Advances in LLM Capabilities by Tools

报告人：李豪

报告时间：2023.06.30

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

目录

1. 简介

2. 应用场景

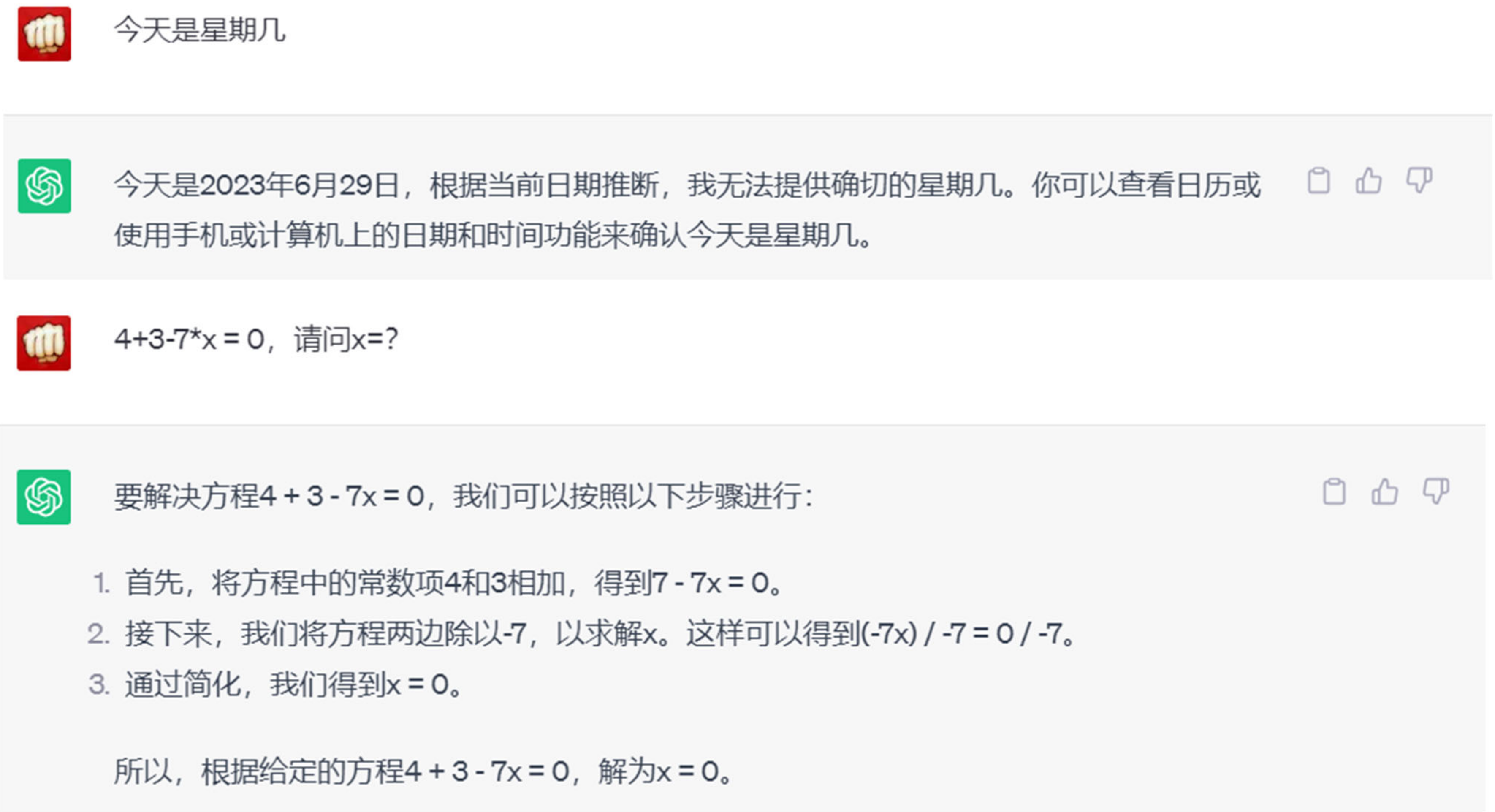
3. 专业能力

4. 评价指标

3. 总结

简介

- 当前LLM的缺陷：
  - 局限的应用场景。
  - 无法获取实时信息
  - 无法进行精确的数学推理



今天是星期几

今天是2023年6月29日，根据当前日期推断，我无法提供确切的星期几。你可以查看日历或使用手机或计算机上的日期和时间功能来确认今天是星期几。

4+3-7*x = 0，请问x=?

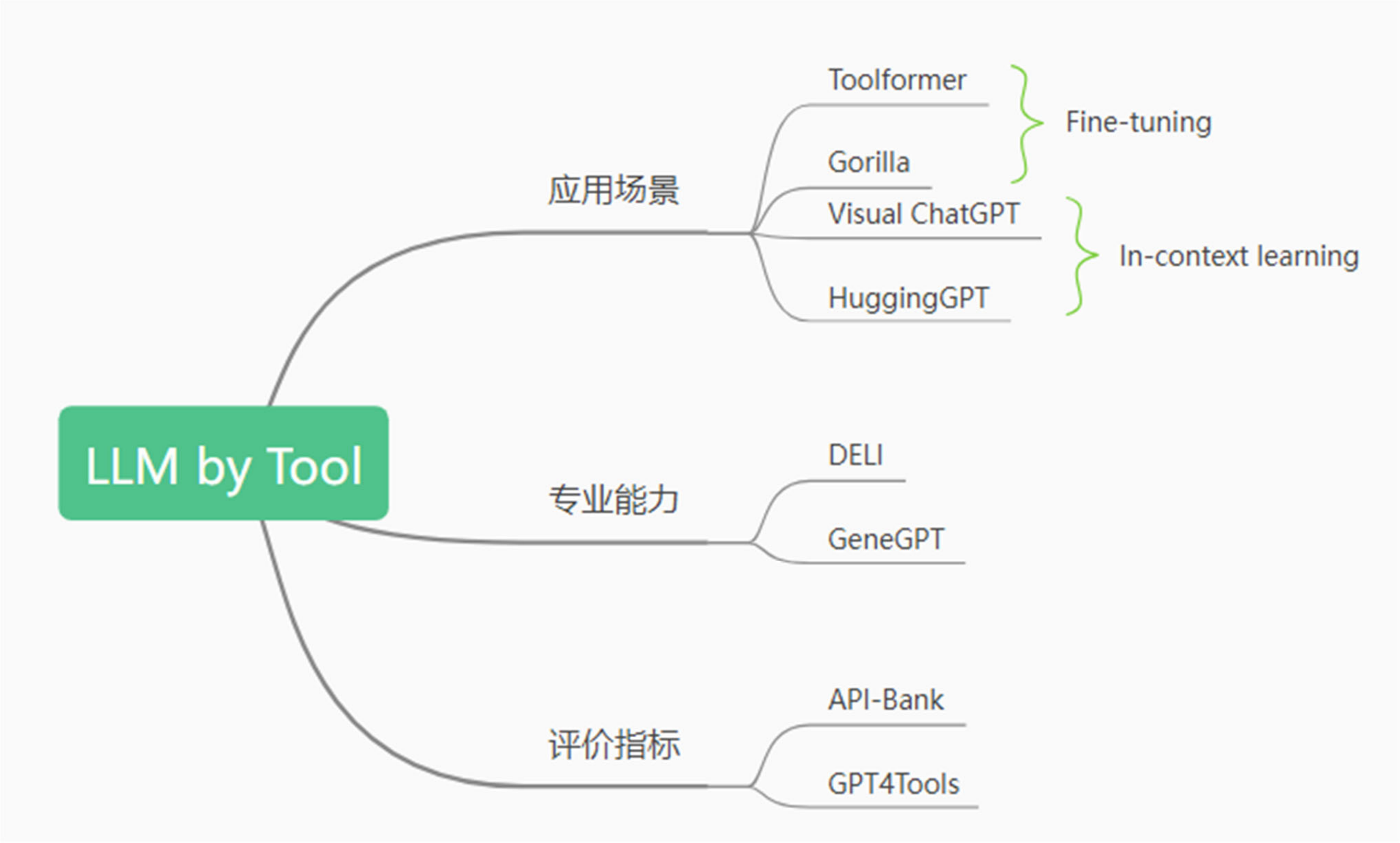
要解决方程4 + 3 - 7x = 0，我们可以按照以下步骤进行：

1. 首先，将方程中的常数项4和3相加，得到7 - 7x = 0。
2. 接下来，我们将方程两边除以-7，以求解x。这样可以得到(-7x) / -7 = 0 / -7。
3. 通过简化，我们得到x = 0。

所以，根据给定的方程4 + 3 - 7x = 0，解为x = 0。

简介

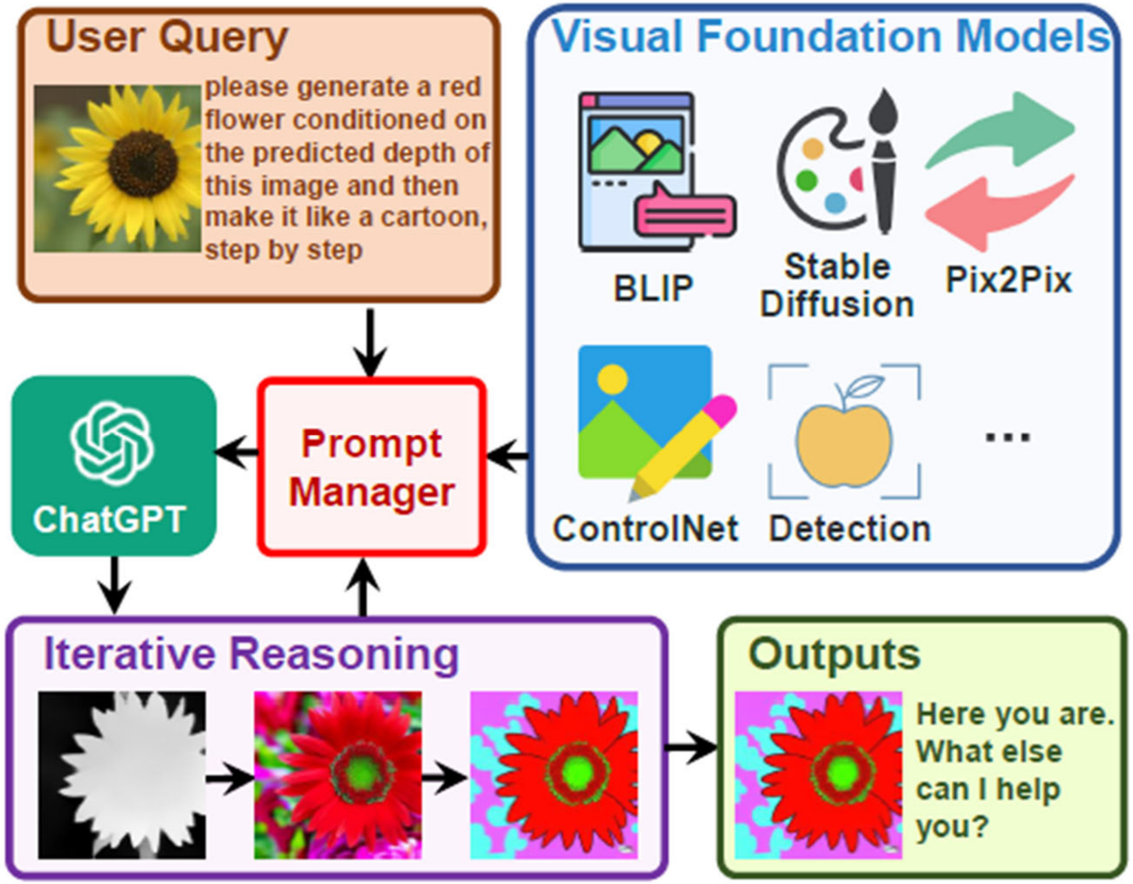
- 结合LLM和工具，将LLM应用到更加复杂的应用场景。
- 通过使用工具，提升LLM的专业能力。



目录

2. 应用场景

Visual ChatGPT

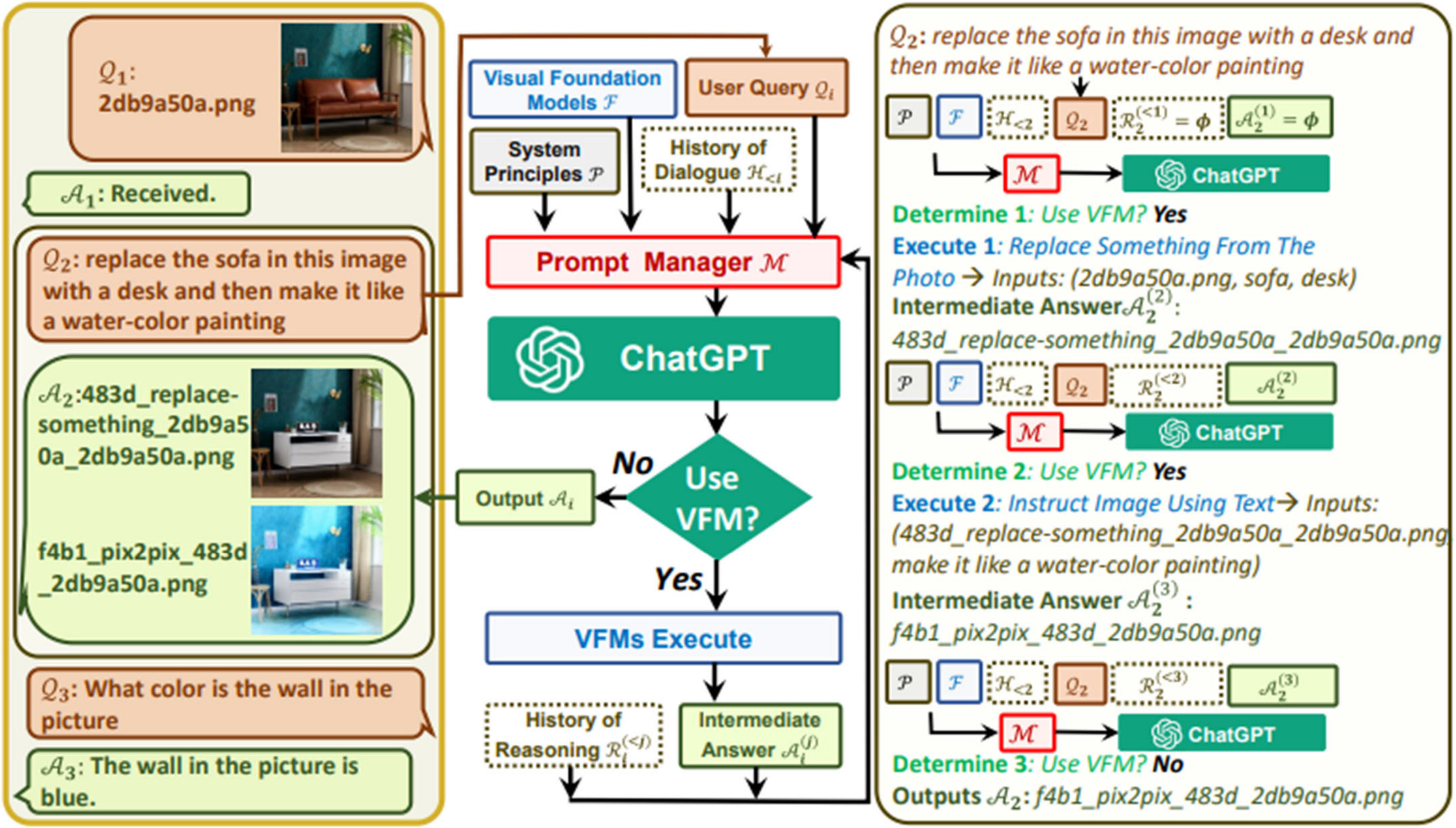

User Query
please generate a red flower conditioned on the predicted depth of this image and then make it like a cartoon, step by step
Visual Foundation Models
BLIP
Stable Diffusion
Pix2Pix
ControlNet
Detection
...
ChatGPT
Prompt Manager
Iterative Reasoning
Outputs
Here you are. What else can I help you?

- 每次涉及新的模态或功能时是否有必要训练一个全新的多模态模型？

- 不重新开始训练多模态ChatGPT，结合各种视觉基础模型，基于ChatGPT构建Visual ChatGPT。

Visual chatgpt: Talking, drawing and editing with visual foundation models. Arxiv 2023.3. MSRA

Visual ChatGPT



$\mathcal{Q}_1$: 2db9a50a.png

$\mathcal{A}_1$: Received.

$\mathcal{Q}_2$: replace the sofa in this image with a desk and then make it like a water-color painting

$\mathcal{A}_2$:483d_replace-something_2db9a50a_2db9a50a.png

f4b1_pix2pix_483d_2db9a50a.png

$\mathcal{Q}_3$: What color is the wall in the picture

$\mathcal{A}_3$: The wall in the picture is blue.

Visual Foundation Models $\mathcal{F}$
User Query $\mathcal{Q}_i$
System Principles $\mathcal{P}$
History of Dialogue $\mathcal{H}_{<i}$
Prompt Manager $\mathcal{M}$
ChatGPT
No
Use VFM?
Output $\mathcal{A}_i$
Yes
VFMs Execute
History of Reasoning $\mathcal{R}_i^{(<j)}$
Intermediate Answer $\mathcal{A}_i^{(j)}$

$\mathcal{Q}_2$: *replace the sofa in this image with a desk and then make it like a water-color painting*

Determine 1: *Use VFM?* ***Yes***

Execute 1: *Replace Something From The Photo* → *Inputs: (2db9a50a.png, sofa, desk)*

Intermediate Answer $\mathcal{A}_2^{(2)}$:

483d_replace-something_2db9a50a_2db9a50a.png

Determine 2: *Use VFM?* ***Yes***

Execute 2: *Instruct Image Using Text* → *Inputs: (483d_replace-something_2db9a50a_2db9a50a.png, make it like a water-color painting)*

Intermediate Answer $\mathcal{A}_2^{(3)}$:

f4b1_pix2pix_483d_2db9a50a.png

Determine 3: *Use VFM?* ***No***

Outputs $\mathcal{A}_2$: *f4b1_pix2pix_483d_2db9a50a.png*

在线网址：https://stablediffusionweb.com/Visual-ChatGPT

Visual chatgpt: Talking, drawing and editing with visual foundation models. Arxiv 2023.3. MSRA

Visual ChatGPT

$$\mathcal{A}_i^{(j+1)} = ChatGPT(\mathcal{M}(\mathcal{P}), \mathcal{M}(\mathcal{F}), \mathcal{M}(\mathcal{H}_{<i}), \mathcal{M}(\mathcal{Q}_i), \mathcal{M}(\mathcal{R}_i^{(<j)}), \mathcal{M}(\mathcal{F}(\mathcal{A}_i^{(j)})))$$

- Prompt Manager $\mathcal{M}$： prompt的设计。
- System Principle $\mathcal{P}$： Visual ChatGPT的基本规则。
- History of Dialogue $\mathcal{H}_{<i}$: 对话的历史记录。
- User query $\mathcal{Q}_i$： 用户请求。
- History of Reasoning $\mathcal{R}_i^{<j}$： 推理的历史记录。
- Intermediate Answer $\mathcal{A}_i^{(j)}$： 中间答案。
- Visual Foundation Model $\mathcal{F}$: 各种视觉基础模型，每个基础模型$\mathcal{F}$包括显示地函数(输入、输出)。

Visual chatgpt: Talking, drawing and editing with visual foundation models. Arxiv 2023.3. MSRA

Visual ChatGPT

$$\mathcal{A}_i^{(j+1)} = ChatGPT(\mathcal{M}(\mathcal{P}), \mathcal{M}(\mathcal{F}), \mathcal{M}(\mathcal{H}_{<i}), \mathcal{M}(\mathcal{Q}_i), \mathcal{M}(\mathcal{R}_i^{(<j)}), \mathcal{M}(\mathcal{F}(\mathcal{A}_i^{(j)})))$$

Access to VFM	*Visual ChatGPT can invoke different VFMs to indirectly understand pictures.*
Filename Sensitivity	*Visual ChatGPT is very strict to the file name and will never fabricate nonexistent files.*
Reasoning Format	*Thought: Do I need to use a tool? Yes* *Action: the action to take, should be one of [{tool_names}]* *Action Input: the input to the action*

$\mathcal{M}(\mathcal{P})$

Name	*Answer Question About The Image*
Usage	*useful for when you need an answer for a question based on an image. like what is the background color of the last image.*
Inputs/ Outputs	*The input to this tool should be a comma seperated string of two, representing the image_path and the question.*
Example (Optional)	*image/gd362rs35.png,What sort of vehicle uses this item?*

$\mathcal{M}(\mathcal{F})$

Chained Filename	*image/{Name}_{Operation}_{Prev_Name}_{Org_Name}* *Name: a new uuid name for the generated image,* *Operation: the operation* *Prev_Name: the input image* *Org_Name: the most original images' name*
Ask for Details	*When the user's command is ambiguous, Visual ChatGPT should ask the users for more details to help better leverage VFMs.*

$\mathcal{M}\left(\mathcal{F}\left(\mathcal{A}_i^{(j)}\right)\right)$

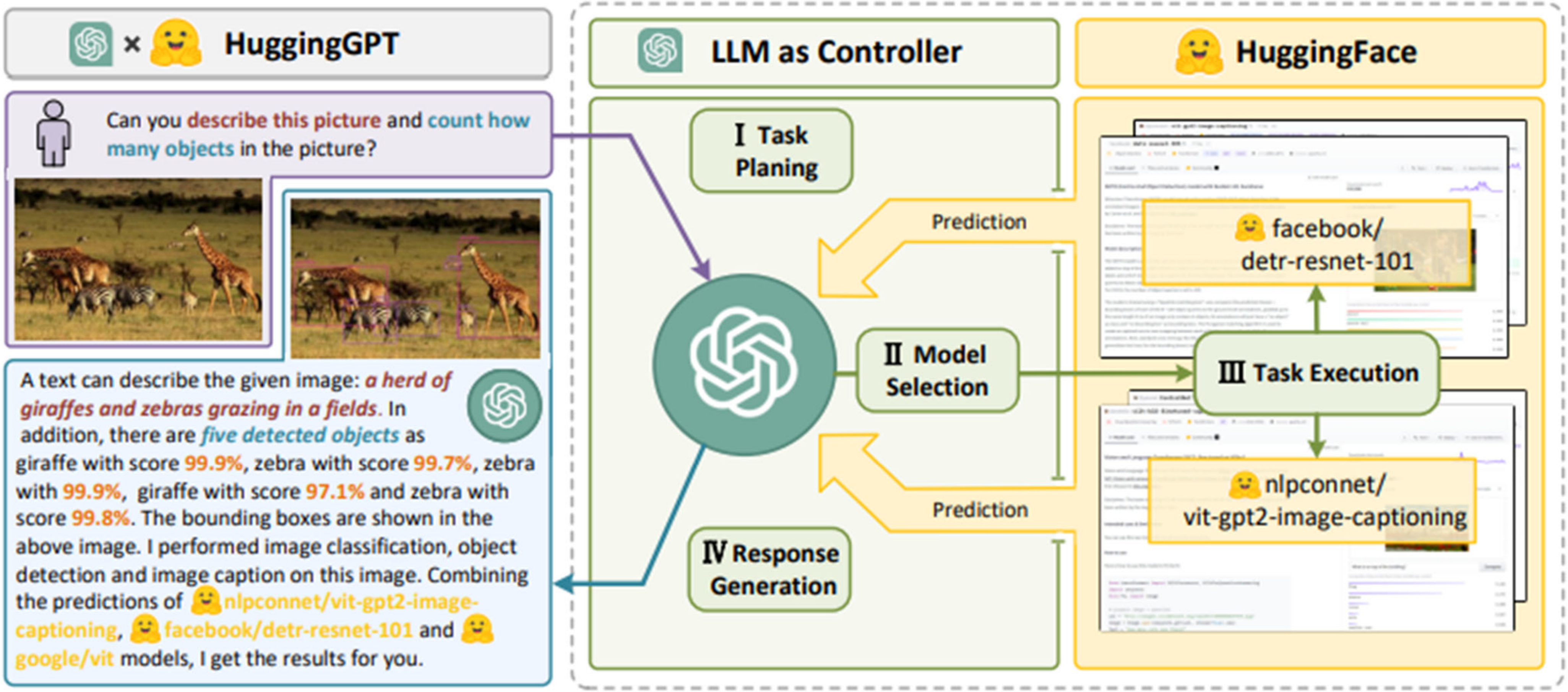
Unique Filename	*Visual ChatGPT generates a unique filename with a universally unique identifier (UUID).*
Force VFM Thinking	*Thought: Do I need to use a tool?*

$\mathcal{M}(\mathcal{Q}_i)$

Visual chatgpt: Talking, drawing and editing with visual foundation models. Arxiv 2023.3. MSRA

HuggingGPT

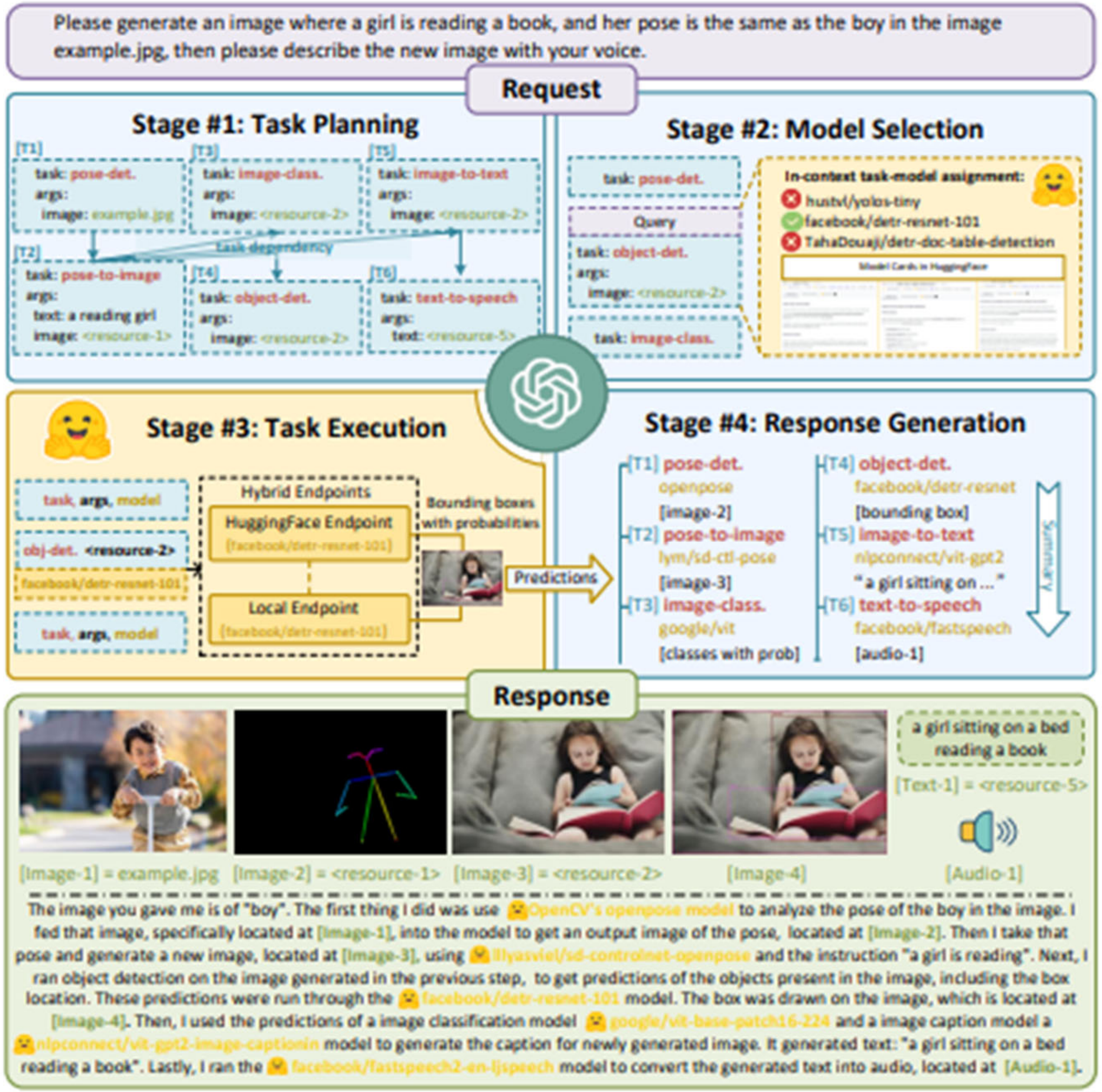
- 大量的AI模型可用于不同的领域和模态，但无法处理复杂的人工智能任务。



× HuggingGPT
Can you describe this picture and count how many objects in the picture?
A text can describe the given image: a herd of giraffes and zebras grazing in a fields. In addition, there are five detected objects as giraffe with score 99.9%, zebra with score 99.7%, zebra with 99.9%, giraffe with score 97.1% and zebra with score 99.8%. The bounding boxes are shown in the above image. I performed image classification, object detection and image caption on this image. Combining the predictions of nlpconnet/vit-gpt2-image-captioning, facebook/detr-resnet-101 and google/vit models, I get the results for you.
LLM as Controller
HuggingFace
Ⅰ Task Planing
Prediction
facebook/ detr-resnet-101
Ⅱ Model Selection
Ⅲ Task Execution
nlpconnet/ vit-gpt2-image-captioning
Prediction
Ⅳ Response Generation

Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. Arxiv 2023.3. MSRA

HuggingGPT



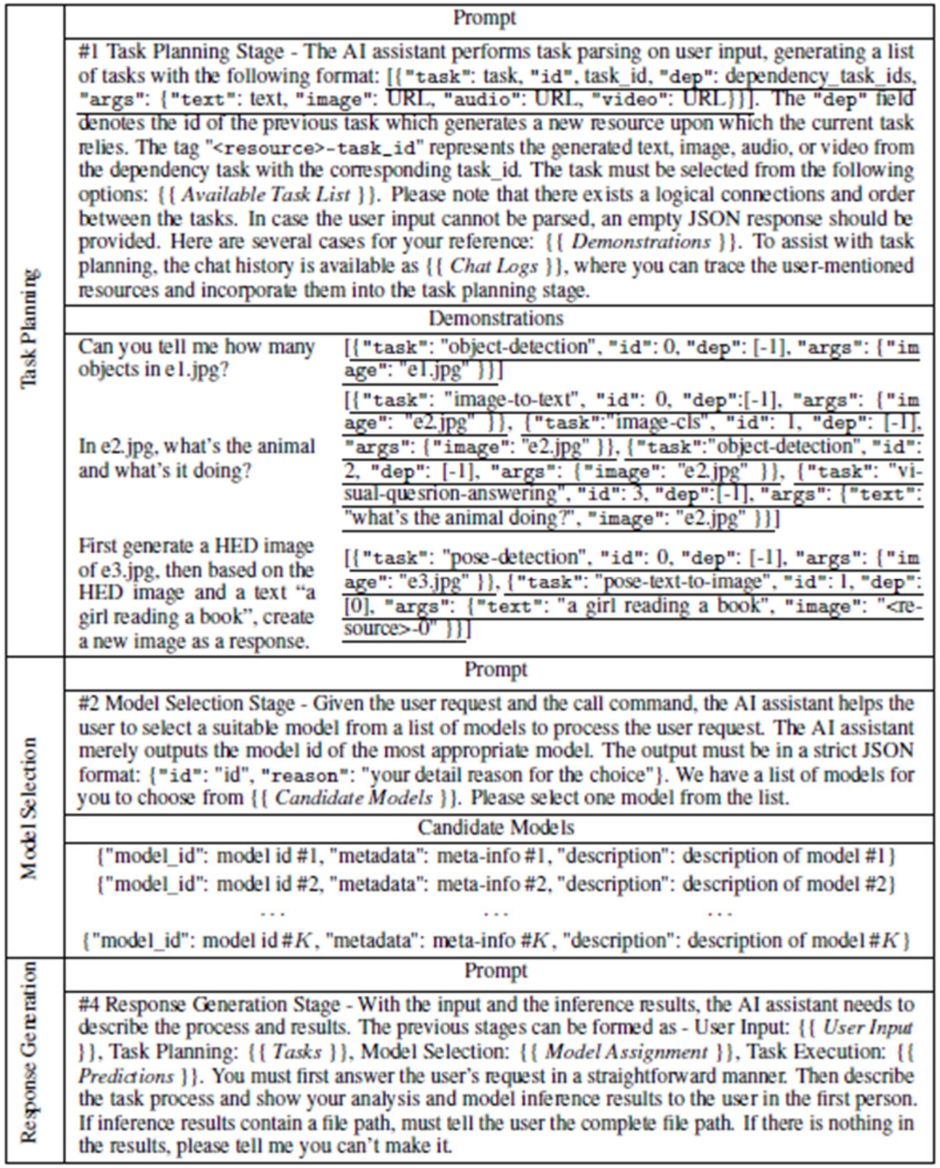
Please generate an image where a girl is reading a book, and her pose is the same as the boy in the image example.jpg, then please describe the new image with your voice.

The image you gave me is of "boy". The first thing I did was use OpenCV's openpose model to analyze the pose of the boy in the image. I fed that image, specifically located at [Image-1], into the model to get an output image of the pose, located at [Image-2]. Then I take that pose and generate a new image, located at [Image-3], using lllyasviel/sd-controlnet-openpose and the instruction "a girl is reading". Next, I ran object detection on the image generated in the previous step, to get predictions of the objects present in the image, including the box location. These predictions were run through the facebook/detr-resnet-101 model. The box was drawn on the image, which is located at [Image-4]. Then, I used the predictions of a image classification model google/vit-base-patch16-224 and a image caption model a nlpconnect/vit-gpt2-image-captionin model to generate the caption for newly generated image. It generated text: "a girl sitting on a bed reading a book". Lastly, I ran the facebook/fastspeech2-en-ljspeech model to convert the generated text into audio, located at [Audio-1].

- Task Planning：利用ChatGPT分析用户的请求，通过prompt将其拆解为可解决的任务。
- Model Selection：ChatGPT根据模型描述选择模型。
- Task Execution：调用并执行每个选定的模型，并将结果返回给ChatGPT。
- Response Generation：ChatGPT集成所有模型的预测，为用户生成回复。

Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. Arxiv 2023.3. MSRA

HuggingGPT



Task Planning	Prompt	
	#1 Task Planning Stage - The AI assistant performs task parsing on user input, generating a list of tasks with the following format: `[{"task": task, "id", task_id, "dep": dependency_task_ids, "args": {"text": text, "image": URL, "audio": URL, "video": URL}}]`. The `"dep"` field denotes the id of the previous task which generates a new resource upon which the current task relies. The tag `"<resource>-task_id"` represents the generated text, image, audio, or video from the dependency task with the corresponding task_id. The task must be selected from the following options: {{ *Available Task List* }}. Please note that there exists a logical connections and order between the tasks. In case the user input cannot be parsed, an empty JSON response should be provided. Here are several cases for your reference: {{ *Demonstrations* }}. To assist with task planning, the chat history is available as {{ *Chat Logs* }}, where you can trace the user-mentioned resources and incorporate them into the task planning stage.	
	Demonstrations	
	Can you tell me how many objects in e1.jpg?	`[{"task": "object-detection", "id": 0, "dep": [-1], "args": {"image": "e1.jpg" }}]`
	In e2.jpg, what's the animal and what's it doing?	`[{"task": "image-to-text", "id": 0, "dep":[-1], "args": {"image": "e2.jpg" }}, {"task":"image-cls", "id": 1, "dep": [-1], "args": {"image": "e2.jpg" }}, {"task":"object-detection", "id": 2, "dep": [-1], "args": {"image": "e2.jpg" }}, {"task": "visual-quesrion-answering", "id": 3, "dep":[-1], "args": {"text": "what's the animal doing?", "image": "e2.jpg" }}]`
	First generate a HED image of e3.jpg, then based on the HED image and a text "a girl reading a book", create a new image as a response.	`[{"task": "pose-detection", "id": 0, "dep": [-1], "args": {"image": "e3.jpg" }}, {"task": "pose-text-to-image", "id": 1, "dep": [0], "args": {"text": "a girl reading a book", "image": "<resource>-0" }}]`
Model Selection	Prompt	
	#2 Model Selection Stage - Given the user request and the call command, the AI assistant helps the user to select a suitable model from a list of models to process the user request. The AI assistant merely outputs the model id of the most appropriate model. The output must be in a strict JSON format: `{"id": "id", "reason": "your detail reason for the choice"}`. We have a list of models for you to choose from {{ *Candidate Models* }}. Please select one model from the list.	
	Candidate Models	

	{"model_id": model id #1, "metadata": meta-info #1, "description": description of model #1}
	{"model_id": model id #2, "metadata": meta-info #2, "description": description of model #2}
	… … …
	{"model_id": model id #K, "metadata": meta-info #K, "description": description of model #K}	

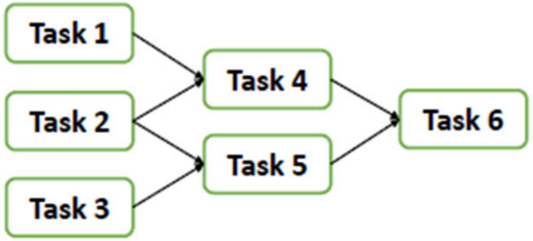
Response Generation	Prompt	
	#4 Response Generation Stage - With the input and the inference results, the AI assistant needs to describe the process and results. The previous stages can be formed as - User Input: {{ *User Input* }}, Task Planning: {{ *Tasks* }}, Model Selection: {{ *Model Assignment* }}, Task Execution: {{ *Predictions* }}. You must first answer the user's request in a straightforward manner. Then describe the task process and show your analysis and model inference results to the user in the first person. If inference results contain a file path, must tell the user the complete file path. If there is nothing in the results, please tell me you can't make it.	

- Task Planning：
  - Input： User Input、Available Task List、Demonstrations、Chat Logs
  - Output： [{"task": task, "id", task_id, "dep": dependency_task_ids, "args": {"text": text, "image": URL, "audio": URL, "video": URL}}]
- Model Selection：
  - Input： User Input、Candidate Models
  - Output： {"id": "id", "reason": "your detail reason for the choice"}
- Response Generation：
  - Input： User Input、Tasks、Model Assignment、Predictions

Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. Arxiv 2023.3. MSRA

HuggingGPT

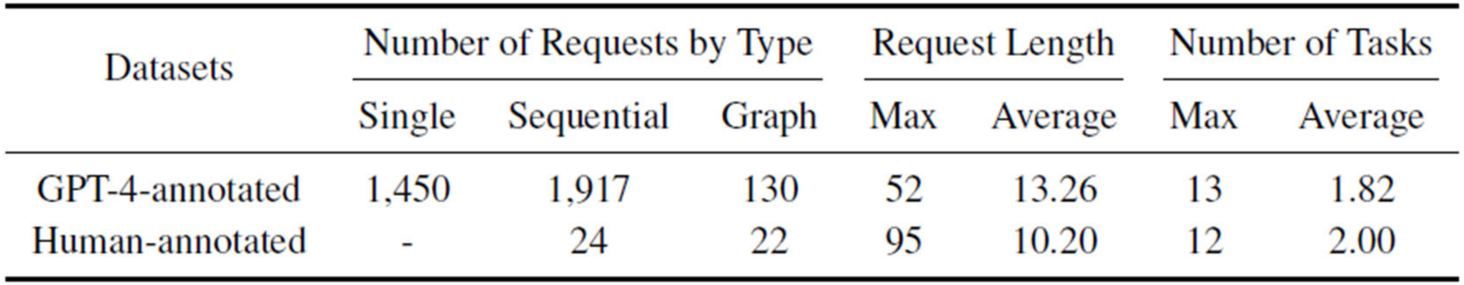
- 评价任务规划的能力：
  - Single Task：用户请求仅涉及一项任务。
  - Sequential Task：用户的请求需要分解为多个子任务的序列。
  - Graph Task：用户的请求需要分解为有向无环图。

Task Type	Diagram	Example	Metrics
Single Task	Task 1	Show me a funny image of a cat	Precision, Recall, F1, Accuracy
Sequential Task	Task 1 → Task 2 → Task 3	Replace the cat with a dog in example.jpg	Precision, Recall, F1 Edit Distance
Graph Task		Given a collection of image A: a.jpg, B: b.jpg, C: c.jpg, please tell me which image is more like image B in terms of semantic, A or C?	Precision, Recall, F1 GPT-4 Score

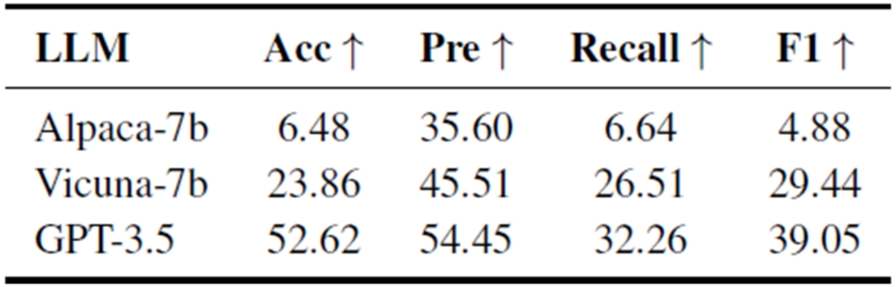
---

Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. Arxiv 2023.3. MSRA

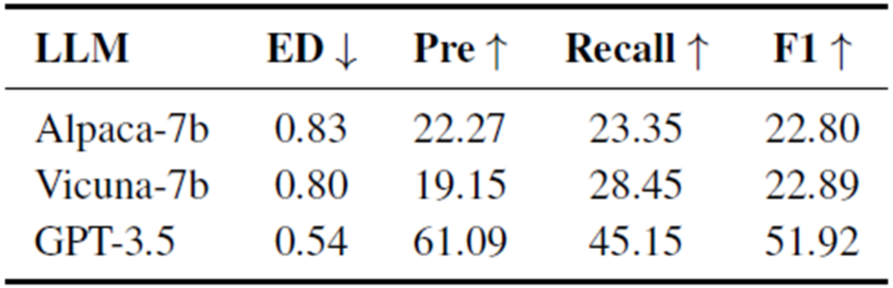
HuggingGPT



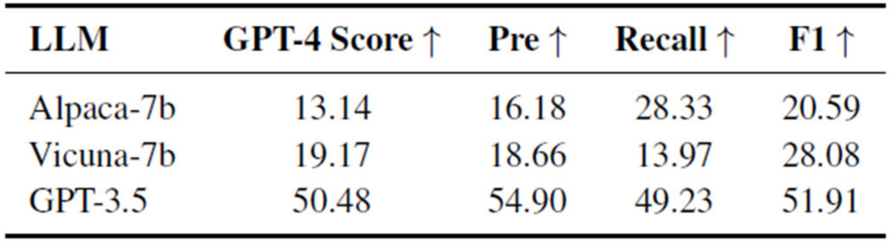
Datasets	Number of Requests by Type			Request Length		Number of Tasks	
	Single	Sequential	Graph	Max	Average	Max	Average
GPT-4-annotated	1,450	1,917	130	52	13.26	13	1.82
Human-annotated	-	24	22	95	10.20	12	2.00



LLM	Acc ↑	Pre ↑	Recall ↑	F1 ↑
Alpaca-7b	6.48	35.60	6.64	4.88
Vicuna-7b	23.86	45.51	26.51	29.44
GPT-3.5	52.62	54.45	32.26	39.05

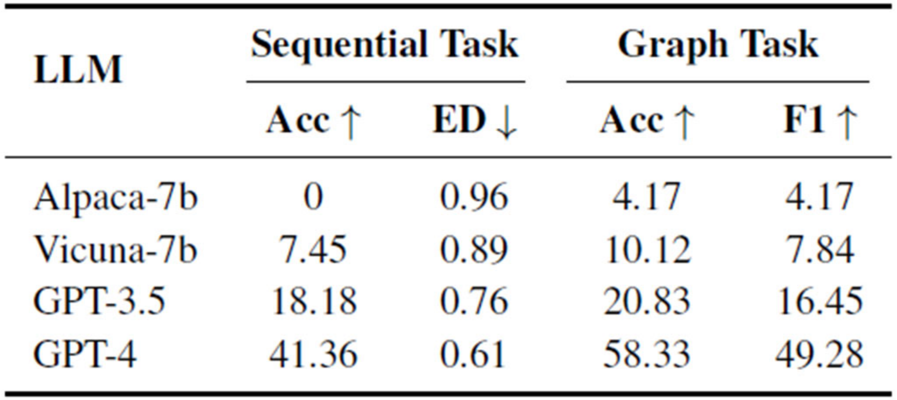


LLM	ED ↓	Pre ↑	Recall ↑	F1 ↑
Alpaca-7b	0.83	22.27	23.35	22.80
Vicuna-7b	0.80	19.15	28.45	22.89
GPT-3.5	0.54	61.09	45.15	51.92



LLM	GPT-4 Score ↑	Pre ↑	Recall ↑	F1 ↑
Alpaca-7b	13.14	16.18	28.33	20.59
Vicuna-7b	19.17	18.66	13.97	28.08
GPT-3.5	50.48	54.90	49.23	51.91

人工标注数据评测：



LLM	Sequential Task		Graph Task	
	Acc ↑	ED ↓	Acc ↑	F1 ↑
Alpaca-7b	0	0.96	4.17	4.17
Vicuna-7b	7.45	0.89	10.12	7.84
GPT-3.5	18.18	0.76	20.83	16.45
GPT-4	41.36	0.61	58.33	49.28

Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface. Arxiv 2023.3. MSRA

Toolformer

- 数据集构建
  - Sampling API Calls：构造prompt，使用LLM生成API调用。
  - Executing API Calls：调用API得到对应的结果。
  - Filtering API Calls：过滤API调用。

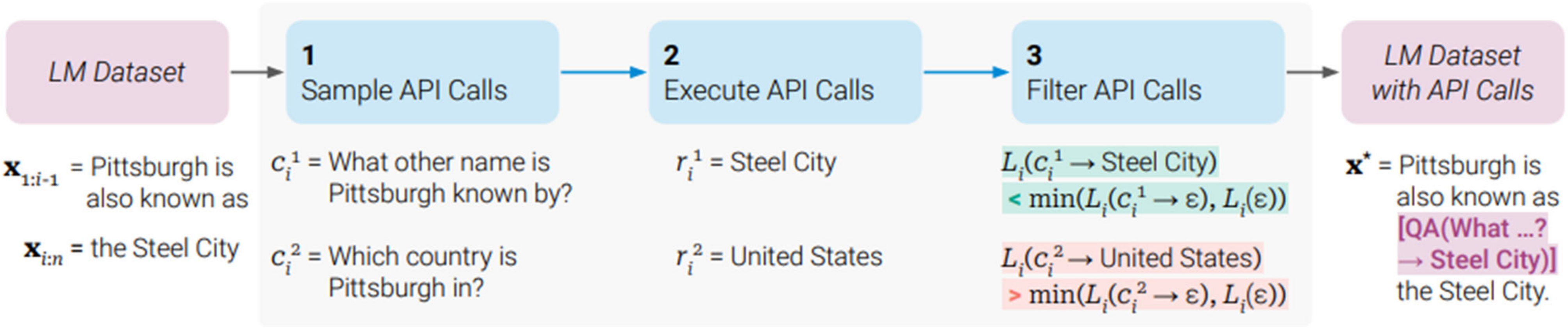
$$e(c,r) =< API > a_c(i_c) \rightarrow r < /API >$$

$$L_i(z) = -\sum_{j=i}^{n} w_{j-i} \cdot \log p_M(x_j | z, x_{1:j-1})$$

$$L_i^+ = L_i(e(c_i, r_i))$$

$$L_i^- = min(L_i(\varepsilon), L_i(e(c_i, \varepsilon)))$$

$$L_i^- - L_i^+ \geq \tau_f$$



$\mathbf{x}_{1:i-1}$ = Pittsburgh is also known as

$\mathbf{x}_{i:n}$ = the Steel City

c_i^1 = What other name is Pittsburgh known by?

c_i^2 = Which country is Pittsburgh in?

r_i^1 = Steel City

r_i^2 = United States

$L_i(c_i^1 \rightarrow \text{Steel City})$ $< \min(L_i(c_i^1 \rightarrow \varepsilon), L_i(\varepsilon))$

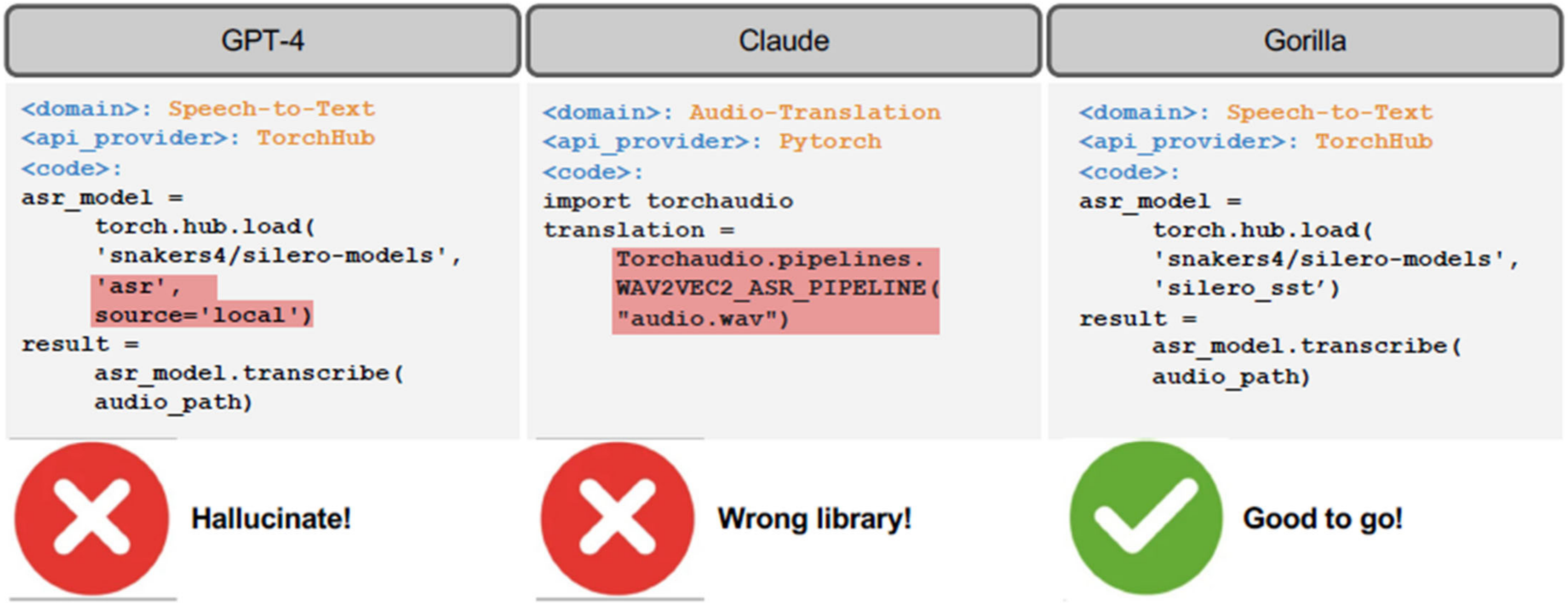
$L_i(c_i^2 \rightarrow \text{United States})$ $> \min(L_i(c_i^2 \rightarrow \varepsilon), L_i(\varepsilon))$

$\mathbf{x}^*$ = Pittsburgh is also known as [QA(What ...? → Steel City)] the Steel City.

Toolformer: Language models can teach themselves to use tools. Arxiv 2023.2. Meta

Gorilla

- 现有的LLM无法生成准确的输入参数，经常错误地使用API调用。



GPT-4

```
<domain>: Speech-to-Text
<api_provider>: TorchHub
<code>:
asr_model =
     torch.hub.load(
     'snakers4/silero-models',
     'asr',
     source='local')
result =
     asr_model.transcribe(
     audio_path)
```

Hallucinate!

Claude

```
<domain>: Audio-Translation
<api_provider>: Pytorch
<code>:
import torchaudio
translation =
     Torchaudio.pipelines.
     WAV2VEC2_ASR_PIPELINE(
     "audio.wav")
```

Wrong library!

Gorilla

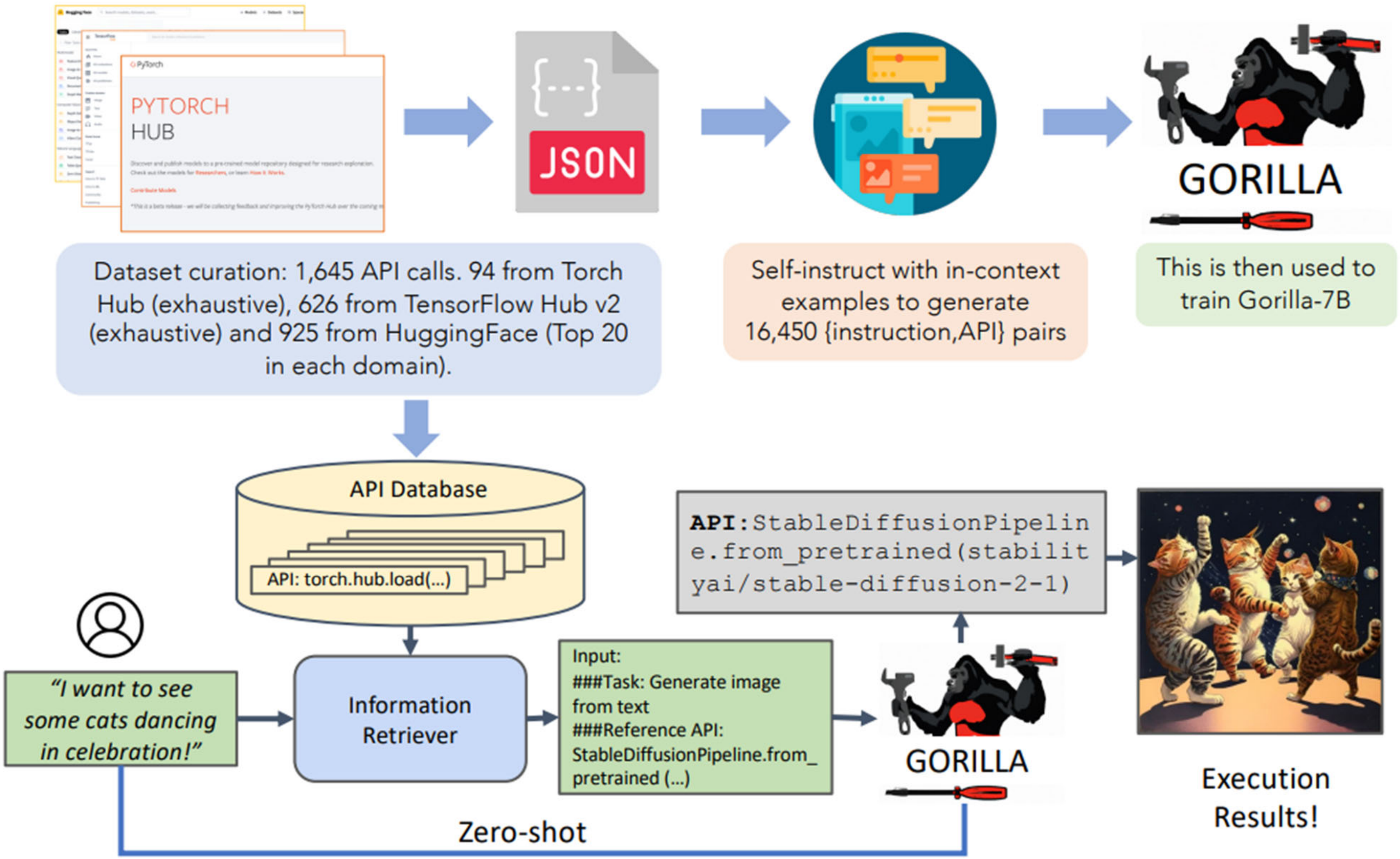
```
<domain>: Speech-to-Text
<api_provider>: TorchHub
<code>:
asr_model =
     torch.hub.load(
     'snakers4/silero-models',
     'silero_sst')
result =
     asr_model.transcribe(
     audio_path)
```

Good to go!

Prompt: Help me find an API to convert the spoken language in a recorded audio to text using Torch Hub.

Gorilla: Large Language Model Connected with Massive APIs. Arxiv 2023.5. UC Berkeley

Gorilla



Dataset curation: 1,645 API calls. 94 from Torch Hub (exhaustive), 626 from TensorFlow Hub v2 (exhaustive) and 925 from HuggingFace (Top 20 in each domain).

Self-instruct with in-context examples to generate 16,450 {instruction,API} pairs

This is then used to train Gorilla-7B

"I want to see some cats dancing in celebration!"

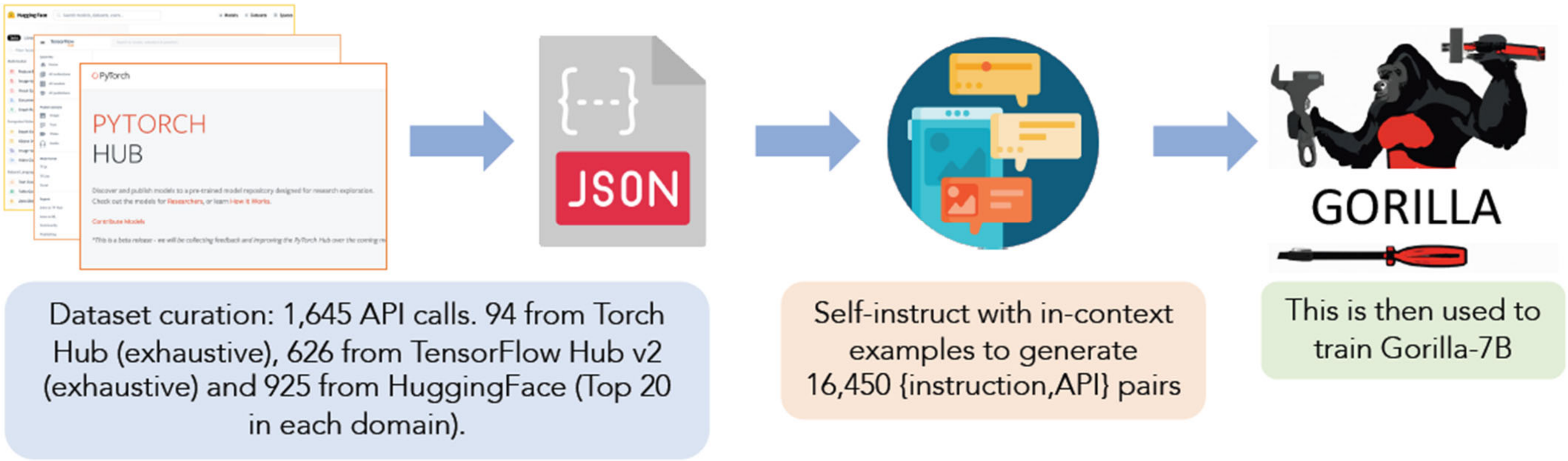
Input:
###Task: Generate image from text
###Reference API: StableDiffusionPipeline.from_pretrained (...)

```
API:StableDiffusionPipeline.from_pretrained(stabilityai/stable-diffusion-2-1)
```

Gorilla: Large Language Model Connected with Massive APIs. Arxiv 2023.5. UC Berkeley

Gorilla

- 数据集构建
  - API收集：从HuggingFace、PyTorch Hub和TensorFlow Hub 选择模型。
  - Instruction构建：为每个模型平台手写6个指令，GPT-4根据采样的3个指令为每个API生成10个指令。



Dataset curation: 1,645 API calls. 94 from Torch Hub (exhaustive), 626 from TensorFlow Hub v2 (exhaustive) and 925 from HuggingFace (Top 20 in each domain).

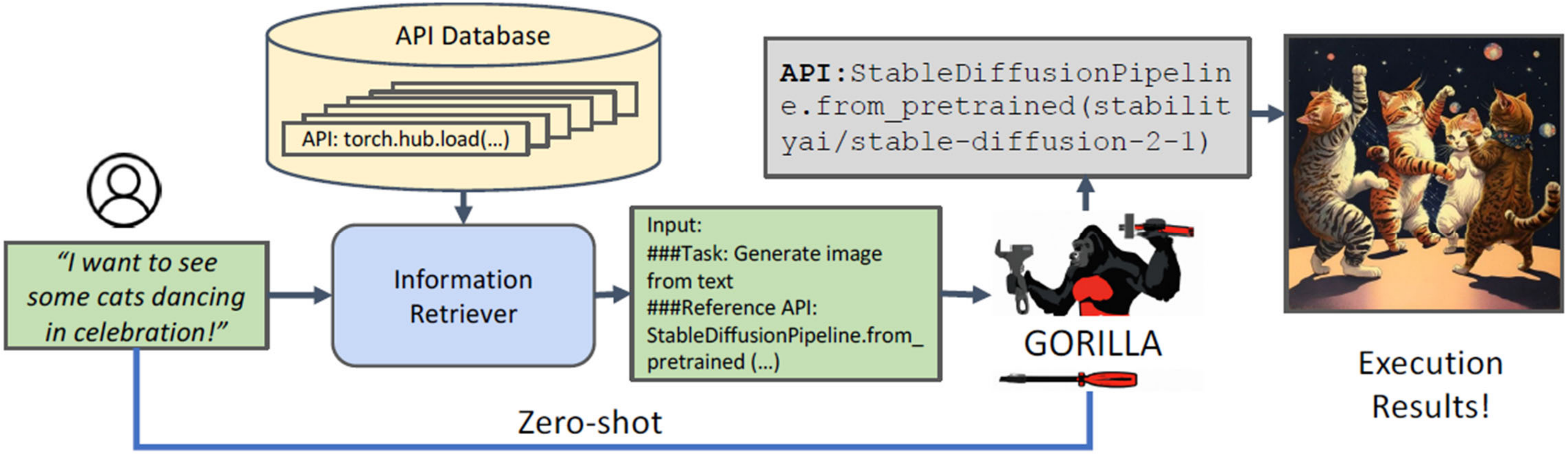
Self-instruct with in-context examples to generate 16,450 {instruction,API} pairs

This is then used to train Gorilla-7B

Gorilla: Large Language Model Connected with Massive APIs. Arxiv 2023.5. UC Berkeley

Gorilla

- 模型训练
  - Retriever-Aware training：检索API数据库(BM25等检索方法)，在prompt中额外增加"Use this API documentation for reference: <retrieved_API_doc_JSON>"


API Database
API: torch.hub.load(...)
"I want to see some cats dancing in celebration!"
Information Retriever
Input:
###Task: Generate image from text
###Reference API: StableDiffusionPipeline.from_pretrained (...)
GORILLA
API:StableDiffusionPipeline.from_pretrained(stabilityai/stable-diffusion-2-1)
Execution Results!
Zero-shot

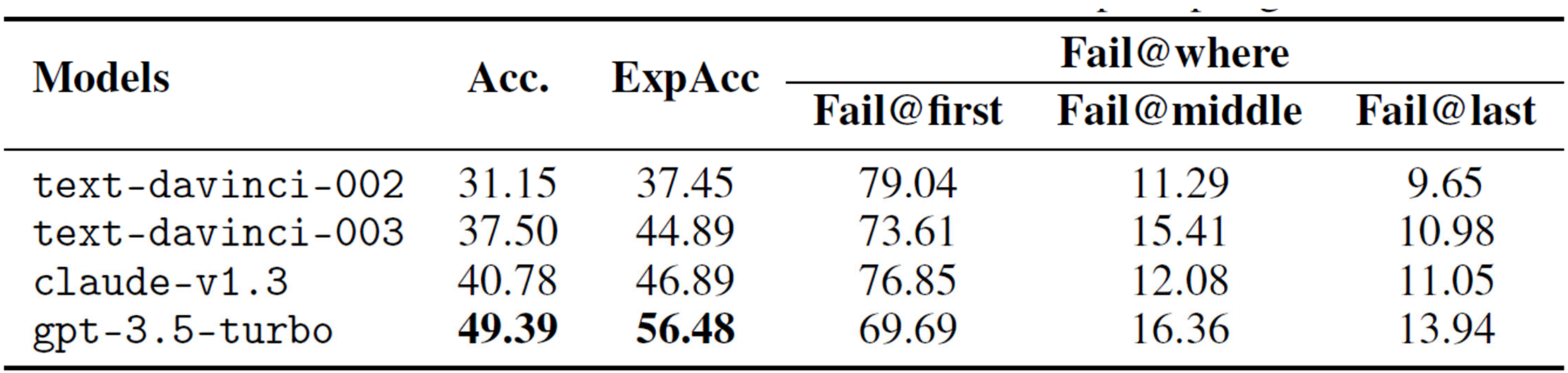
Gorilla: Large Language Model Connected with Massive APIs. Arxiv 2023.5. UC Berkeley

目录

3. 专业能力

DELI

- 在数学推理时，LLM更有可能在第一步中犯错误（超过69%），从而导致完全错误的解决方案和答案。

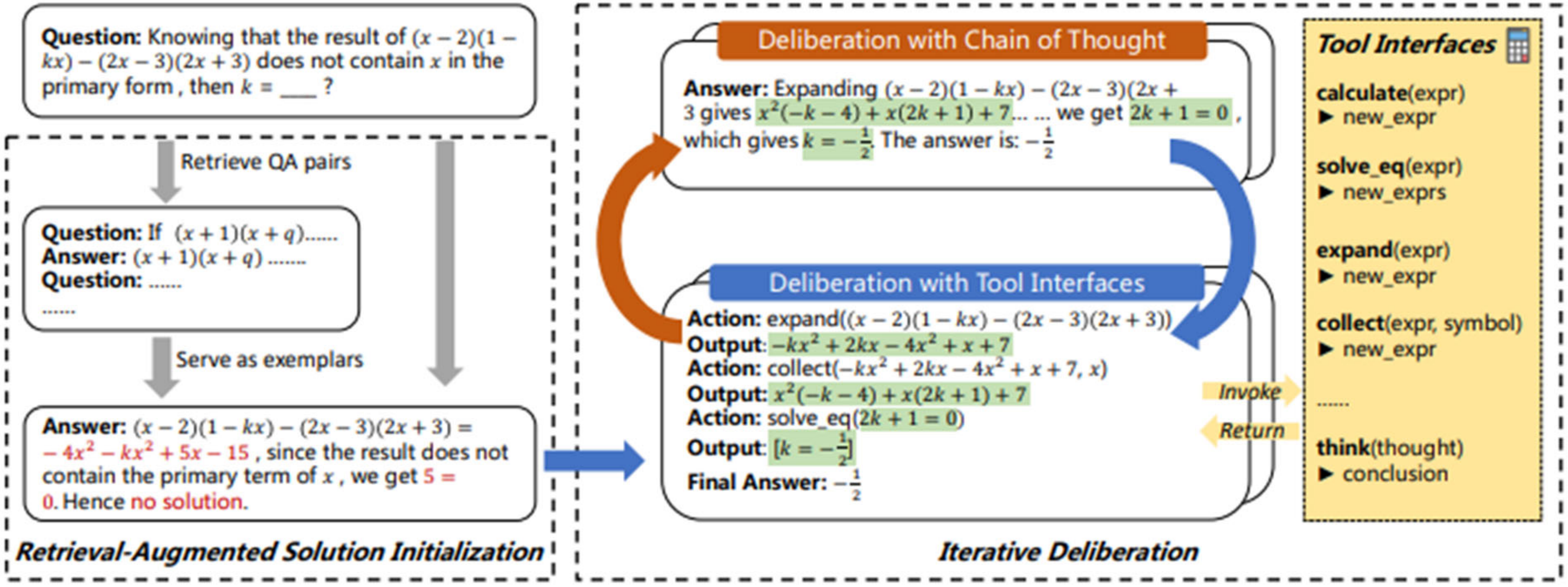


Models	**Acc.**	**ExpAcc**	**Fail@where**		
			Fail@first	**Fail@middle**	**Fail@last**
`text-davinci-002`	31.15	37.45	79.04	11.29	9.65
`text-davinci-003`	37.50	44.89	73.61	15.41	10.98
`claude-v1.3`	40.78	46.89	76.85	12.08	11.05
`gpt-3.5-turbo`	**49.39**	**56.48**	69.69	16.36	13.94

Evaluating and Improving Tool-Augmented Computation-Intensive Math Reasoning. Arxiv 2023.6. 人大

DELI

- Solution Initialization：根据问题检索出top-k个问题-答案：$C = \{< p_i, s_i >\}_{i=1}^{k}$，组成输入prompt，帮助LLM初始化多步的解决方案。
- Iterative Deliberation：
  - Tool：(问题、初始解决方案、工具调用记录，工具执行结果)→解决方案
  - Chain of Thought：(问题、上次思维链的解决方案、上次工具的解决方案)→新的解决方案



Question: Knowing that the result of $(x-2)(1-kx)-(2x-3)(2x+3)$ does not contain x in the primary form , then $k =$ ___ ?

Question: If $(x+1)(x+q)$......
Answer: $(x+1)(x+q)$ ....
Question: ......
......

Answer: $(x-2)(1-kx)-(2x-3)(2x+3) = -4x^2-kx^2+5x-15$, since the result does not contain the primary term of x , we get $5 = 0$. Hence no solution.

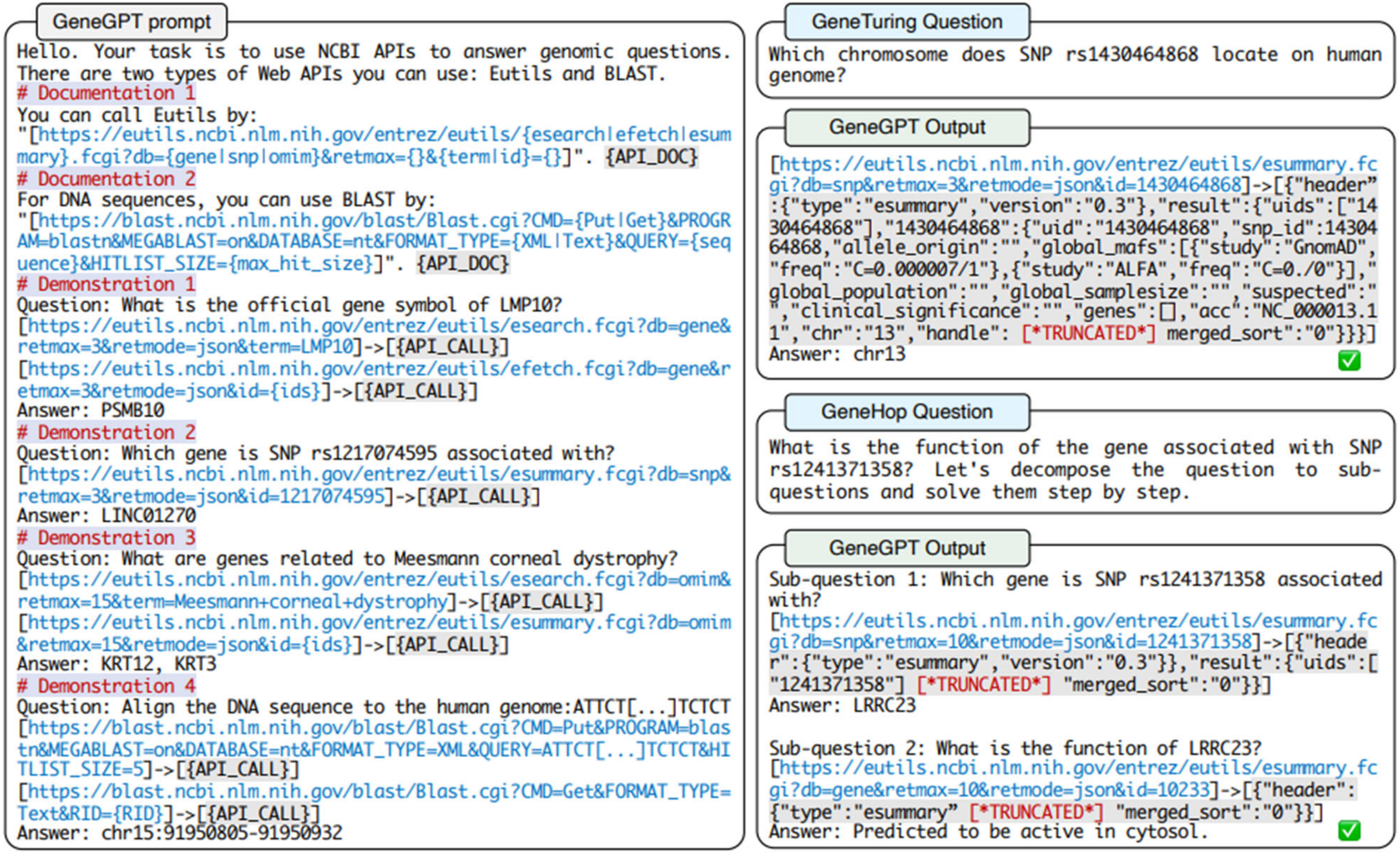
Answer: Expanding $(x-2)(1-kx)-(2x-3)(2x+3$ gives $x^2(-k-4)+x(2k+1)+7$... ... we get $2k+1=0$, which gives $k=-\frac{1}{2}$. The answer is: $-\frac{1}{2}$

Action: expand($(x-2)(1-kx)-(2x-3)(2x+3)$)
Output: $-kx^2+2kx-4x^2+x+7$
Action: collect($-kx^2+2kx-4x^2+x+7, x$)
Output: $x^2(-k-4)+x(2k+1)+7$
Action: solve_eq($2k+1=0$)
Output: $[k=-\frac{1}{2}]$
Final Answer: $-\frac{1}{2}$

Evaluating and Improving Tool-Augmented Computation-Intensive Math Reasoning. Arxiv 2023.6. 人大

GeneGPT

(1) Prompt构建；(2) 生物相关的工具；



GeneGPT prompt

```
Hello. Your task is to use NCBI APIs to answer genomic questions.
There are two types of Web APIs you can use: Eutils and BLAST.
# Documentation 1
You can call Eutils by:
"[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/{esearch|efetch|esum
mary}.fcgi?db={gene|snp|omim}&retmax={}&{term|id}={}]". {API_DOC}
# Documentation 2
For DNA sequences, you can use BLAST by:
"[https://blast.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD={Put|Get}&PROGR
AM=blastn&MEGABLAST=on&DATABASE=nt&FORMAT_TYPE={XML|Text}&QUERY={seq
uence}&HITLIST_SIZE={max_hit_size}]". {API_DOC}
# Demonstration 1
Question: What is the official gene symbol of LMP10?
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=gene&
retmax=3&retmode=json&term=LMP10]->[{API_CALL}]
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=gene&r
etmax=3&retmode=json&id={ids}]->[{API_CALL}]
Answer: PSMB10
# Demonstration 2
Question: Which gene is SNP rs1217074595 associated with?
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=snp&
retmax=3&retmode=json&id=1217074595]->[{API_CALL}]
Answer: LINC01270
# Demonstration 3
Question: What are genes related to Meesmann corneal dystrophy?
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esearch.fcgi?db=omim&
retmax=15&term=Meesmann+corneal+dystrophy]->[{API_CALL}]
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fcgi?db=omim
&retmax=15&retmode=json&id={ids}]->[{API_CALL}]
Answer: KRT12, KRT3
# Demonstration 4
Question: Align the DNA sequence to the human genome:ATTCT[...]TCTCT
[https://blast.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Put&PROGRAM=blas
tn&MEGABLAST=on&DATABASE=nt&FORMAT_TYPE=XML&QUERY=ATTCT[...]TCTCT&HI
TLIST_SIZE=5]->[{API_CALL}]
[https://blast.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Get&FORMAT_TYPE=
Text&RID={RID}]->[{API_CALL}]
Answer: chr15:91950805-91950932
```

GeneTuring Question

```
Which chromosome does SNP rs1430464868 locate on human
genome?
```

GeneGPT Output

```
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fc
gi?db=snp&retmax=3&retmode=json&id=1430464868]->[{"header”
:{"type":"esummary","version":"0.3"},"result":{"uids":["14
30464868"],"1430464868":{"uid":"1430464868","snp_id":14304
64868,"allele_origin":"","global_mafs":[{"study":"GnomAD",
"freq":"C=0.000007/1"},{"study":"ALFA","freq":"C=0./0"}],"
global_population":"","global_samplesize":"","suspected":"
","clinical_significance":"","genes":[],"acc":"NC_000013.1
1","chr":"13","handle": [*TRUNCATED*] merged_sort":"0"}}}]
Answer: chr13                                            ✅
```

GeneHop Question

```
What is the function of the gene associated with SNP
rs1241371358? Let's decompose the question to sub-
questions and solve them step by step.
```

GeneGPT Output

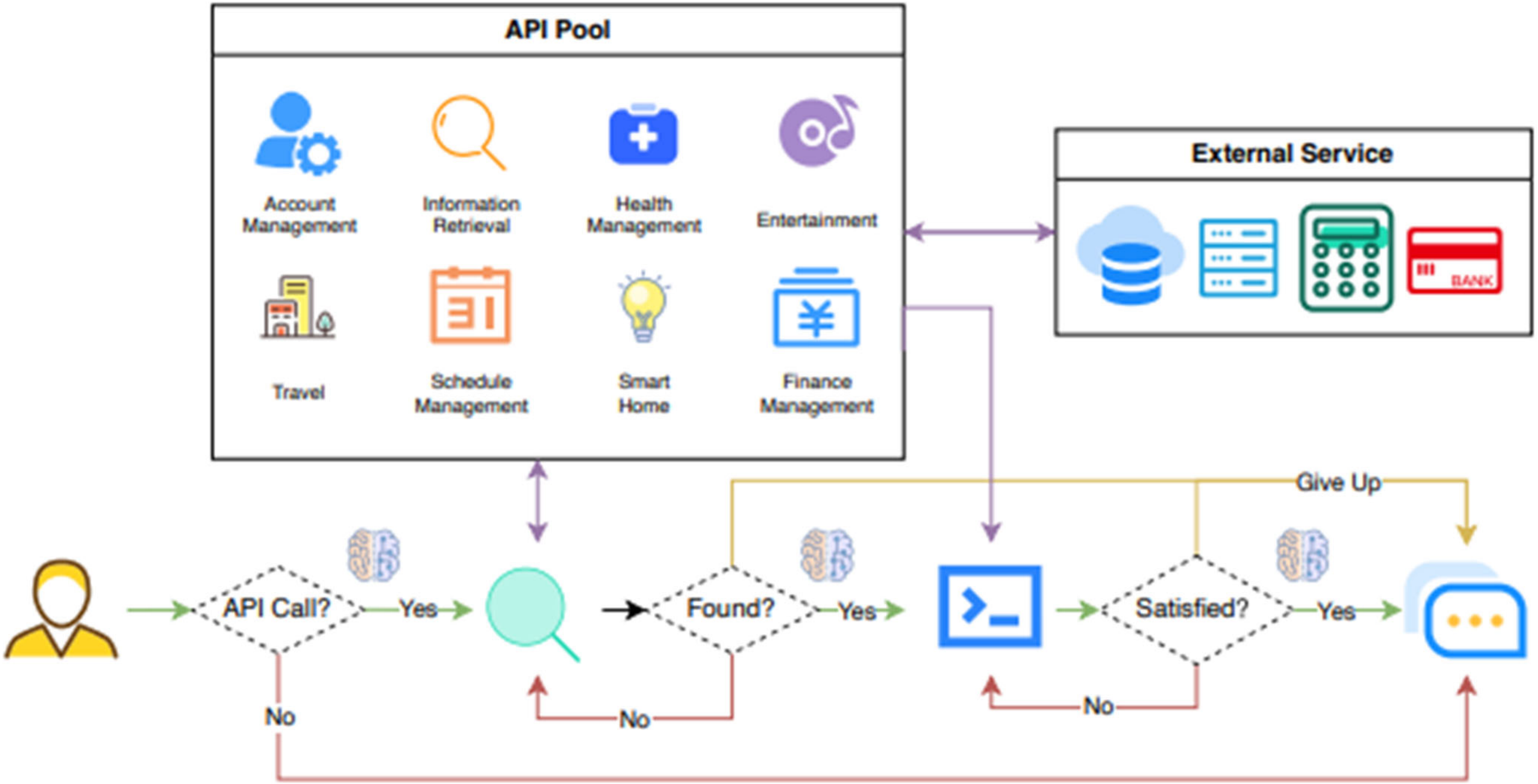
```
Sub-question 1: Which gene is SNP rs1241371358 associated
with?
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fc
gi?db=snp&retmax=10&retmode=json&id=1241371358]->[{"heade
r":{"type":"esummary","version":"0.3"}},"result":{"uids":[
"1241371358"] [*TRUNCATED*] "merged_sort":"0"}}]
Answer: LRRC23

Sub-question 2: What is the function of LRRC23?
[https://eutils.ncbi.nlm.nih.gov/entrez/eutils/esummary.fc
gi?db=gene&retmax=10&retmode=json&id=10233]->[{"header":
{"type":"esummary” [*TRUNCATED*] "merged_sort":"0"}}]
Answer: Predicted to be active in cytosol.             ✅
```

GeneGPT: Augmenting Large Language Models with Domain Tools for Improved Access to Biomedical Information. Arxiv 2023.05. National Institutes of Health

目录

4.　评价指标

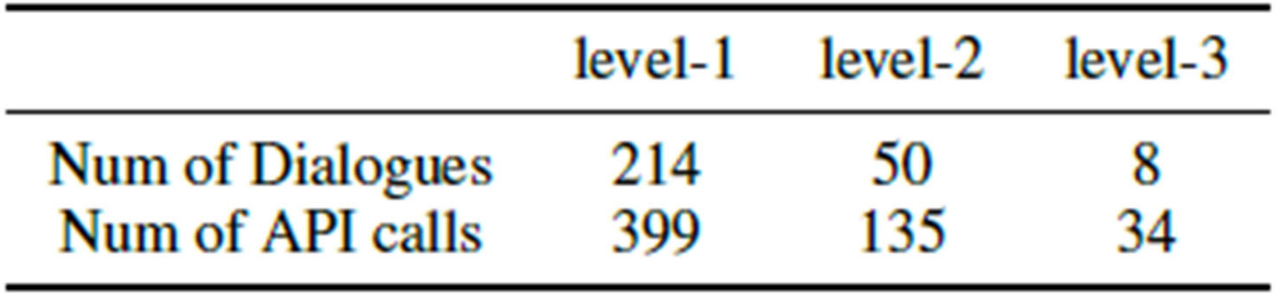
API-Bank



API-Bank: A Benchmark for Tool-Augmented LLMs. Arxiv 2023.4. 阿里达摩院

API-Bank

- 评价指标
  - level-1：评估LLM 调用 API 的能力（Accuracy）；给定API的描述，模型需要确定是否调用该API。
  - level-2：进一步评估LLM检索API的能力（Rouge）；模型需要检索可能解决用户需求的API。
  - level-3：考察LLM规划API的能力（number of turns）。



	level-1	level-2	level-3
Num of Dialogues	214	50	8
Num of API calls	399	135	34

API-Bank: A Benchmark for Tool-Augmented LLMs. Arxiv 2023.4. 阿里达摩院

GPT4Tools

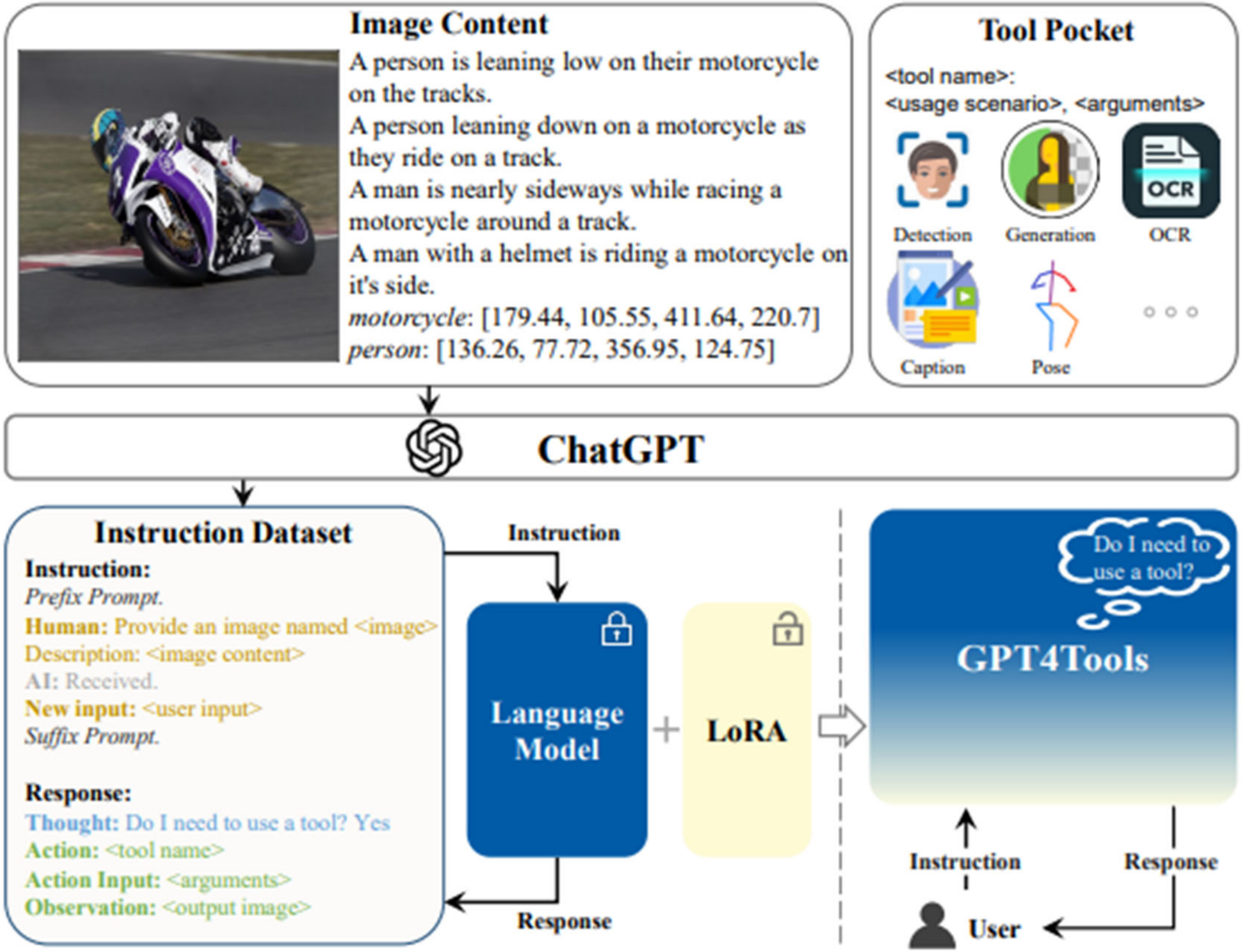


Image Content

A person is leaning low on their motorcycle on the tracks.
A person leaning down on a motorcycle as they ride on a track.
A man is nearly sideways while racing a motorcycle around a track.
A man with a helmet is riding a motorcycle on it's side.
motorcycle: [179.44, 105.55, 411.64, 220.7]
person: [136.26, 77.72, 356.95, 124.75]

Tool Pocket

<tool name>:
<usage scenario>, <arguments>

Instruction Dataset

Instruction:
Prefix Prompt.
Human: Provide an image named <image>
Description: <image content>
AI: Received.
New input: <user input>
Suffix Prompt.

Response:
Thought: Do I need to use a tool? Yes
Action: <tool name>
Action Input: <arguments>
Observation: <output image>

GPT4Tools: Teaching Large Language Model to Use Tools via Self-instruction. Arxiv 2023.5. 清华

GPT4Tools

- 评价指标
  - Successful Rate of Thought：

$$SR_t = \frac{1}{N}\sum_{i=1}^{N}\Pi(\tau_i)$$

  - Successful Rate of Action：

$$SR_{act} = \frac{1}{N}\sum_{i=1}^{N}\Pi(\alpha_i)$$

  - Successful Rate of Arguments：

$$SR_{args} = \frac{1}{N}\sum_{i=1}^{N}\eta_i, \quad where \quad \eta_i = \frac{1}{K}\sum_{j}^{K}\eta_{i,j}$$

  - Successful Rate：

$$SR = \frac{1}{N}\sum_{i=1}^{N}\Pi(\tau_i)\cdot\Pi(\alpha_i)\cdot\Pi(\eta_i > 0.5)$$

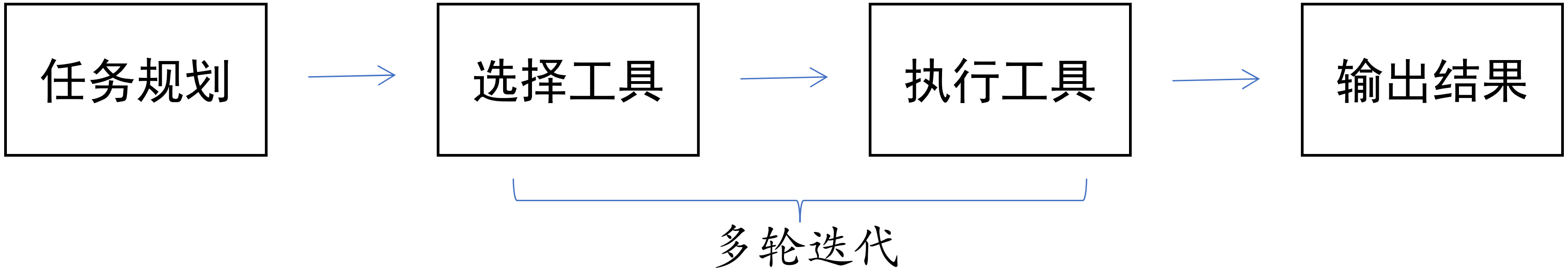
GPT4Tools: Teaching Large Language Model to Use Tools via Self-instruction. Arxiv 2023.5. 清华

目录

5. 总结

总结

- 基本流程：



任务规划 → 选择工具 → 执行工具 → 输出结果

多轮迭代

- 实现方式：In-context learning、 Fine-tuning

- 现有缺陷：
  - 工具使用能力的评价指标
  - 高质量的数据集
  - Prompt设计

参考文献

[1] Wu C, Yin S, Qi W, et al. Visual chatgpt: Talking, drawing and editing with visual foundation models[J]. arXiv preprint arXiv:2303.04671, 2023.
[2] Shen Y, Song K, Tan X, et al. Hugginggpt: Solving ai tasks with chatgpt and its friends in huggingface[J]. arXiv preprint arXiv:2303.17580, 2023.
[3] Schick T, Dwivedi-Yu J, Dessì R, et al. Toolformer: Language models can teach themselves to use tools[J]. arXiv preprint arXiv:2302.04761, 2023.
[4] Patil S G, Zhang T, Wang X, et al. Gorilla: Large Language Model Connected with Massive APIs[J]. arXiv preprint arXiv:2305.15334, 2023.
[5] Zhang B, Zhou K, Wei X, et al. Evaluating and Improving Tool-Augmented Computation-Intensive Math Reasoning[J]. arXiv preprint arXiv:2306.02408, 2023.
[6] Jin Q, Yang Y, Chen Q, et al. GeneGPT: Augmenting Large Language Models with Domain Tools for Improved Access to Biomedical Information[J]. ArXiv, 2023.
[7] Li M, Song F, Yu B, et al. Api-bank: A benchmark for tool-augmented llms[J]. arXiv preprint arXiv:2304.08244, 2023.
[8] Yang R, Song L, Li Y, et al. GPT4Tools: Teaching Large Language Model to Use Tools via Self-instruction[J]. arXiv preprint arXiv:2305.18752, 2023.

谢谢大家

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS