



# SearchLVLMs: Augmenting LVLMs by Searching Up-to-Date Knowledge

Chuanhao Li  
Beijing Institute of Technology  
[lichuanhao@bit.edu.cn](mailto:lichuanhao@bit.edu.cn)

2024-11-13

# Challenging Problems

---

## ➤ Novel Concepts/Events

- **Cognitive Boundary Theory**

- Humans develop certain cognitive “boundaries” or “limits” in the processes of thinking, learning, and decision-making.

## ➤ Motivation

- Large Vision-Language Model
  - Expensive training → Infrequent updates
- Real World
  - New information/knowledge
- Result
  - LVLMs failure



**Question:** Where is the Meditation Spot located at this Temple?

**Multi-Choices:**

- (A) Temple Garden
- (B) Temple Pavilion
- (C) Temple Entrance ✓
- (D) Temple Sanctuary

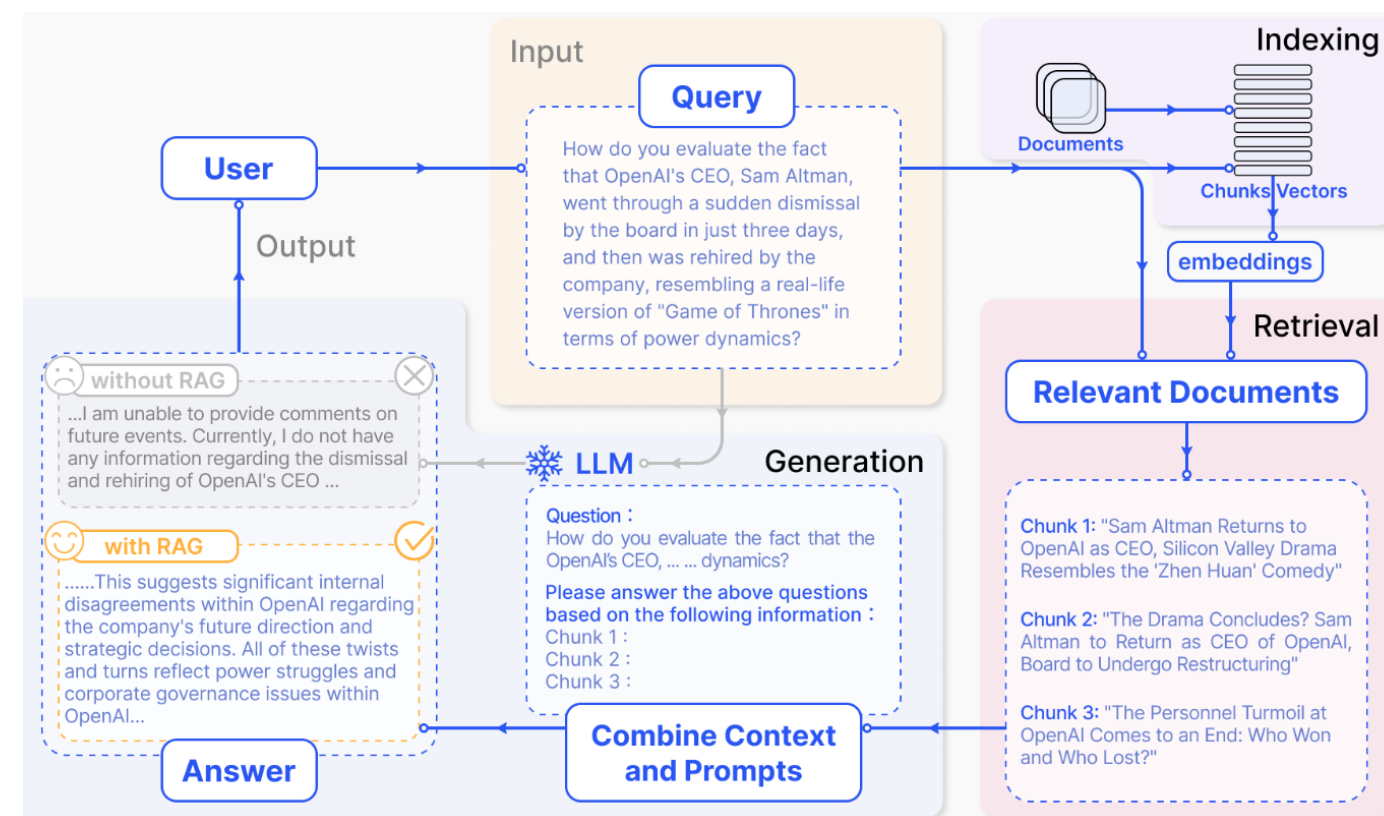


# Challenging Problems

## ➤ Novel Concepts/Events

### • Existing Work

- Retrieval-Augmented Generation.
- LVLMs with Internet augmentation.
- Methods based on External Tools.



Gao et al. "RAG for LLMs: A Survey." arXiv. 2024.

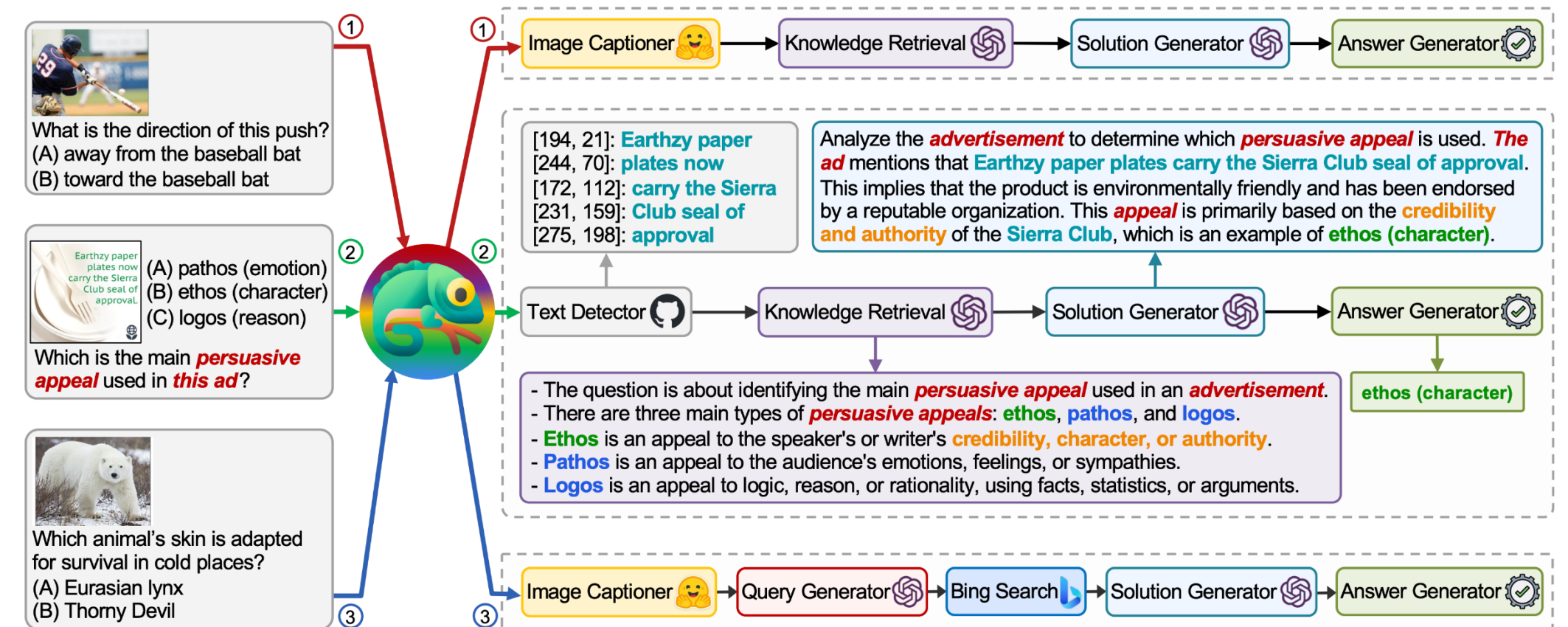


联网搜索后回答我，她叫什么名字。

她是黑神话悟空中的角色

I'm unable to identify the person in the image. If you have a description or more context about the individual, I can help with similar styling or information.

谢谢你的补充！黑神话悟空中的角色们都有独特的风格，这位角色的黑暗、精致的装扮确实符合游戏的美学。黑神话悟空的角色通常会有丰富的细节和浓厚的神话色彩，如果你想了解更多关于这个角色的具体信息或游戏中的设定，可以随时告诉我。



Lu et al. "Chameleon: Plug-and-Play Compositional Reasoning with LLMs." NeurIPS. 2023.

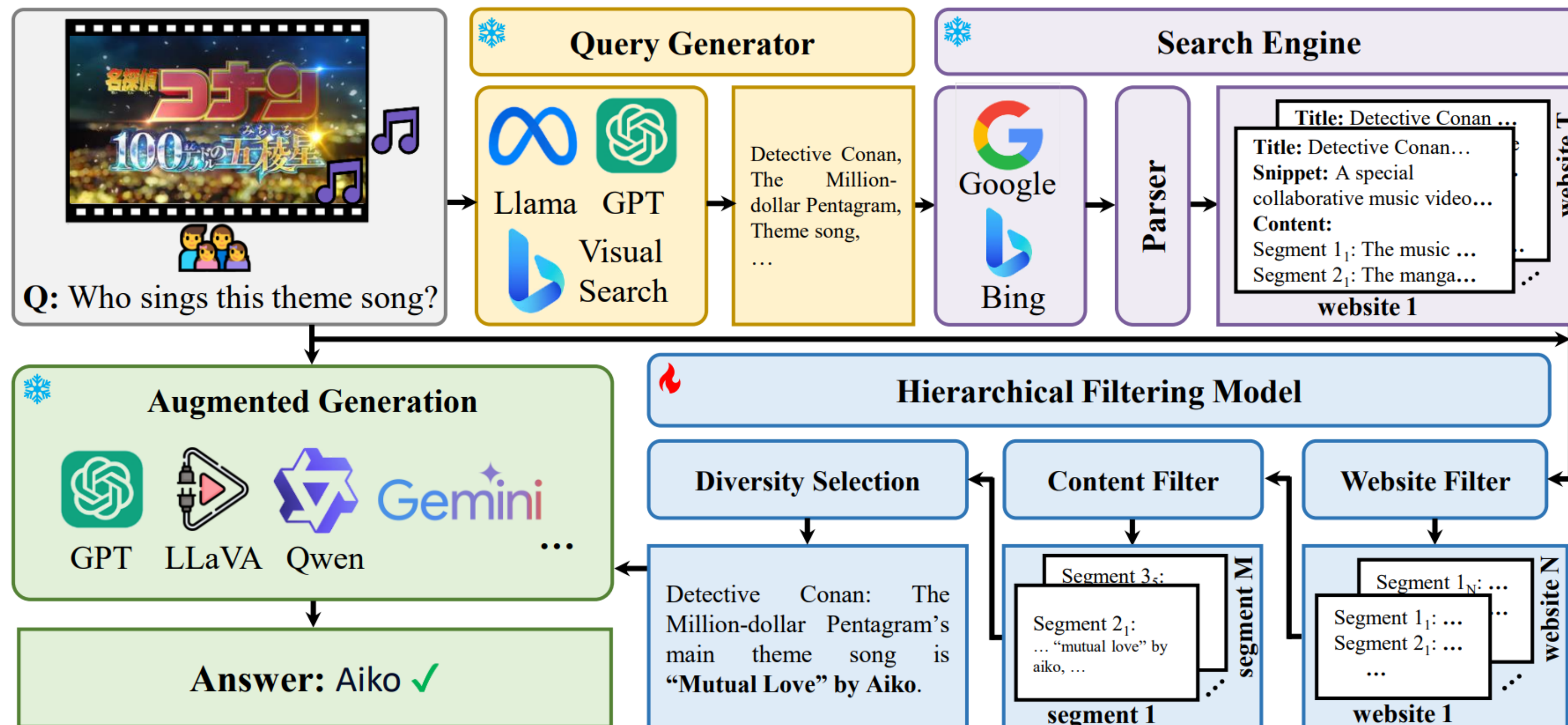
### • Our Solution

- Internet-Augmented Generation.

# SearchLVLMs: Augmenting LVLMs by Searching Up-to-Date Knowledge

## ➤ Solution: SearchLVLMs Framework

- Hierarchical Filtering Model
  - Most LVLMs are poor at handling such long contexts
  - Handling long contexts is computationally intensive and time-consuming

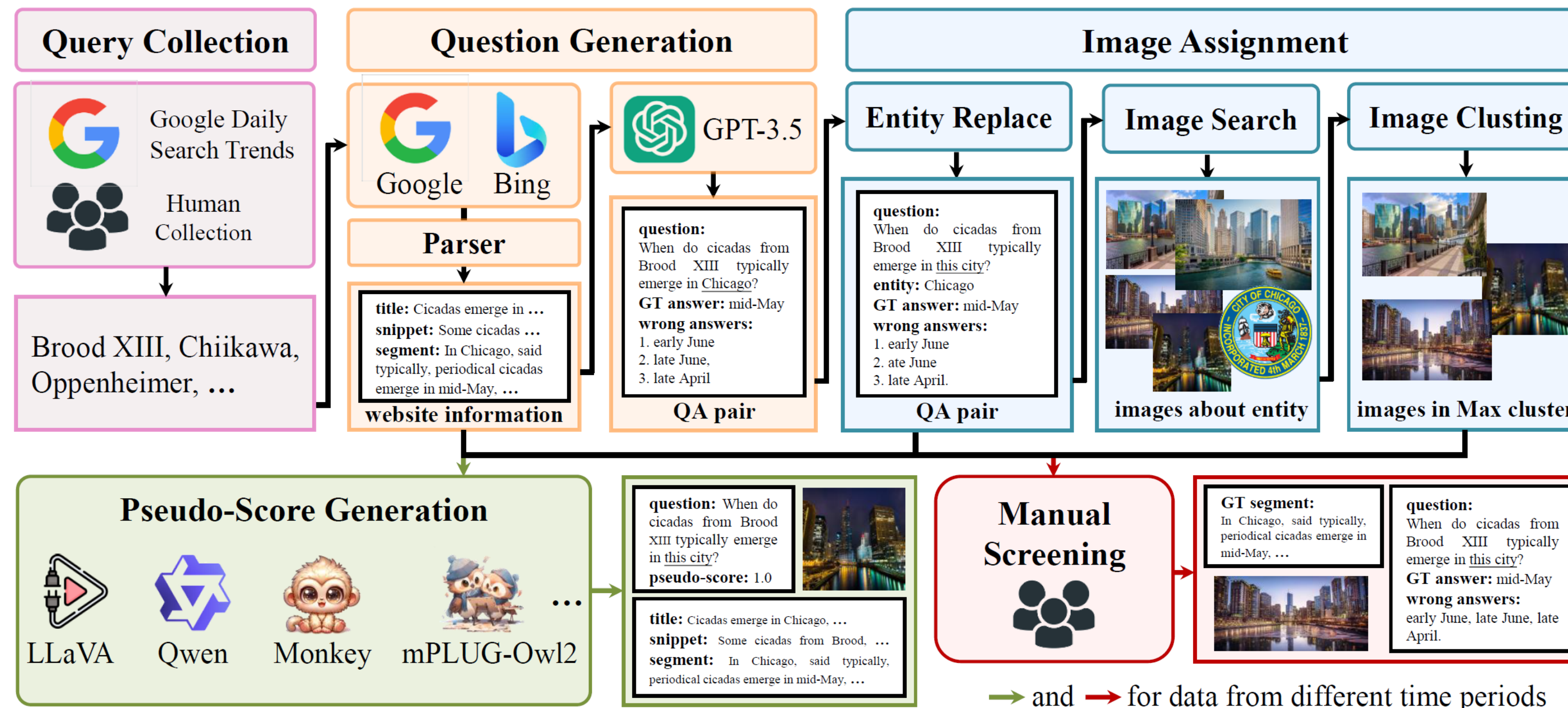




# SearchLVLMs: Augmenting LVLMs by Searching Up-to-Date Knowledge

## ➤ Sample Generation Pipeline





- Pseudo-Score Generation: for Training
- Manual Screening: for Evaluation





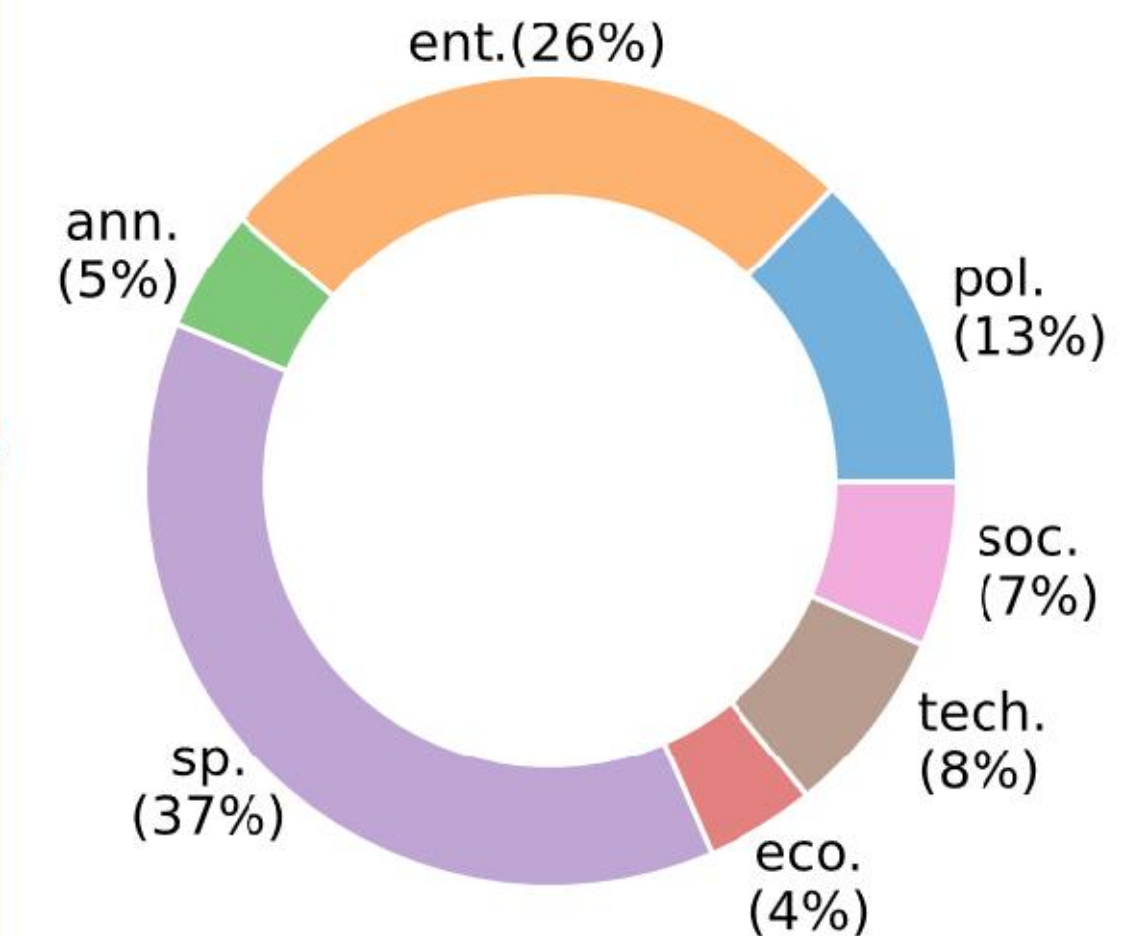
# SearchLVLMs: Augmenting LVLMs by Searching Up-to-Date Knowledge

## ➤ UDK-VQA Dataset

	<p><b>title:</b> Where did Auburn finish in the ... <b>snippet:</b> Auburn basketball ended its SEC, ... <b>segment:</b> The Tigers will cap the 2023-24 season by being ranked No. 17, ... <b>question:</b> Where did this College basketball team finish in the final USA TODAY Sports Coaches Poll for the 2023-24 season? <b>pseudo-score:</b> <math>R_c=1.0, R_w=1.0</math></p>		<p><b>GT segment:</b> Tems announced that her debut album Born in the Wild was done, ... <b>question:</b> What was the name of Tems' debut album announced at this music festival? <b>GT answer:</b> Born in the Wild <b>wrong answers:</b> For Broken Ears, Essence, If Orange Was a Place.</p>
	<p><b>title:</b> Everything about Banff National Park <b>snippet:</b> Simplicity marks the origin ... <b>segment:</b> Simplicity marks the origin of Banff Canada's first national park, ... <b>question:</b> What caused the stranded passengers in this National Park? <b>pseudo-score:</b> <math>R_c=0.4, R_w=0.6</math></p>		<p><b>GT segment:</b> ..., acts as a barrier for over-land runoff of nutrients and eroded soil, ... <b>question:</b> What is the purpose of planting new woodland in this River's SAC catchment? <b>GT answer:</b> a barrier for over-land runoff <b>wrong answers:</b> provide shade, reduce river temperature, adapt to climate change.</p>

Training Samples.

Test Samples.



Category statistics.

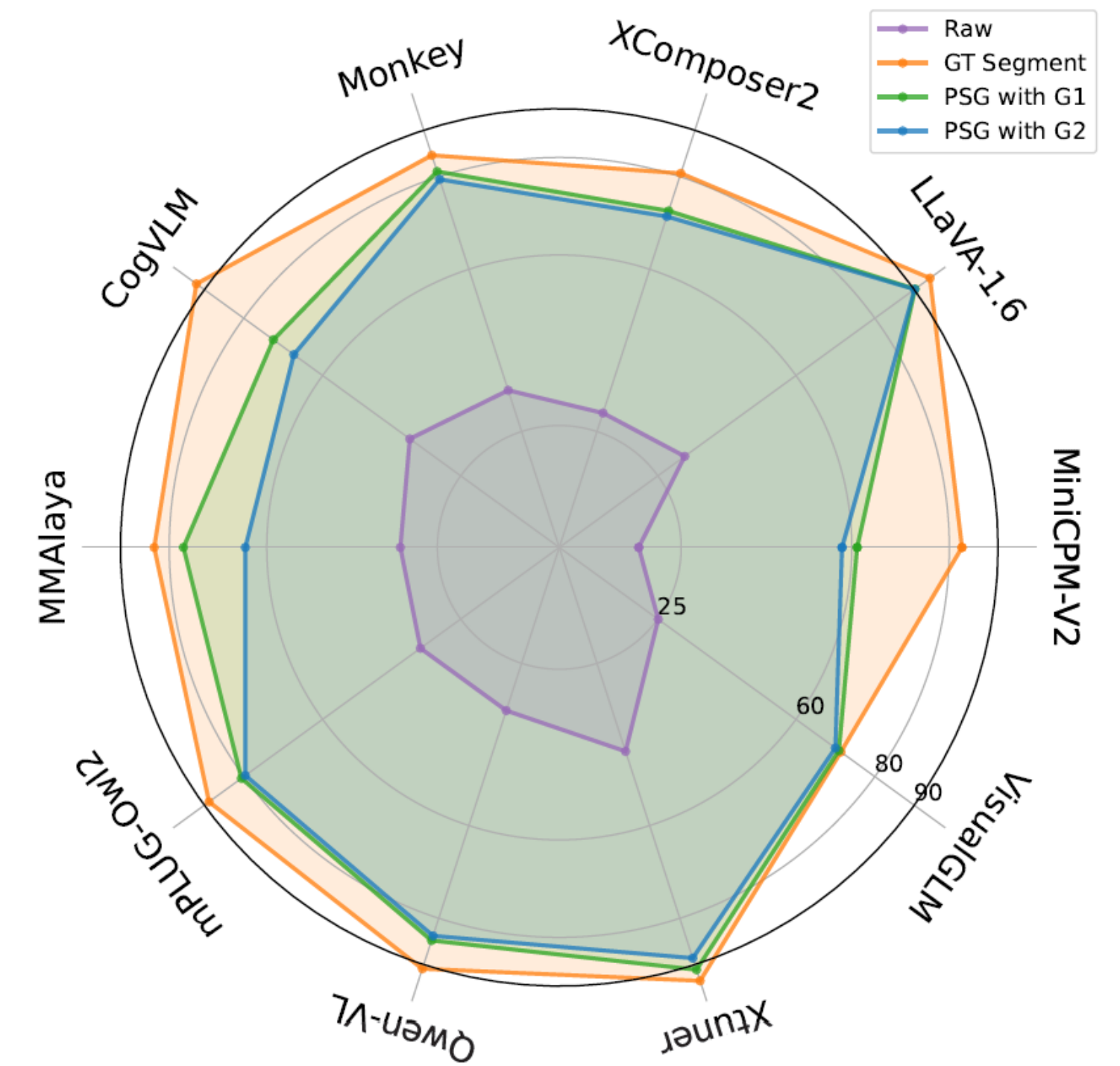


# SearchLVLMs: Augmenting LVLMs by Searching Up-to-Date Knowledge

## ➤ Test for Up-to-date Knowledge

### Our UDK-VQA Dataset

Model Variant	pol.	ent.	ann.	sp.	eco.	tech.	soc.	overall	
Gemini 1.5 Pro	Raw	6.2 (5.7)	15.8 (16.3)	10.2 (11.9)	7.4 (8.1)	2.3 (2.5)	8.0 (6.2)	3.0 (3.7)	9.1 (9.5)
	LC	61.7 (65.7)	71.5 (77.3)	73.5 (76.2)	77.2 (79.2)	72.7 (72.5)	81.3 (83.1)	62.1 (66.7)	76.4 (76.1)
	<b>Ours</b>	<b>82.8 (82.9)</b>	<b>79.6 (79.0)</b>	<b>91.8 (92.9)</b>	<b>81.5 (80.1)</b>	<b>97.7 (97.5)</b>	<b>84.0 (83.1)</b>	<b>90.9 (88.9)</b>	<b>83.3 (82.3)</b>
GPT 4V	Raw	21.1 (23.8)	31.5 (30.9)	16.3 (19.0)	16.7 (17.5)	15.9 (17.5)	41.3 (38.5)	21.2 (22.2)	24.2 (23.8)
	IAG	62.5 (68.0)	61.9 (63.6)	63.3 (66.7)	62.4 (63.3)	70.5 (67.5)	80.0 (78.5)	69.7 (70.4)	64.5 (65.9)
	<b>Ours</b>	<b>76.6 (83.8)</b>	<b>85.8 (85.8)</b>	<b>89.8 (90.5)</b>	<b>86.2 (86.4)</b>	<b>97.7 (97.5)</b>	<b>92.0 (90.8)</b>	<b>87.9 (92.6)</b>	<b>87.2 (87.4)</b>
GPT 4o	Raw	36.7 (40.0)	34.2 (36.1)	42.9 (47.6)	28.6 (30.4)	40.9 (42.5)	65.3 (67.7)	36.4 (38.9)	37.2 (37.8)
	IAG	61.7 (63.1)	57.3 (58.4)	69.4 (64.3)	48.4 (47.9)	70.5 (75.0)	81.3 (81.5)	62.1 (61.1)	57.8 (57.9)
	<b>Ours</b>	<b>86.7 (92.4)</b>	<b>89.6 (91.4)</b>	<b>98.0 (100)</b>	<b>83.9 (88.0)</b>	<b>97.7 (100)</b>	<b>90.7 (96.9)</b>	<b>89.4 (94.4)</b>	<b>91.8 (91.6)</b>
LLaVA 1.6	Raw	43.8 (45.7)	32.3 (31.8)	22.4 (23.8)	24.9 (23.2)	25.0 (25.0)	53.3 (55.4)	33.3 (31.5)	31.8 (31.2)
	Gen.	39.1 (40.0)	31.5 (28.8)	18.4 (19.0)	25.7 (25.3)	36.4 (37.5)	44.0 (44.6)	39.4 (37.0)	31.3 (30.4)
	Cham.	58.6 (58.1)	57.3 (57.5)	53.1 (57.1)	65.3 (67.2)	52.3 (52.5)	72.0 (67.7)	74.2 (72.2)	62.3 (62.7)
	C→F*	55.5 (56.2)	56.5 (57.9)	34.7 (35.7)	54.0 (54.5)	54.5 (52.5)	62.7 (64.6)	56.1 (53.7)	54.7 (55.3)
	<b>Ours</b>	<b>86.7 (87.6)</b>	<b>91.9 (91.8)</b>	<b>93.9 (97.6)</b>	<b>88.1 (88.0)</b>	<b>90.9 (90.0)</b>	<b>93.3 (93.8)</b>	<b>95.5 (100)</b>	<b>90.2 (90.7)</b>
Intern VL 1.5	Raw	43.8 (43.8)	53.1 (52.4)	49.0 (47.6)	29.9 (30.7)	34.1 (35.0)	73.3 (76.9)	37.9 (35.2)	42.6 (42.8)
	Gen.	29.7 (30.5)	28.1 (26.2)	28.6 (26.2)	22.8 (23.5)	31.8 (32.5)	42.7 (46.2)	28.8 (27.8)	27.6 (27.6)
	Cham.	59.4 (58.1)	61.9 (61.8)	55.1 (57.1)	55.6 (55.4)	52.3 (52.5)	65.3 (67.7)	71.2 (70.4)	59.3 (59.2)
	C→F*	59.4 (58.1)	65.0 (64.8)	44.9 (42.9)	54.2 (55.7)	47.7 (50.0)	65.3 (66.2)	53.0 (50.0)	57.7 (58.0)
	<b>Ours</b>	<b>90.6 (89.5)</b>	<b>95.4 (95.3)</b>	<b>98.0 (97.6)</b>	<b>88.9 (88.0)</b>	<b>100 (100)</b>	<b>96.0 (95.4)</b>	<b>98.5 (98.1)</b>	<b>92.9 (92.3)</b>



# Q & A

---

Thanks for your listening!



GitHub