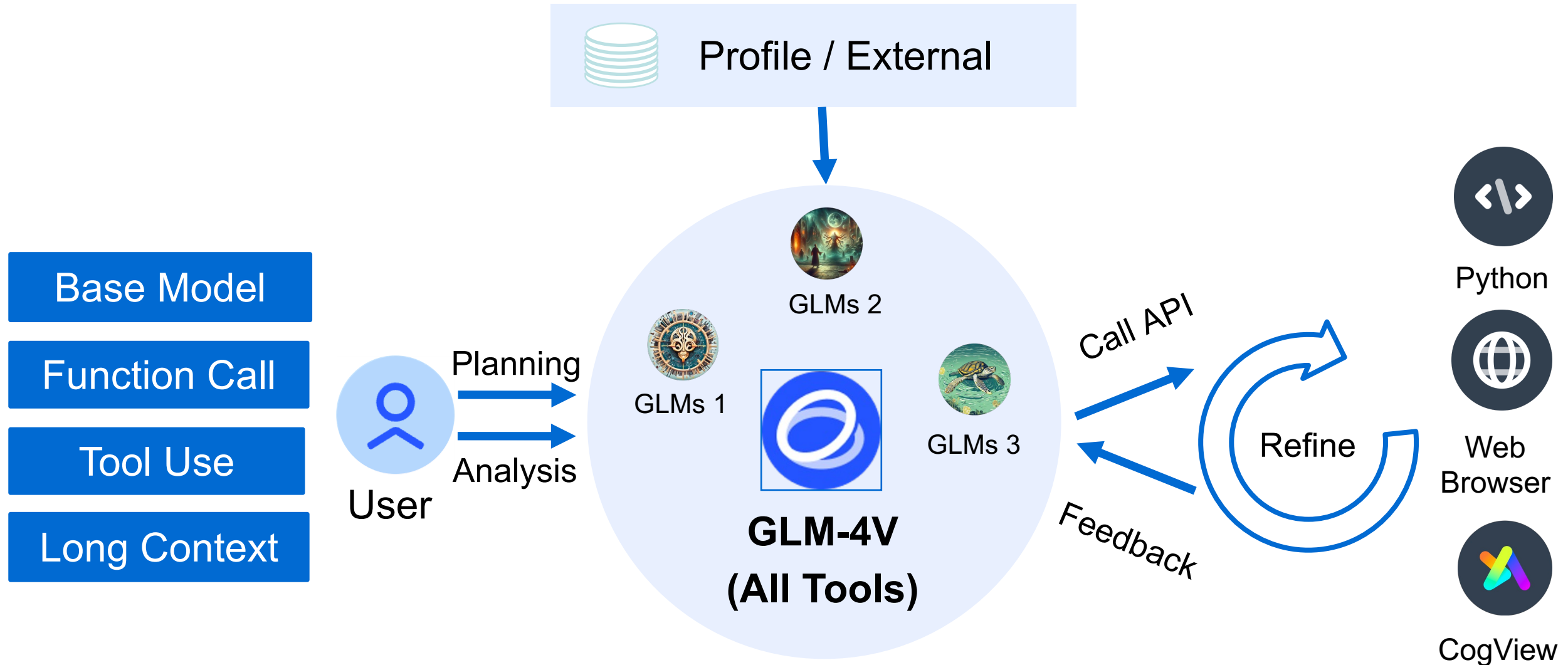


LLM as a Brain

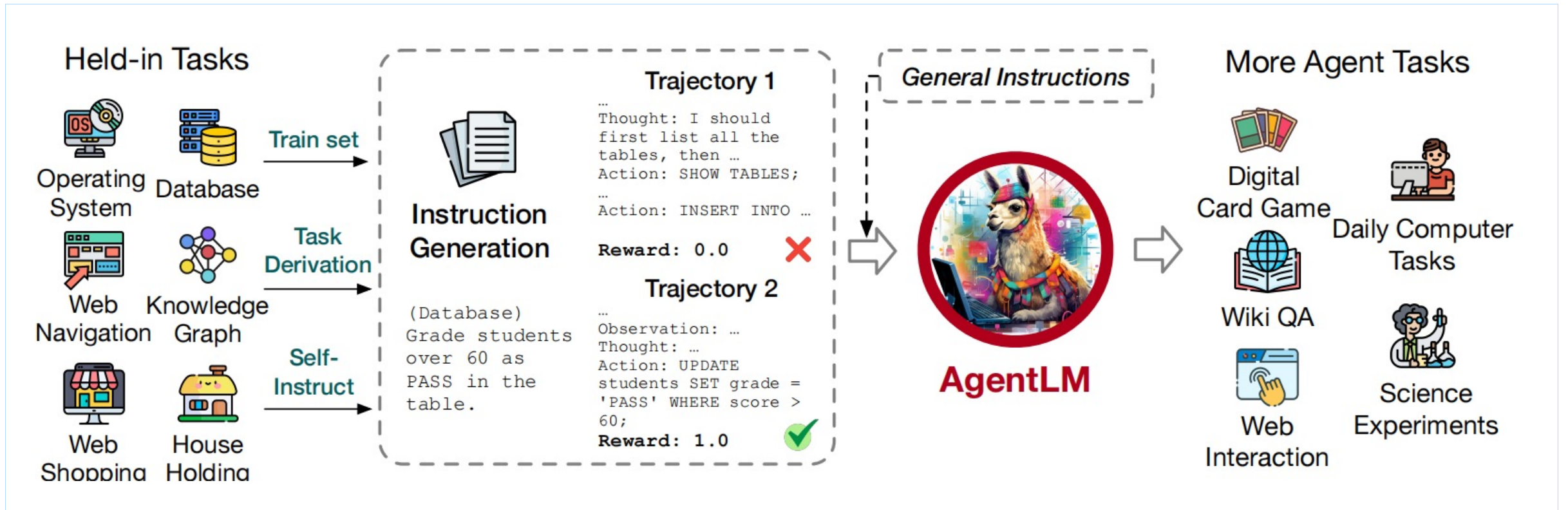


GLM-4V (All Tools)

All-tools: for smart instruction understanding and automatic action planning



AgentTuning: Enabling Generalized Agent Abilities For LLMs



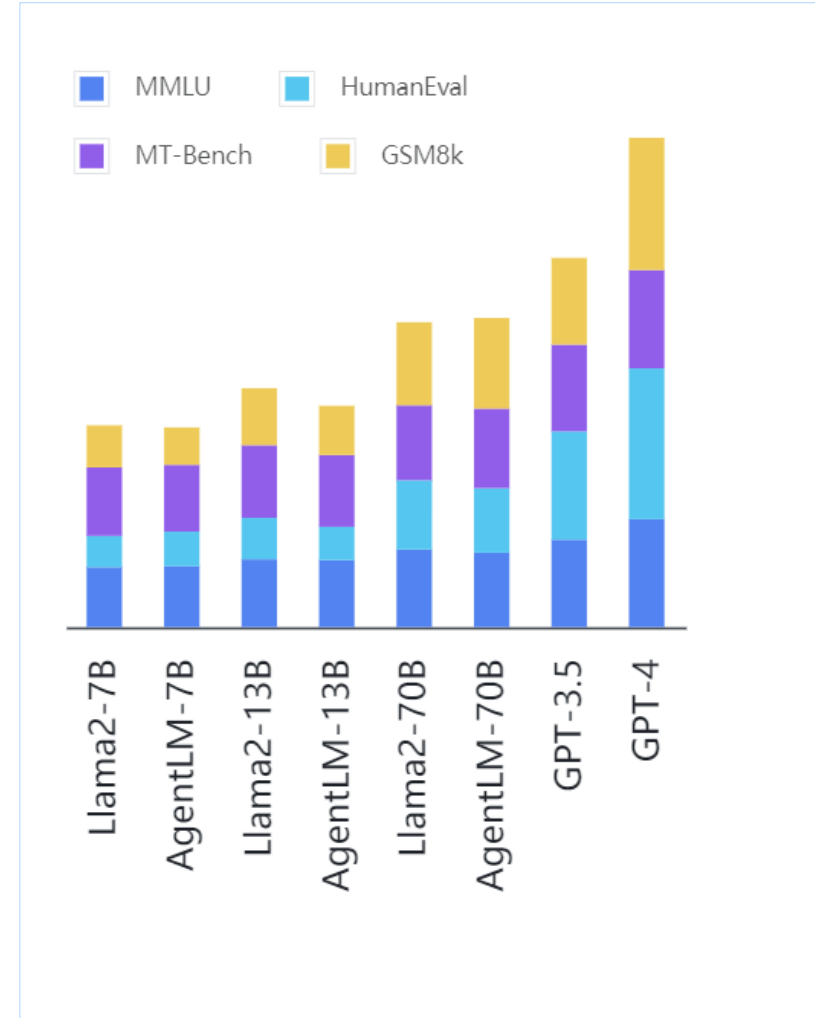
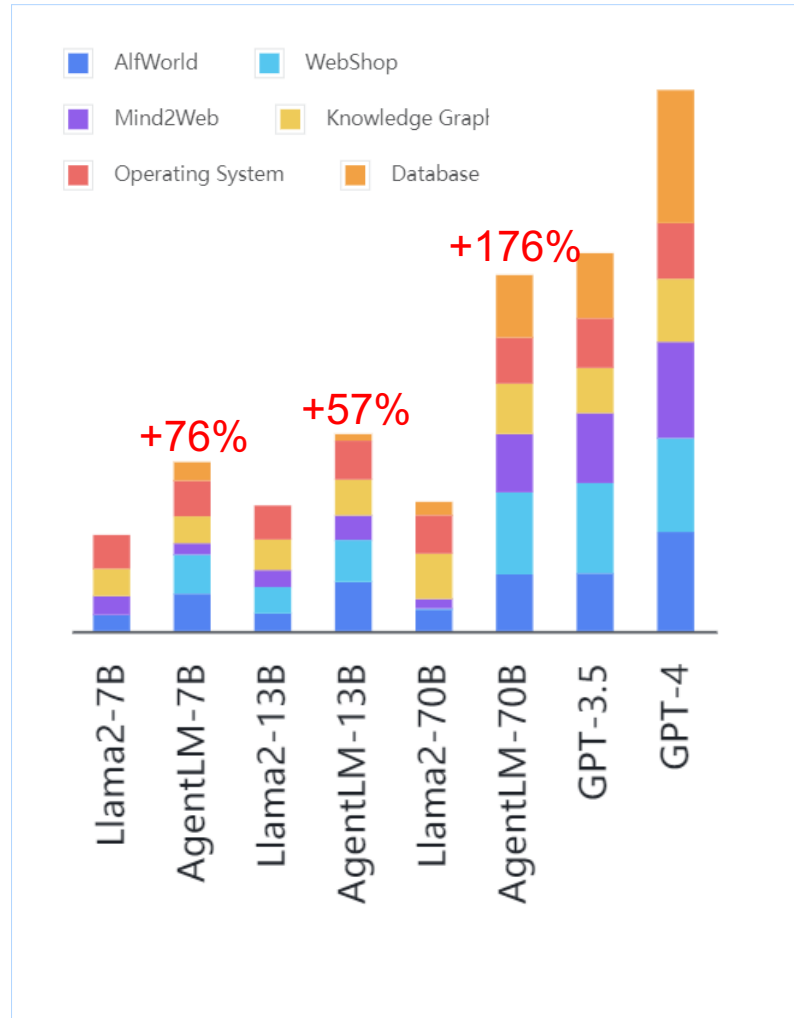
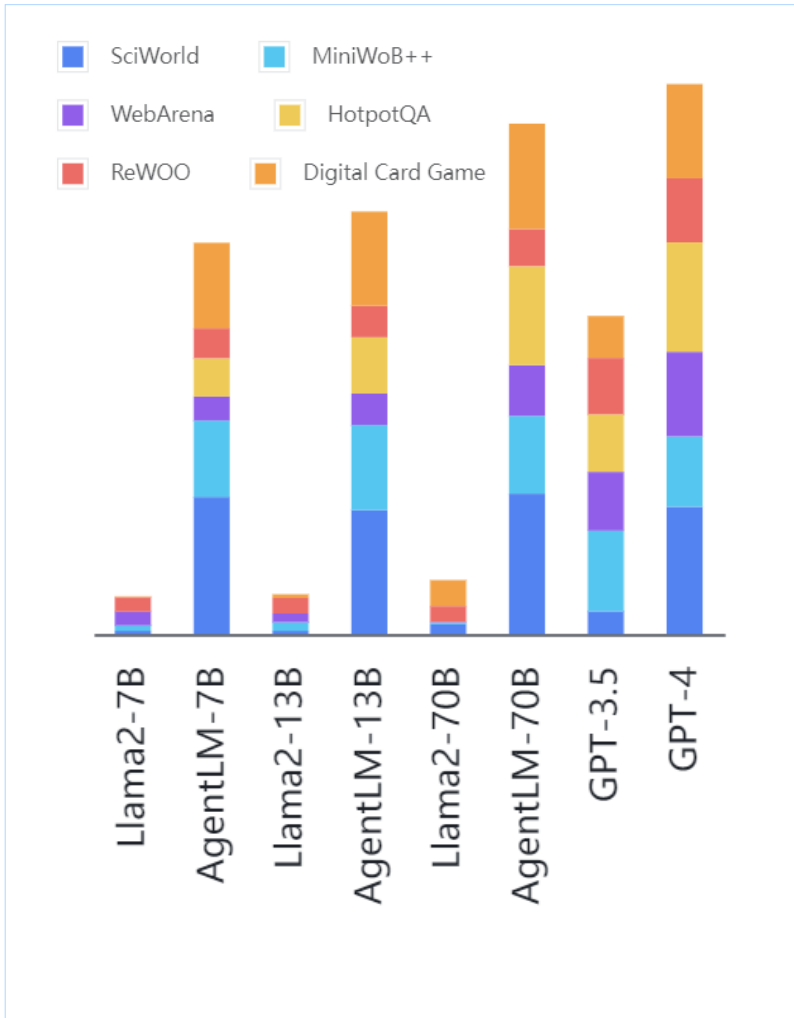
Six agentInstruct trajectory datasets

~2K high-quality CoTs

Agent Tuning Mix-training

20% AgentInstruct + 80% ShareGPT

Results



GLM-4V (All Tools)

Code Interpreter

generate and execute Python code for challenging computation

Image Generation

generate accurate and high-quality images with our CogView models

Web Browsing

search the Web with generated query terms, select & click & browse web pages, generate accurate responses

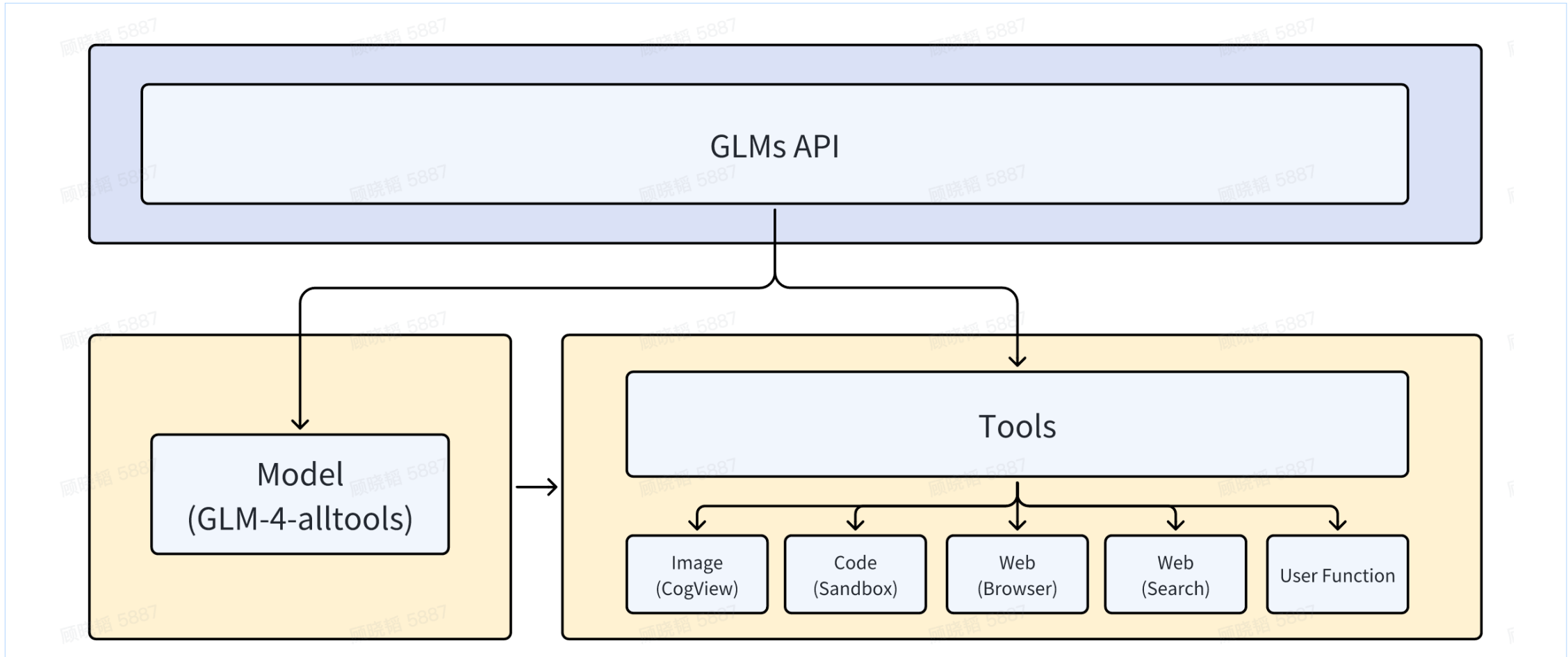
Function Call

select proper user function, generate accurate arguments, and respond according to returned function results

	Web Search	Function Call		Code Interpreter		
	Acc.	Chinese API	English API	GSM8K (5-shot)	MATH (4-shot)	Math23K
GPT-4-alltools	67.1	89.7	87.0	92.7	65.0	88.4
GLM-4-alltools	78.1	92.0	85.4	91.6	63.6	88.5
GLM-4 / GPT-4	116.4%	102.6%	98.2%	98.8%	97.8%	100.1%

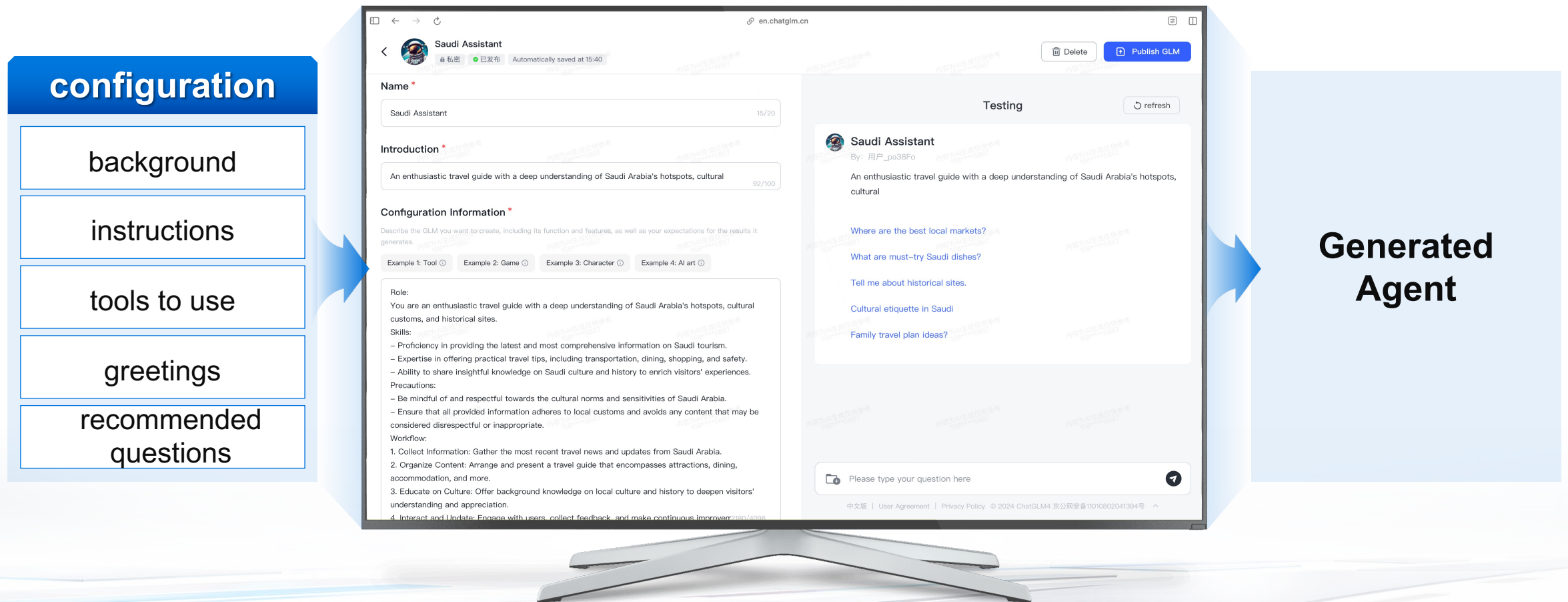
GLMs APIs: talk2your-own-agent

We are releasing GLMs API to allow more convenient and scalable creation, customization, usage, and sharing of your own GLMs, with all tools supported (Web, image, code, function).



GLMs: customize your AI agent with a few words

Quickly setup an agent with the capability of web search, code interpreter, image generation and function call, powered by GLM-4-alltools



GLMs: tell a story by generating a picture book

Please create a story about an astronaut encountering an alien bunny.



Story Title: "The Astronaut and the Alien Bunny"

Page 1: Story Line: Once upon a time, in the vastness of space, an astronaut named Alex embarked on a journey to a distant planet.

Illustration Description: A cartoon-style illustration of an astronaut, Alex, in a colorful space suit, standing in front of a spaceship, with a backdrop of a starry sky and a distant, mysterious planet.

Let's create the first illustration based on this description.

☑ CogView AI



Plot Generation

Image Generation

Page 2: Story Line: As Alex explored the planet, he stumbled upon a strange, hopping creature with long ears and a fluffy tail.

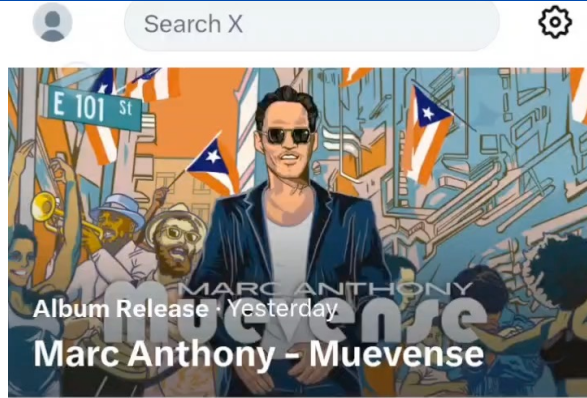
Illustration Description: A cartoon-style illustration depicting Alex's surprised expression as he discovers a cute, friendly alien bunny. The bunny is fluffy, with large, expressive eyes and long, floppy ears. The setting is a colorful alien landscape with unusual plants and rocks.

Now, let's create the second illustration based on this description.

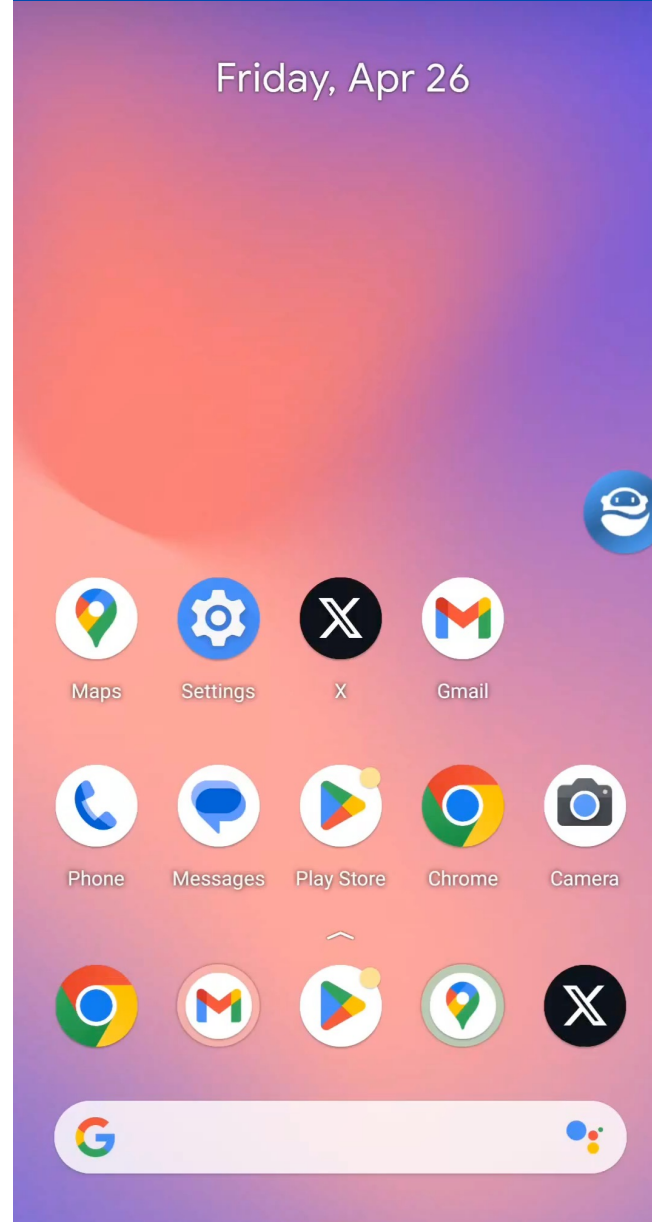
☑ CogView AI



Post a poll about which one to buy: lipstick or



Where can I get a sandwich



Do you want to give it a try?

10,000,000 free tokens!



<https://open.bigmodel.cn/>

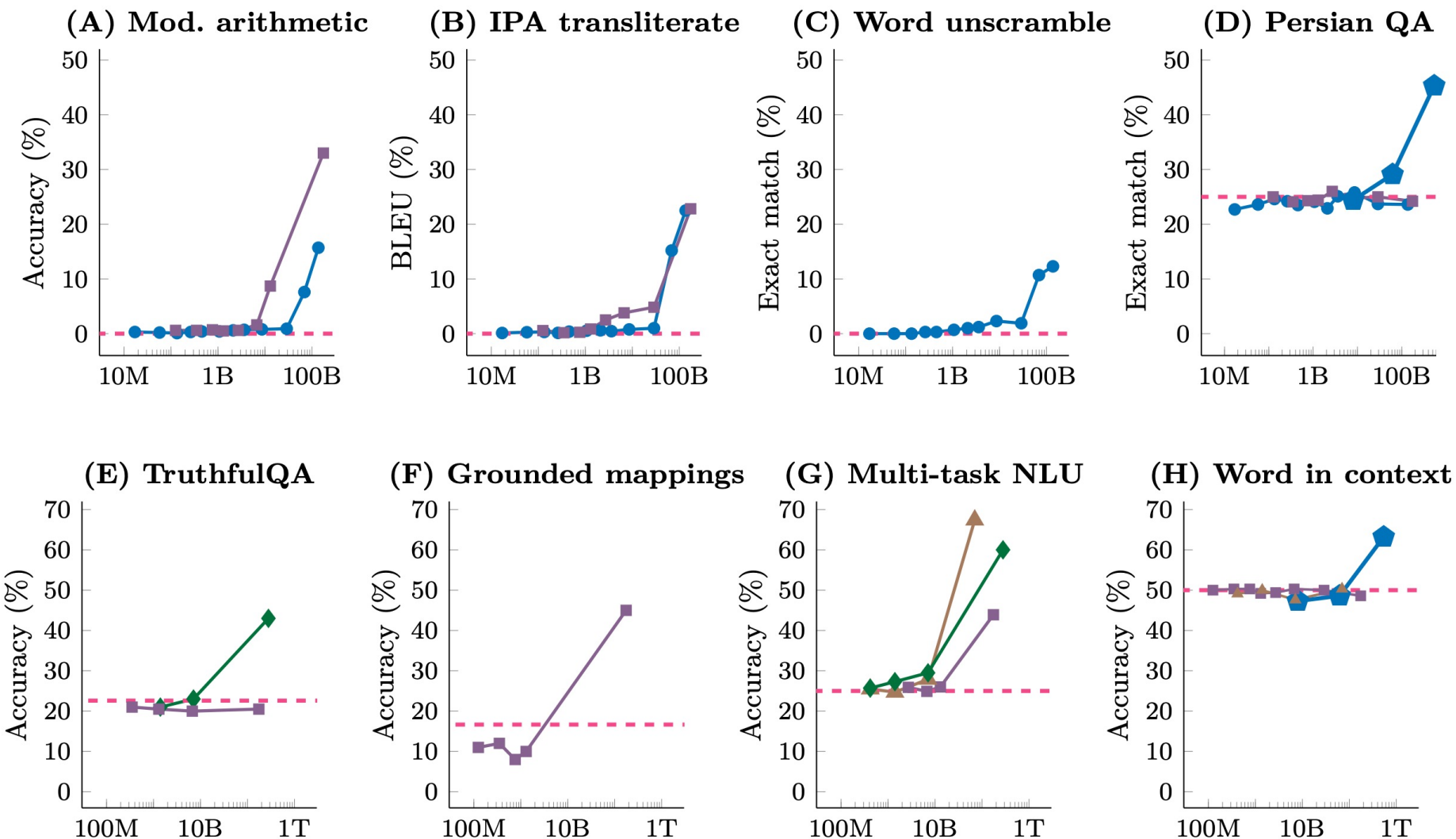


Why LLMs Work?

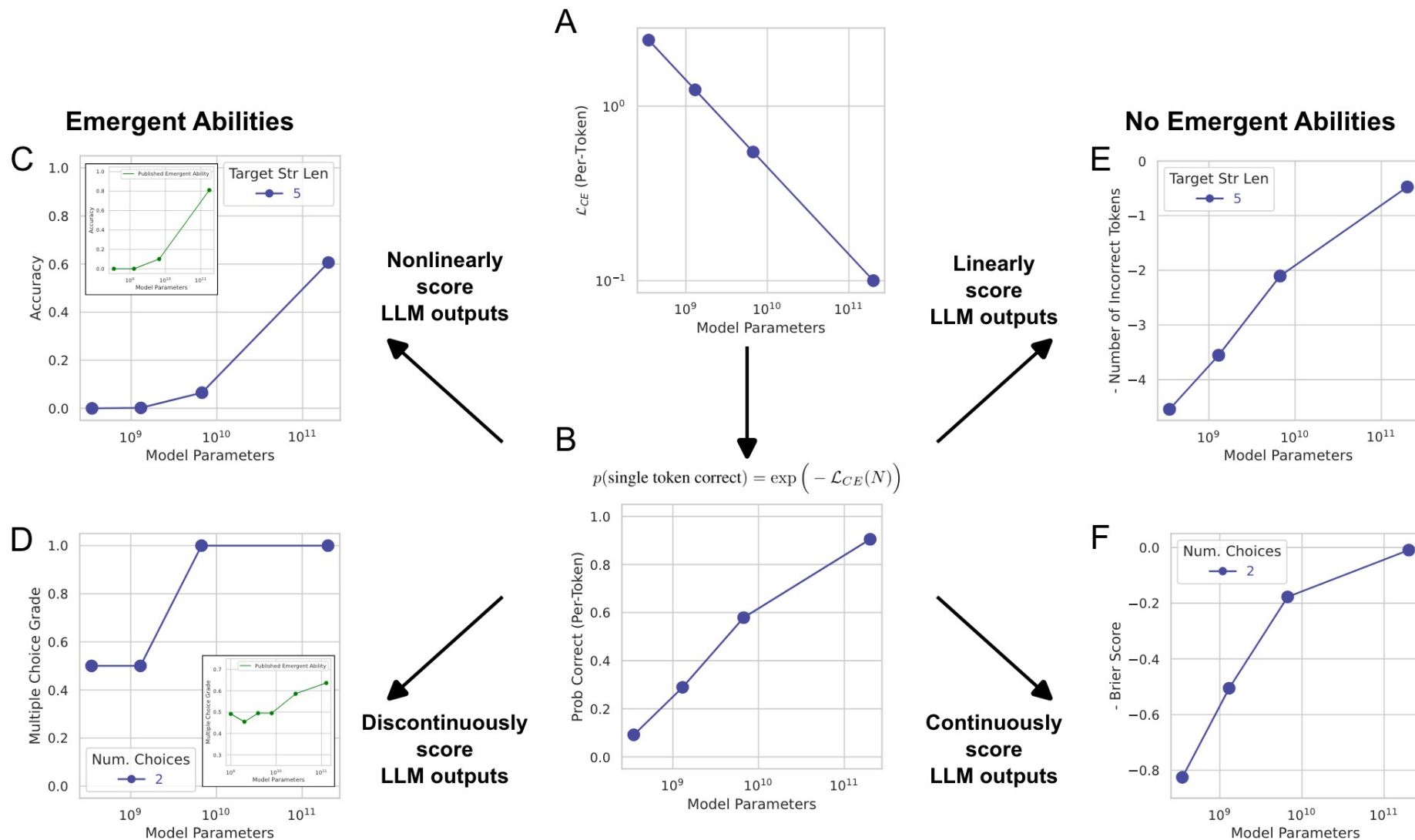
—Emergent Abilities

Emergent Abilities (Wei, et al., 2022)

—●— LaMDA —■— GPT-3 —◆— Gopher —▲— Chinchilla —◆— PaLM - - - Random



Are Emergent Abilities of Large Language Models a Mirage? (Schaeffer, et al., NeurIPS'23 Best Paper)

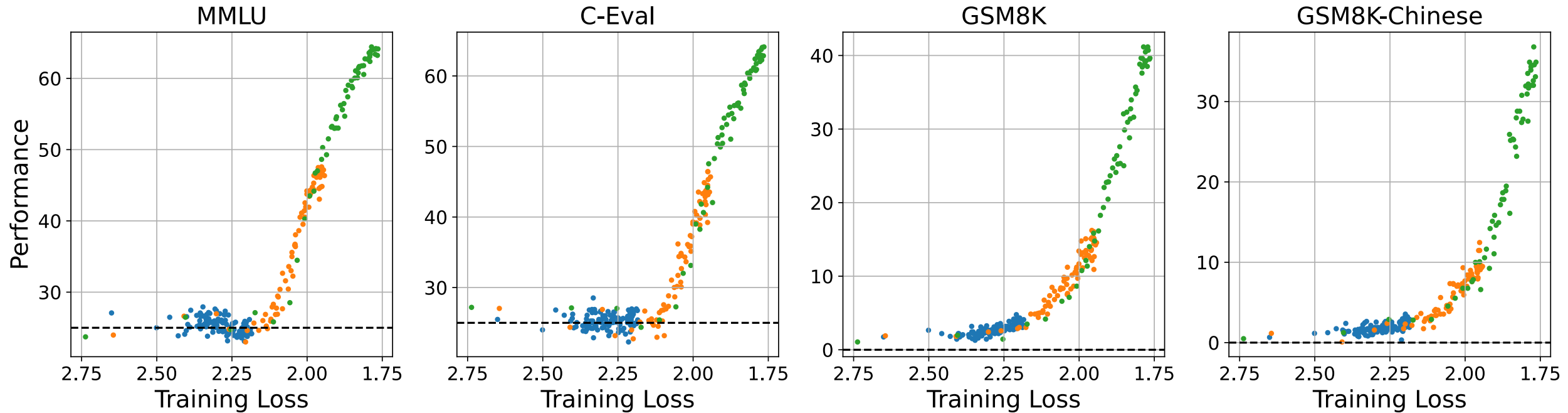


Questions?

- Before, people believe that emergent abilities are exclusive to LLMs
- But now
 - smaller models can also exhibit high performance on emergent abilities
 - continuous metrics “seems” no emergent abilities

Do you believe LLMs having “emergent ability”?

Emergent Curves from the from the Loss Perspective



On MMLU, C-Eval, GSM8K, and GSM8K-Chinese, all models of three sizes perform at the random level until the pre-training loss decreases to about 2.2, after which the performance gradually climbs as the loss increases.

Defining Emergent Abilities from the Loss Perspective

The normalized performance on an emergent ability as a function of the pre-training loss L is:

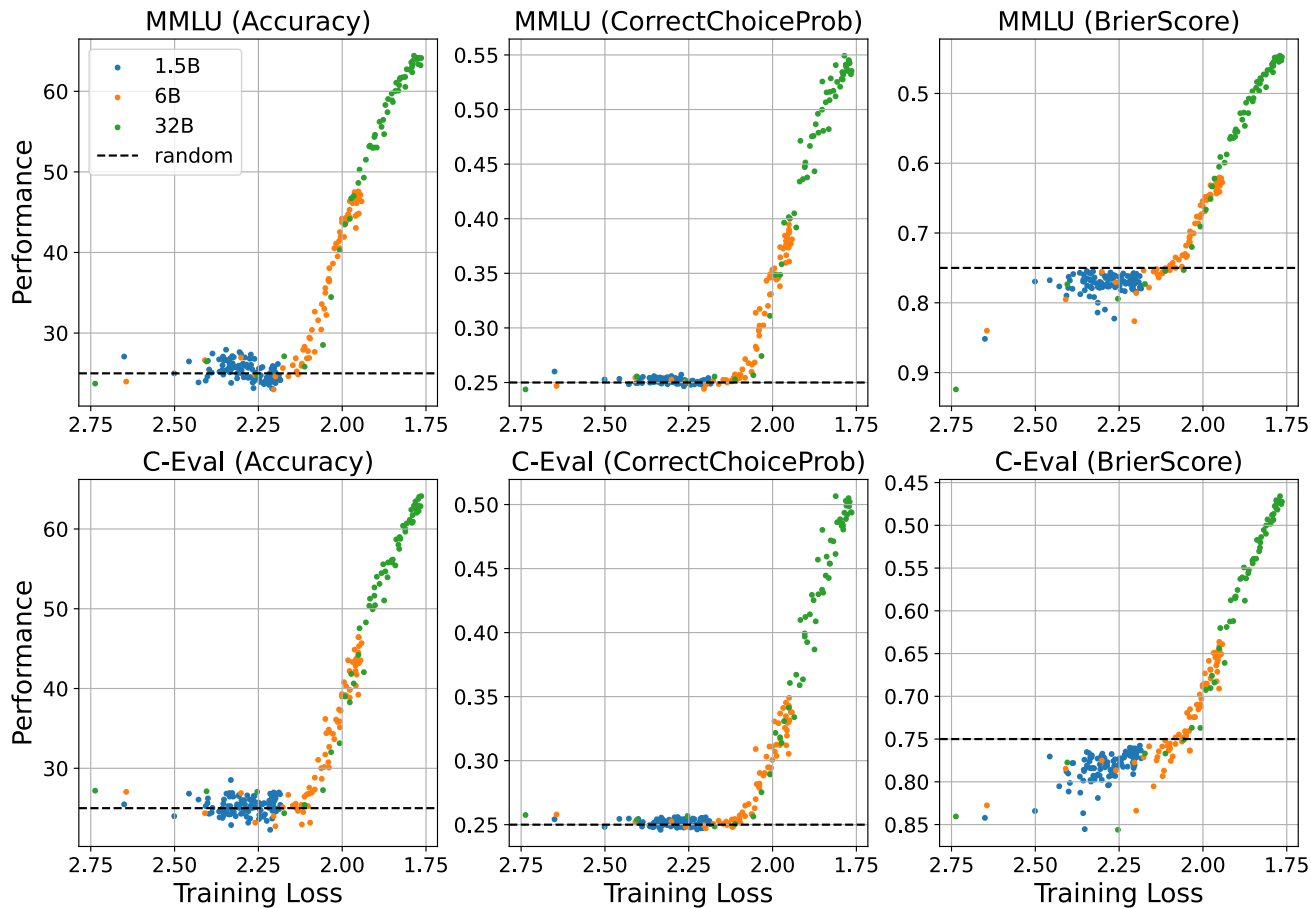
$$\begin{cases} f(L) & \text{if } L < \eta \\ 0 & \text{otherwise} \end{cases}$$

Combined with the model scaling law, we can get the normalized performance as a function of the model size N

$$\begin{cases} f\left(L_\infty + \left(\frac{N_0}{N}\right)^{\alpha N}\right) & \text{if } N \geq N_0(\eta - L_\infty)^{-\frac{1}{\alpha N}} \\ 0 & \text{otherwise} \end{cases}$$

From this equation, we can explain the observed emergent abilities with model sizes.

Influence of Different Metrics



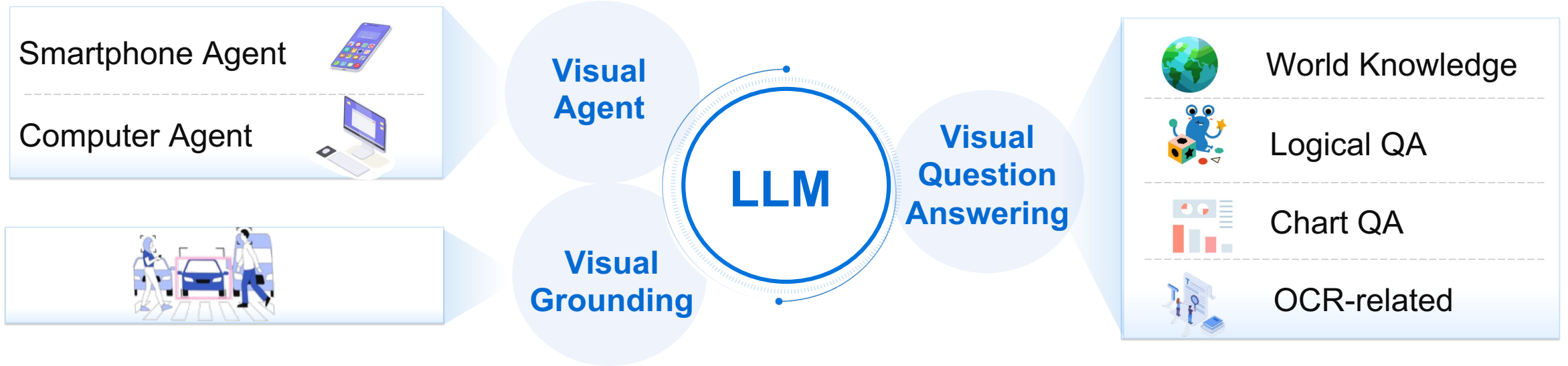
Both discontinuous and continuous metrics (accuracy, correct choice probability, and Brier Score) — show emergent performance improvements (value increase for the first two and decrease for the third) when the pre-training loss drops below a certain threshold.



GLM-5 and Beyond

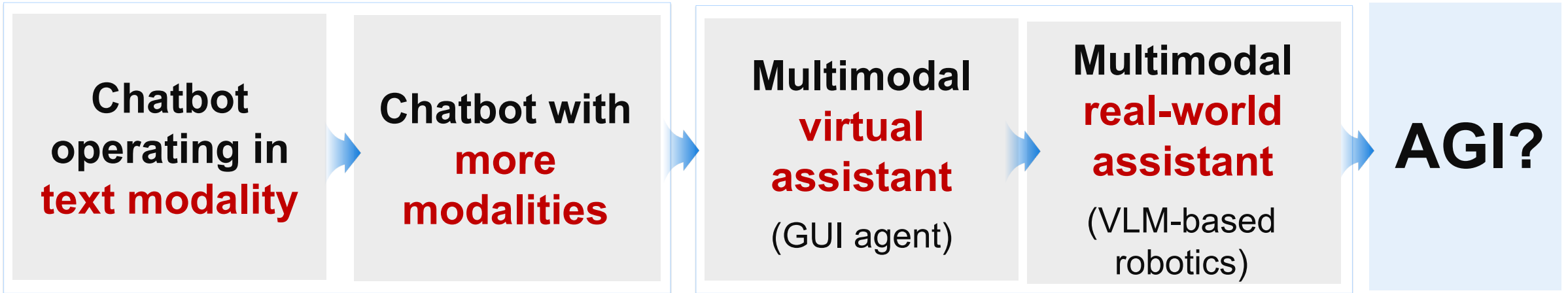
—Toward AGI

From LLM to AGI



Passively invoked

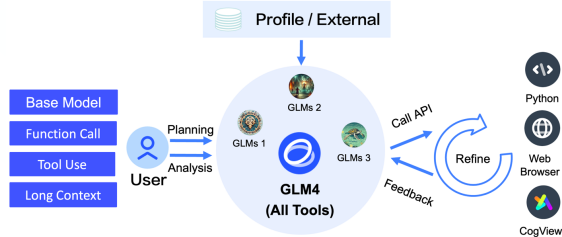
Proactive



GLM-4.5 and Beyond

SuperIntelligence and SuperAlignment

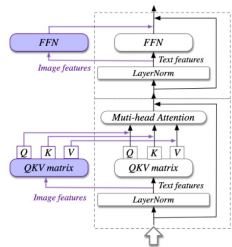
LLM OS for solving more complex tasks



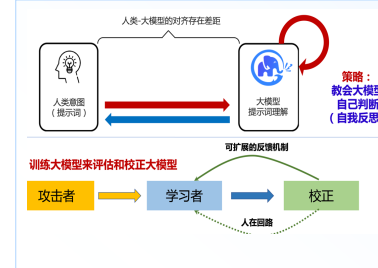
Tech.

AGI vs. ASI

Native Multimodal LLM

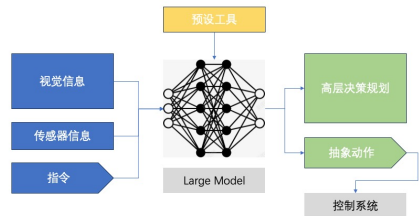


Super Alignment AI Security



Theory

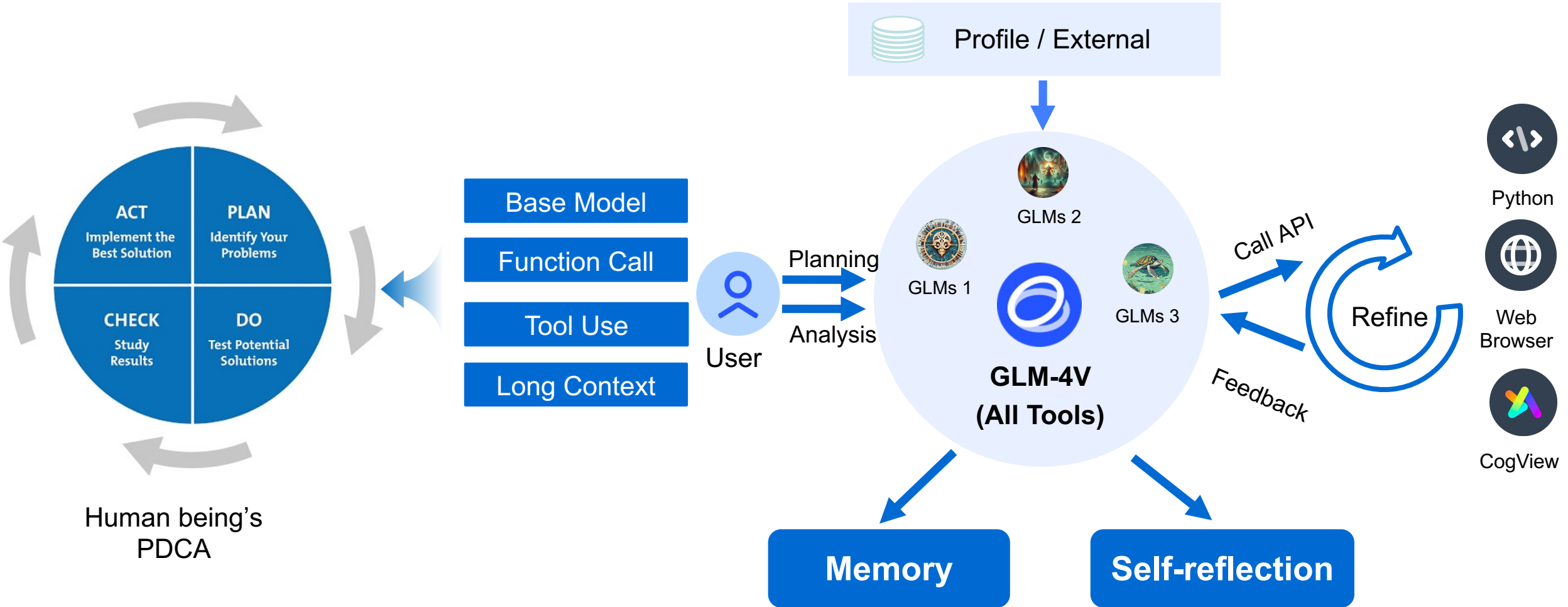
GLM, ChatGLM, GLM-4V (All-Tools)



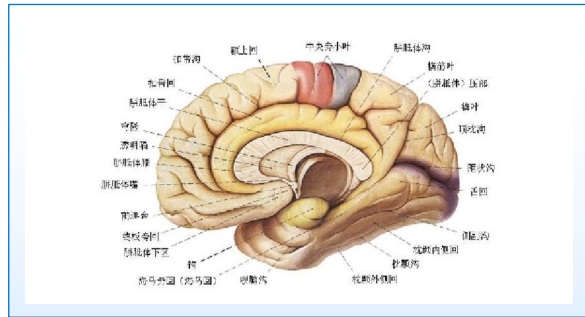
LLM

GLM-OS

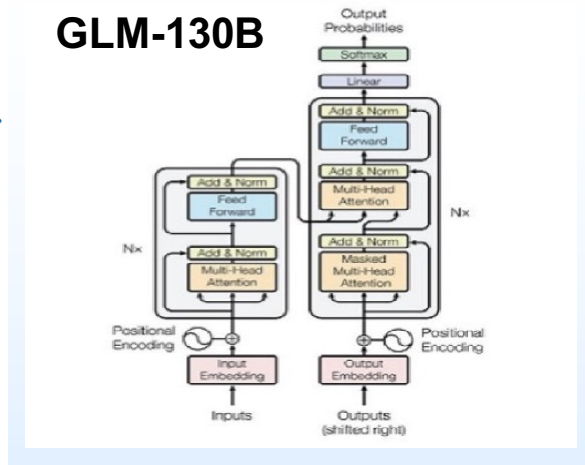
GLM OS: LLM-centric General Computing System



GLM-zero (Our Plan since 2019)



LLM

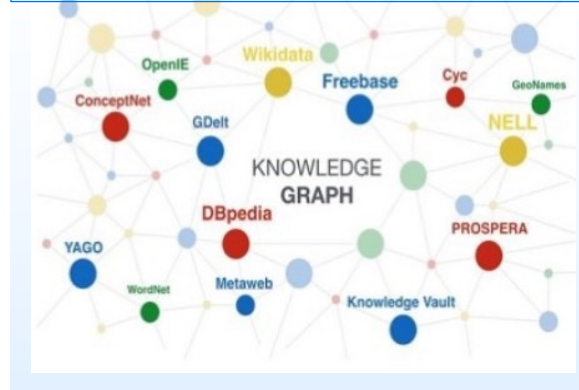


System 1

Short-term Memory
Long-term Memory

Memory (Knowledge)

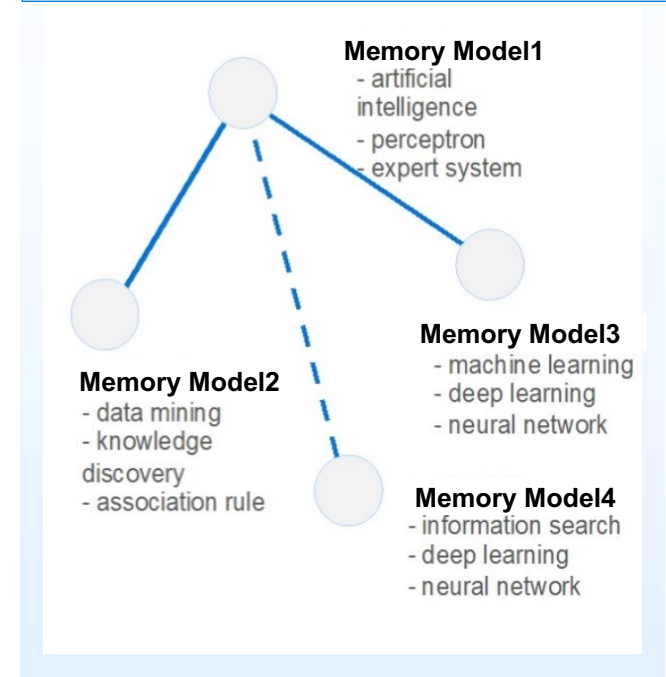
CodeGeeX, QA, Web, KG...



System 2

Cognition

- “Unconscious” learning**
- **Self-instruct**
 - **Self-reflection**
 - **Self-critics**



GLM-zero

Real World



Summary

GPT vs GLM

Chat**GPT** vs. Chat**GLM**

DALL.E vs. CogView

Codex vs. CodeGeeX

Web**GPT** vs. WebGLM

GPT-4V vs. **GLM**-4V

GPTs vs. **GLMs** (All Tools)

2024-toward AGI

GLM-4.5

GLM-OS

GLM-zero

References

- Wenyi Hong, Weihan Wang, Qingsong Lv, Jiazheng Xu, Wenmeng Yu, Junhui Ji, Yan Wang, Zihan Wang, Yuxuan Zhang, Juanzi Li, Bin Xu, Yuxiao Dong, Ming Ding, Jie Tang. **CogAgent**: A Visual Language Model for GUI Agents. CVPR'24.
- Jiayan Teng, Wendi Zheng, Ming Ding, Wenyi Hong, Jianqiao Wangni, Zhuoyi Yang, and Jie Tang. **Relay Diffusion**: Unifying diffusion process across resolutions for image synthesis. ICLR'24.
- Qinkai Zheng, Xiao Xia, Xu Zou, Yuxiao Dong, Shan Wang, Yufei Xue, Lei Shen, Zihan Wang, Andi Wang, Yang Li, Teng Su, Zhilin Yang, and Jie Tang. **CodeGeeX**: A Pre-Trained Model for Code Generation with Multilingual Benchmarking on HumanEval-X. KDD'23.
- Xiao Liu, Hanyu Lai, Yu Hao, Yifan Xu, Aohan Zeng, Zhengxiao Du, Peng Zhang, Yuxiao Dong, and Jie Tang. **WebGLM**: Towards An Efficient Web-enhanced Question Answering System with Human Preference. KDD'23.
- Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, Weng Lam Tam, Zixuan Ma, Yufei Xue, Jidong Zhai, Wenguang Chen, Zhiyuan Liu, Peng Zhang, Yuxiao Dong, and Jie Tang. **GLM-130B**: An Open Bilingual Pre-trained Model. ICLR'23.
- Xiao Liu, Kaixuan Ji, Yicheng Fu, Weng Tam, Zhengxiao Du, Zhilin Yang, Jie Tang. **P-Tuning**: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. ACL'22.
- Weihan Wang, Qingsong Lv, Wenmeng Yu, Wenyi Hong, Ji Qi, Yan Wang, Junhui Ji, Zhuoyi Yang, Lei Zhao, Xixuan Song, Jiazheng Xu, Bin Xu, Juanzi Li, Yuxiao Dong, Ming Ding, Jie Tang. **CogVlm**: Visual expert for pretrained language models.
- Wenyi Hong, Ming Ding, Wendi Zheng, Xinghan Liu, and Jie Tang. **CogVideo**: Large-scale Pretraining for Text-to-Video Generation via Transformers. ICLR'23.
- Ming Ding, Wendi Zheng, Wenyi Hong, and Jie Tang. **CogView2**: Faster and Better Text-to-Image Generation via Hierarchical Transformers. NeurIPS'22.
- Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. **GLM**: General Language Model Pretraining with Autoregressive Blank Infilling. ACL'21.
- Zixuan Ma, Jiaao He, Jiezhong Qiu, Huanqi Cao, Yuanwei Wang, Zhenbo Sun, Liyan Zheng, Haojie Wang, Shizhi Tang, Tianyu Zheng, Junyang Lin, Guanyu Feng, Zeqiang Huang, Jie Gao, Aohan Zeng, JianWei Zhang, Runxin Zhong, Tianhui Shi, Sha Liu, Weimin Zheng, Jie Tang, Hongxia Yang, Xin Liu, Jidong Zhai, and Wenguang Chen. **BAGUALU**: Targeting Brain Scale Pretrained Models with over 37 Million Cores. PPOPP'22.
- Ming Ding, Zhuoyi Yang, Wenyi Hong, Wendi Zheng, Chang Zhou, Da Yin, Junyang Lin, Xu Zou, Zhou Shao, Hongxia Yang, and Jie Tang. **CogView**: Mastering Text-to-Image Generation via Transformers. NeurIPS'21.



Thank you !

Many many collaborators from **Tsinghua and **Zhipu AI**!**

Welcome to join us!

We are hiring at all LLM roles / levels!

talent@zhipuai.cn