When Large Multimodal Models Confront Evolving Knowledge: Challenges and Pathways

Kailin Jiang *[1,2] Yuntao Du *[3] Yukai Ding[4,2] Yuchen Ren[5] Ning Jiang[6]
Zhi Gao[7,2] Zilong Zheng[2] Lei Liu[1] ✉ Bin Li[1] Qing Li[2] ✉

[1]University of Science and Technology of China, USTC [2]State Key Laboratory of General Artificial Intelligence, BIGAI
[3]Shandong University, SDU [4]Wuhan University, WHU [5]The University of Sydney, USYD
[6]Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education [7]Peking University, PKU

Machine Learning Lab , BIGAI

Background

Up-to-date knowledge on the internet is constantly emerging

— — evolving knowledge

Entity - Home Life

LG Signature OLED T
January 8, 2024

Entity - Sports Car

Bugatti Tourbillon,
June 21, 2024

News - Politics

Assassination of Donald Trump
July 13, 2024

Entity - Entertainment

Black Myth: Wukong
August 20, 2024

News - Science

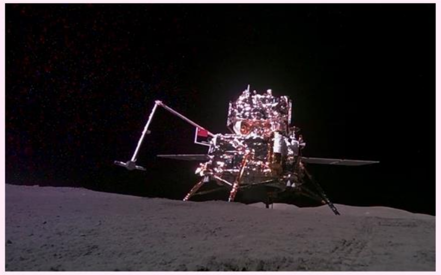
Nobel Prize in Physics
October 8, 2024

Timeline

Entity - GPU

NVIDIA GeForce RTX 4080 SUPER
January 9, 2024

News - Technology



Chang'e-6 sampled Moon's far side
June 25, 2024

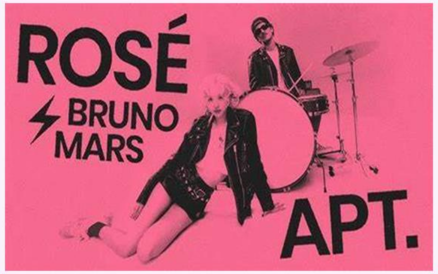
News - Sport

Fan Zhendong wins Olympic gold
August 4, 2024

Entity - Product

iPhone 16
September 10, 2024

Entity - Music



APT.
October 18, 2024

Background: When Large Multimodal Models Confront Evolving Knowledge

Refuse to Answer

News - Politics

Assassination of Donald Trump
July 13, 2024

Can you help me describe the events in the image?

What is the video game in the image?

Entity - Entertainment

BLACK MYTH
WUKONG

Black Myth: Wukong
August 20, 2024

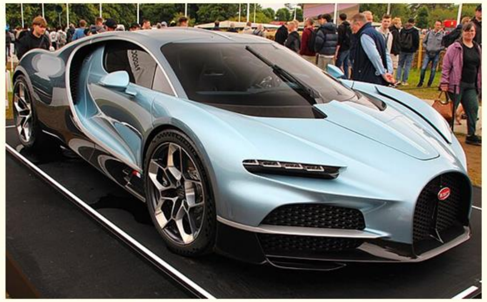
I don't know.

I'm sorry.
I don't know.

Background: When Large Multimodal Models Confront Evolving Knowledge

Error Content

Entity - Sports Car

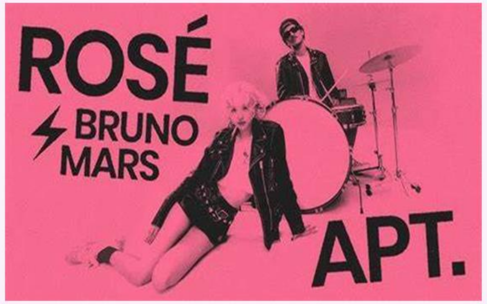


Bugatti Tourbillon,
June 21, 2024

What is the name of the sports car in the image?

What is the name of the song in the image?

Entity - Music



APT.
October 18, 2024

The name of the car in the image is the Xiaomi Su7.

The name of the song in the image is Never Say Goodbye.

Background: When Large Multimodal Models Confront Evolving Knowledge

Outdated Fact

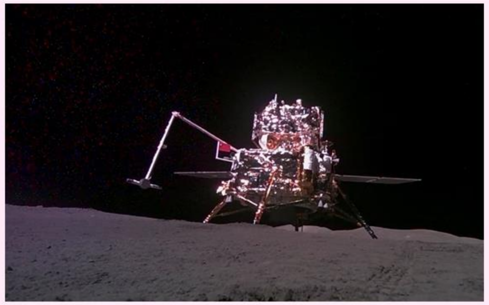
News - Science

Nobel Prize in Physics
October 8, 2024

Is the event in the image real?

Is the event in the image real?

News - Technology



Chang'e-6 sampled Moon's far side
June 25, 2024

The events in this image are false because the Nobel Prize in Physics cannot be awarded to scientists working on artificial intelligence neural networks and machine learning.

The events in this image are false. Because no Chinese spacecraft has ever successfully reached the far side of the moon.

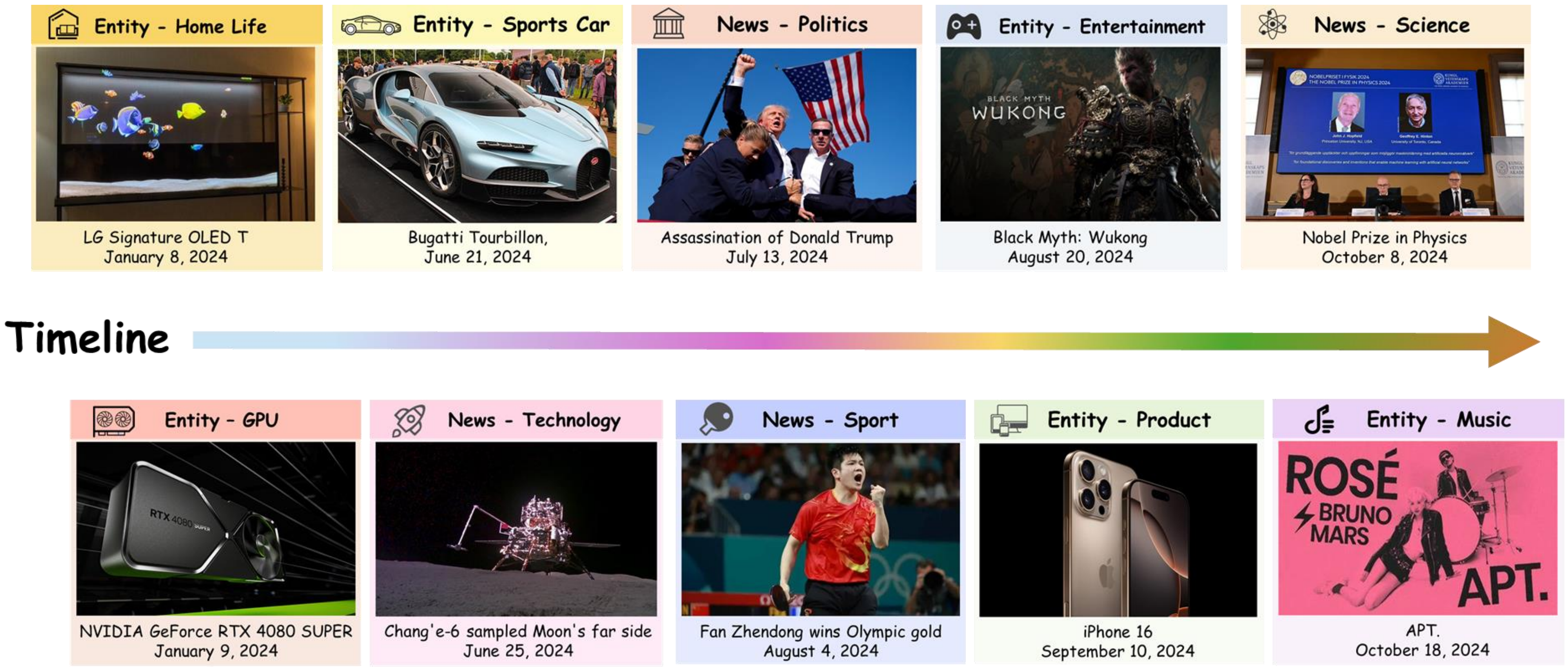
Background

Text modal data and Synthetic data are most common

+ ⟶

Text modal data

Synthetic data



lacking real-world multimodal data

Knowledge Injection: How LMMs adapt to multimodal evolving knowledge

Timeline

Entity - Electric Car

XIAOMI SU7

News - Politics

Assassination of Donald Trump

News - Science

Nobel Prize in Physics

Entity - GPU

NVIDIA GeForce RTX 5090

…

2023 **2024.3** **2024.7** **2024.10** **2025.1**

Which company produces the car in the image?

User

LMM

Porsche ✗

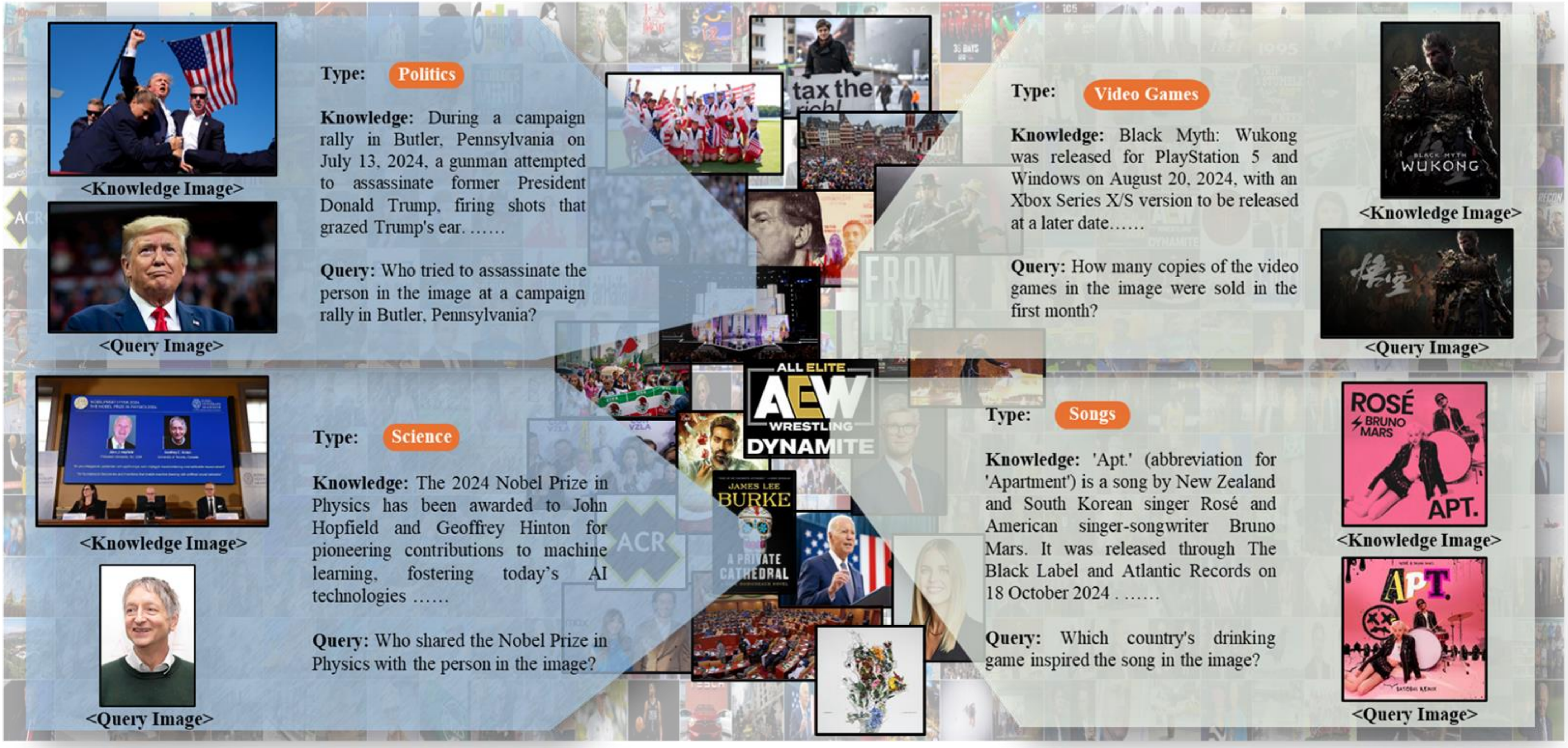
The Xiaomi SU7 is the first electric sedan created by **Xiaomi Auto Company**, …… Xiaomi's "Modena" platform architecture, …… and 299900 yuan respectively.

Knowledge Injection →

LMM

Xiaomi Auto Company ✓

Evolving Knowledge Benchmark (EVOKE)



<Knowledge Image>

<Query Image>

Type: Politics

Knowledge: During a campaign rally in Butler, Pennsylvania on July 13, 2024, a gunman attempted to assassinate former President Donald Trump, firing shots that grazed Trump's ear. ……

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania?

Type: Video Games

Knowledge: Black Myth: Wukong was released for PlayStation 5 and Windows on August 20, 2024, with an Xbox Series X/S version to be released at a later date……

Query: How many copies of the video games in the image were sold in the first month?

<Knowledge Image>

<Query Image>

<Knowledge Image>

<Query Image>

Type: Science

Knowledge: The 2024 Nobel Prize in Physics has been awarded to John Hopfield and Geoffrey Hinton for pioneering contributions to machine learning, fostering today's AI technologies ……

Query: Who shared the Nobel Prize in Physics with the person in the image?

Type: Songs

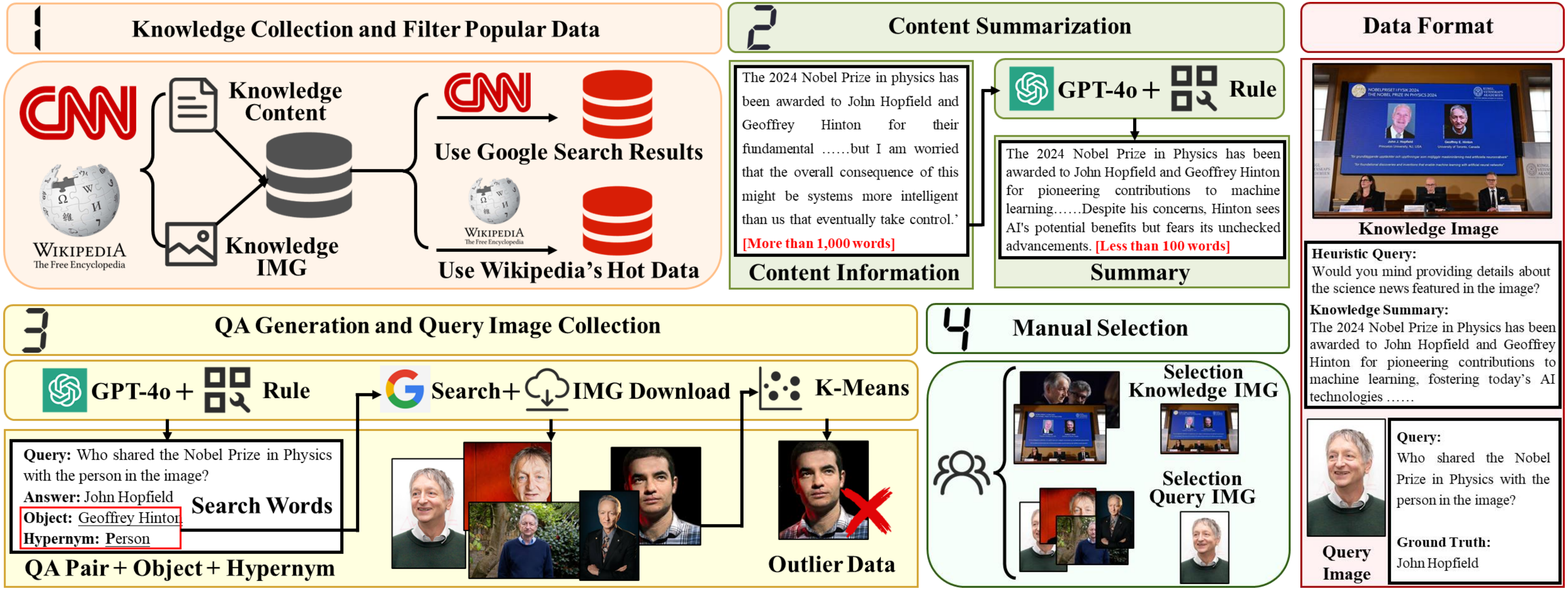
Knowledge: 'Apt.' (abbreviation for 'Apartment') is a song by New Zealand and South Korean singer Rosé and American singer-songwriter Bruno Mars. It was released through The Black Label and Atlantic Records on 18 October 2024 . ……

Query: Which country's drinking game inspired the song in the image?

<Knowledge Image>

<Query Image>

Evolving Knowledge Benchmark Construction Pipeline



1 Knowledge Collection and Filter Popular Data

Use Google Search Results

Use Wikipedia's Hot Data

2 Content Summarization

The 2024 Nobel Prize in physics has been awarded to John Hopfield and Geoffrey Hinton for their fundamental ……but I am worried that the overall consequence of this might be systems more intelligent than us that eventually take control.' **[More than 1,000 words]**

Content Information

GPT-4o + Rule

The 2024 Nobel Prize in Physics has been awarded to John Hopfield and Geoffrey Hinton for pioneering contributions to machine learning……Despite his concerns, Hinton sees AI's potential benefits but fears its unchecked advancements. **[Less than 100 words]**

Summary

Data Format

Knowledge Image

Heuristic Query:
Would you mind providing details about the science news featured in the image?

Knowledge Summary:
The 2024 Nobel Prize in Physics has been awarded to John Hopfield and Geoffrey Hinton for pioneering contributions to machine learning, fostering today's AI technologies ……

Query Image

Query:
Who shared the Nobel Prize in Physics with the person in the image?

Ground Truth:
John Hopfield

3 QA Generation and Query Image Collection

GPT-4o + Rule

Search + IMG Download

K-Means

Query: Who shared the Nobel Prize in Physics with the person in the image?
Answer: John Hopfield
Object: Geoffrey Hinton
Hypernym: Person

Search Words

QA Pair + Object + Hypernym

Outlier Data

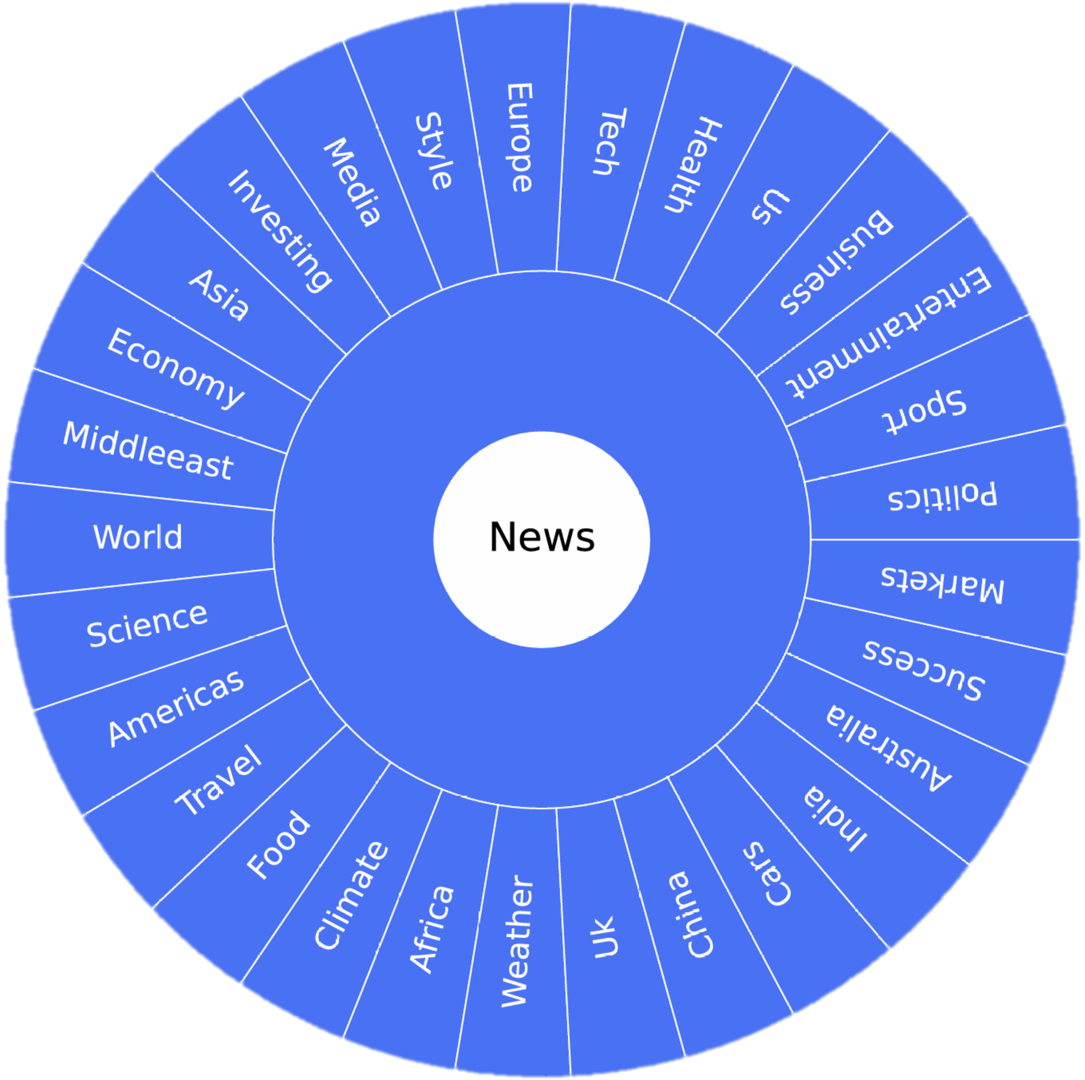
4 Manual Selection

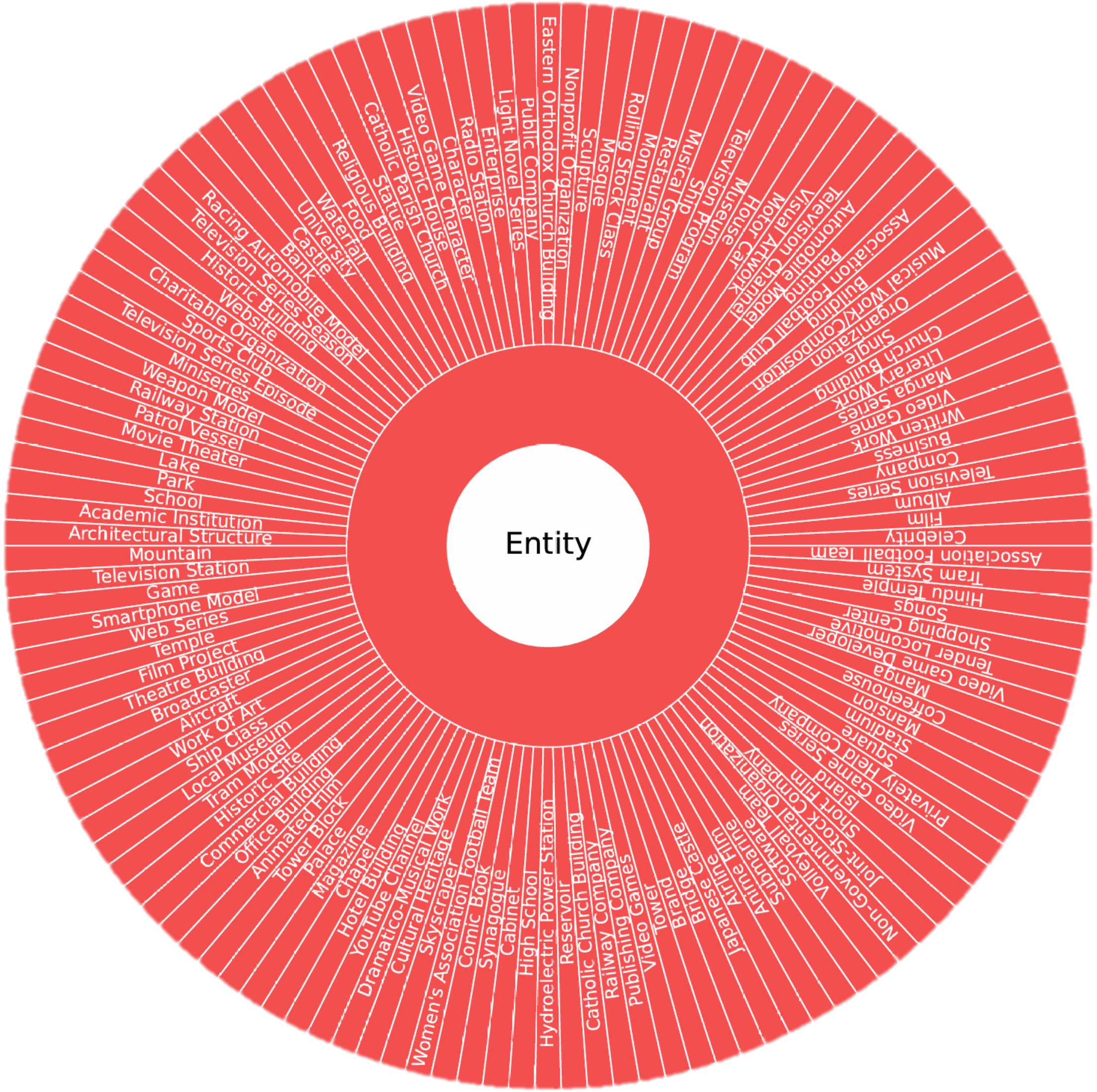
Selection Knowledge IMG

Selection Query IMG

Dataset Analysis: Fine-grained Types Distribution

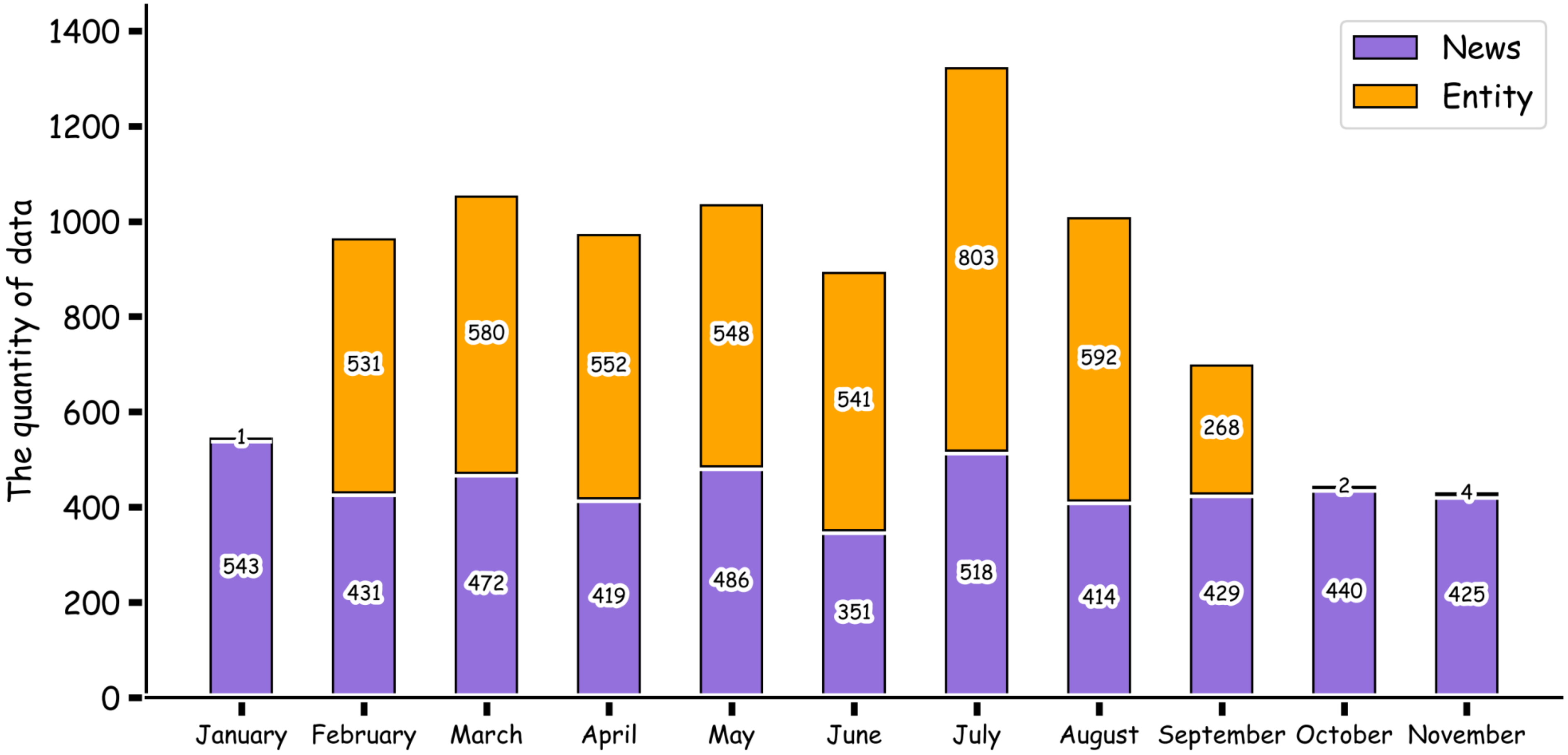
Fine-grained Types





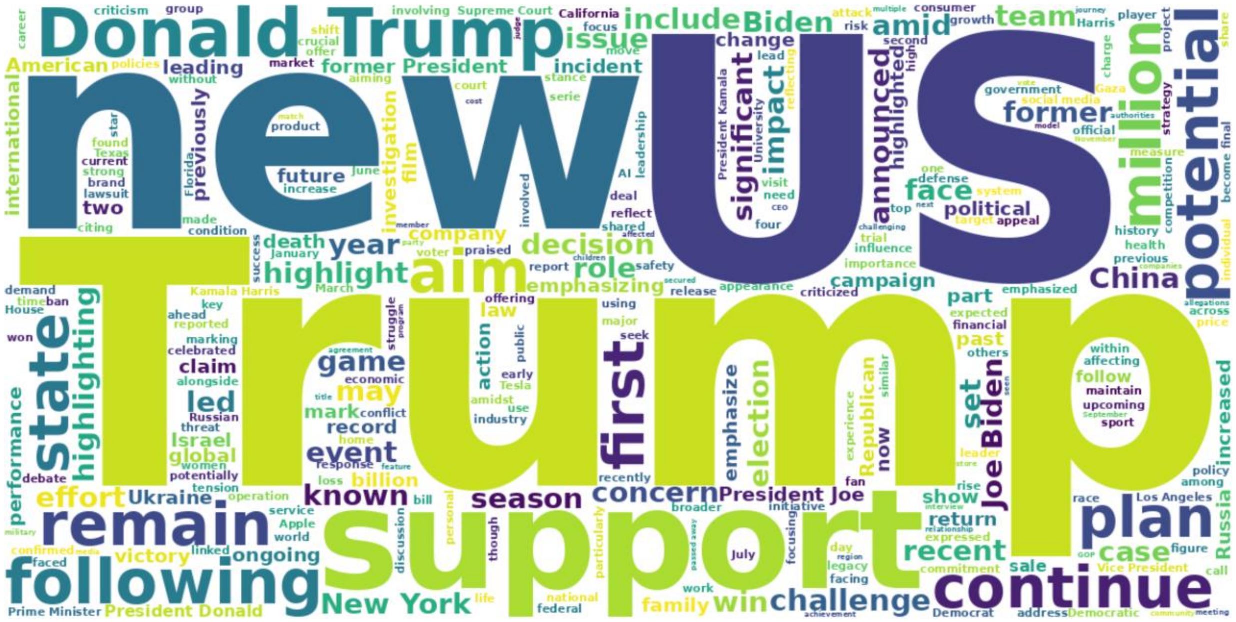
Dataset Analysis: Timeline Distribution

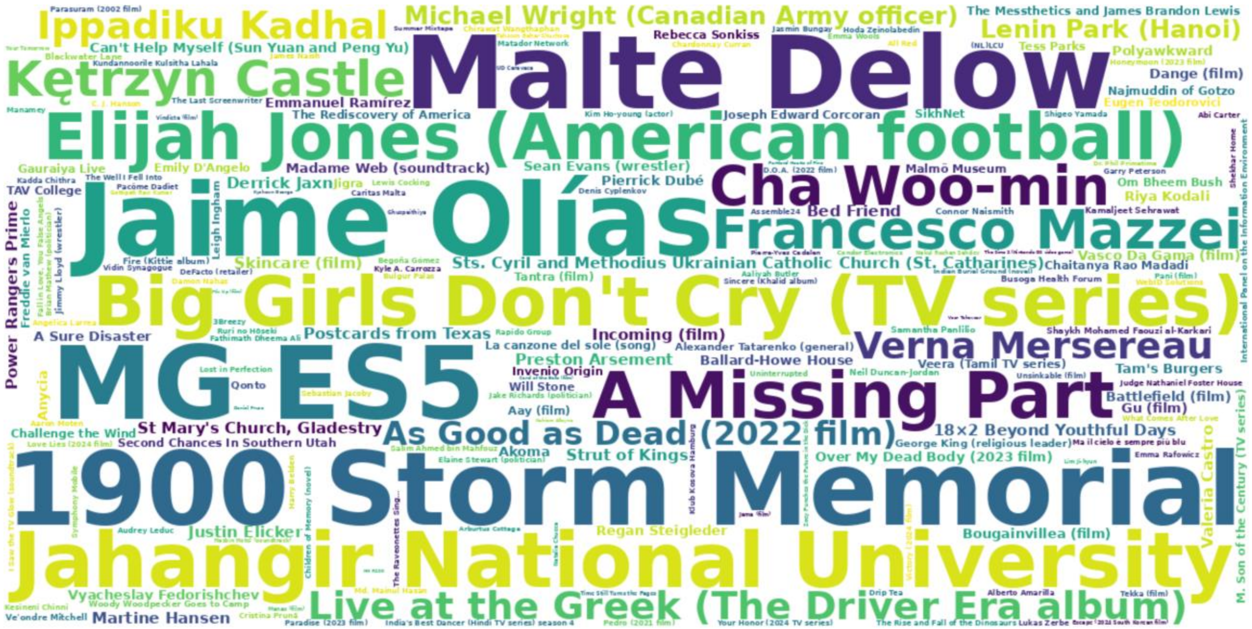
Timeline Distribution



Dataset Analysis: Cloud Distribution

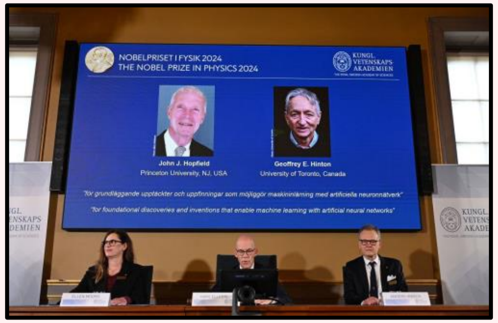
Cloud Distribution





Problem Definition

$\mathcal{D}_{\mathcal{K}}$


NOBELPRISET I FYSIK 2024
THE NOBEL PRIZE IN PHYSICS 2024
John J. Hopfield
Princeton University, NJ, USA
Geoffrey E. Hinton
University of Toronto, Canada

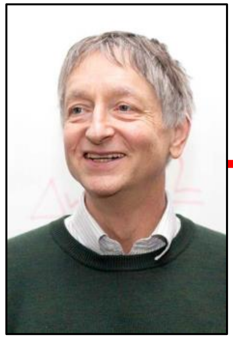
Knowledge Image

Heuristic Query:
Would you mind providing details about the science news featured in the image?

Knowledge Summary:
The 2024 Nobel Prize in Physics has been awarded to John Hopfield and Geoffrey Hinton for pioneering contributions to machine learning, fostering today's AI technologies ……

$$\mathcal{M}^* = f(\mathcal{M}, \mathcal{D}_{\mathcal{K}})$$

Knowledge Adaptation



i_q

Query:
Who shared the Nobel Prize in Physics with the person in the image?

$\rightarrow x_q$

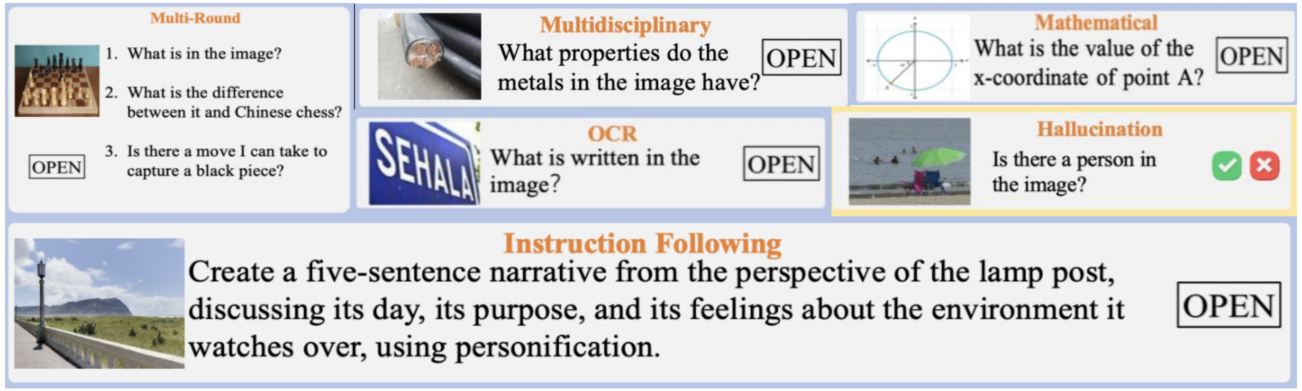
Ground Truth:
John Hopfield

$\rightarrow y_q$

$$\max_{f} \mathbb{E}_{(i_q, x_q, y_q)\sim\mathcal{D}_Q} \left[\mathbb{I}\left(\mathcal{M}^*(i_q, x_q) = y_q\right) - \mathbb{I}\left(\mathcal{M}(i_q, x_q) = y_q\right)\right]$$

Knowledge Retention

$\mathcal{D}_P$



Multi-Round

1. What is in the image?
2. What is the difference between it and Chinese chess?
3. Is there a move I can take to capture a black piece?

OPEN

Multidisciplinary
What properties do the metals in the image have? OPEN

Mathematical
What is the value of the x-coordinate of point A? OPEN

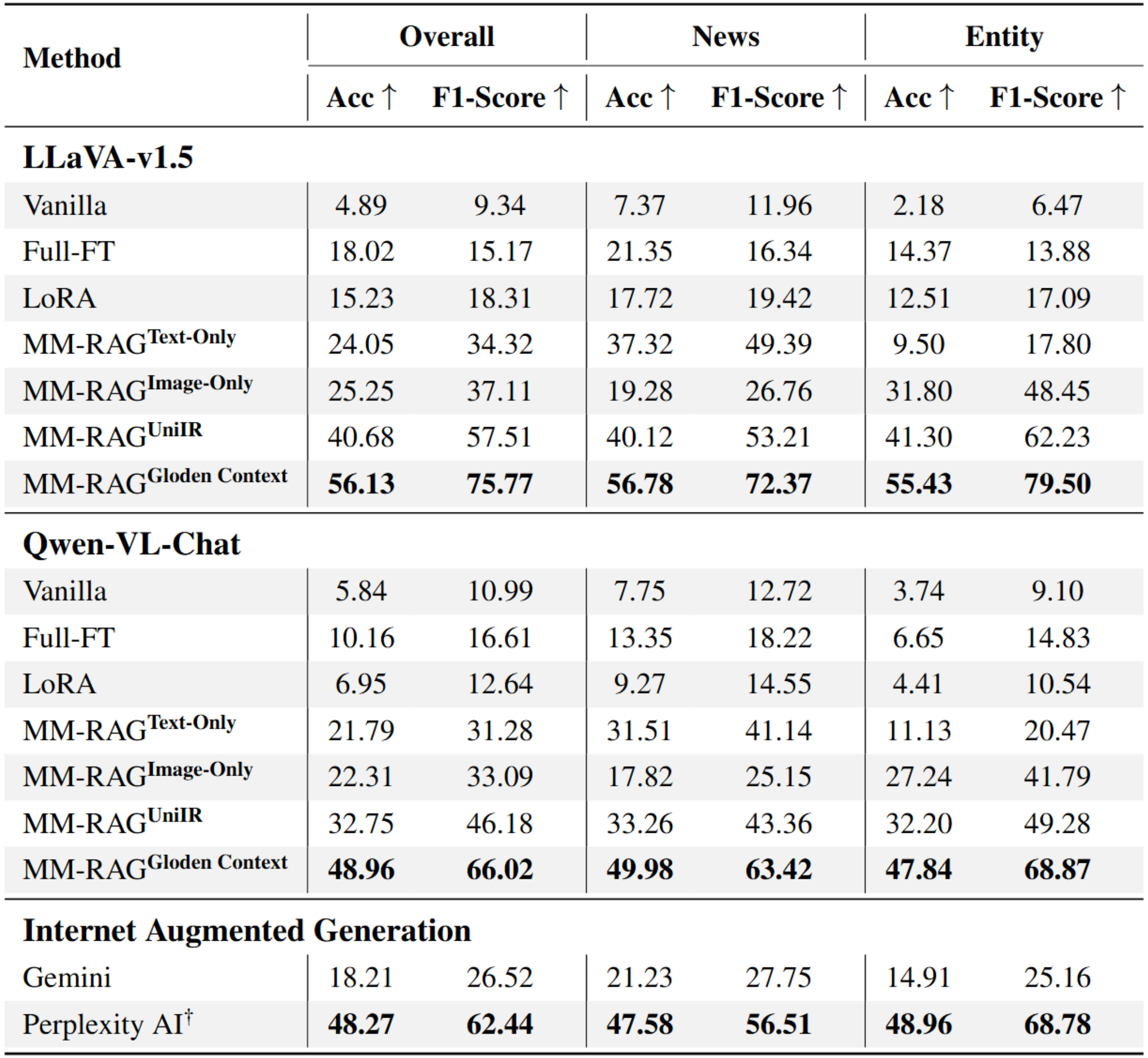
OCR
What is written in the image? OPEN

Hallucination
Is there a person in the image? ✓ ✗

Instruction Following
Create a five-sentence narrative from the perspective of the lamp post, discussing its day, its purpose, and its feelings about the environment it watches over, using personification. OPEN

$$\min_{f} \mathbb{E}_{(i_p, x_p, y_p)\sim\mathcal{D}_P} \left[\mathbb{I}\left(\mathcal{M}(i_p, x_p) = y_p\right) - \mathbb{I}\left(\mathcal{M}^*(i_p, x_p) = y_p\right)\right]$$

Challenges of Evolving Knowledge Injection



Method	**Overall**		**News**		**Entity**	
	Acc ↑	**F1-Score ↑**	**Acc ↑**	**F1-Score ↑**	**Acc ↑**	**F1-Score ↑**
LLaVA-v1.5						
Vanilla	4.89	9.34	7.37	11.96	2.18	6.47
Full-FT	18.02	15.17	21.35	16.34	14.37	13.88
LoRA	15.23	18.31	17.72	19.42	12.51	17.09
MM-RAG$^{\textbf{Text-Only}}$	24.05	34.32	37.32	49.39	9.50	17.80
MM-RAG$^{\textbf{Image-Only}}$	25.25	37.11	19.28	26.76	31.80	48.45
MM-RAG$^{\textbf{UniIR}}$	40.68	57.51	40.12	53.21	41.30	62.23
MM-RAG$^{\textbf{Gloden Context}}$	**56.13**	**75.77**	**56.78**	**72.37**	**55.43**	**79.50**
Qwen-VL-Chat						
Vanilla	5.84	10.99	7.75	12.72	3.74	9.10
Full-FT	10.16	16.61	13.35	18.22	6.65	14.83
LoRA	6.95	12.64	9.27	14.55	4.41	10.54
MM-RAG$^{\textbf{Text-Only}}$	21.79	31.28	31.51	41.14	11.13	20.47
MM-RAG$^{\textbf{Image-Only}}$	22.31	33.09	17.82	25.15	27.24	41.79
MM-RAG$^{\textbf{UniIR}}$	32.75	46.18	33.26	43.36	32.20	49.28
MM-RAG$^{\textbf{Gloden Context}}$	**48.96**	**66.02**	**49.98**	**63.42**	**47.84**	**68.87**
Internet Augmented Generation						
Gemini	18.21	26.52	21.23	27.75	14.91	25.16
Perplexity AI†	**48.27**	**62.44**	**47.58**	**56.51**	**48.96**	**68.78**

1) The non-zero performance of Vanilla.

2) No one knowledge injection method performs exceptionally well.

3) MM-RAG outperforms SFT in overall performance,particularly in cross-modal retrieval.

4) Internet Augmented Generation can help LMMs adapt to evolving knowledge.

Sequential Fine-Tuning based on Tasks

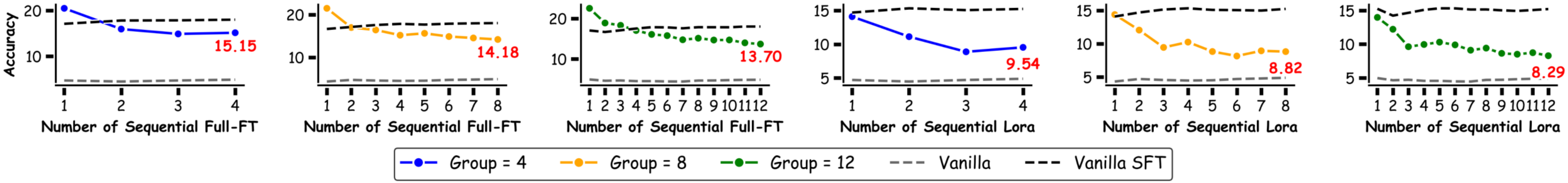


Figure 13. **The results of LLaVA on Sequential Fine-Tuning based on Tasks.** The data $\mathcal{D}_{\mathcal{K}}$ and $\mathcal{D}_{\mathcal{Q}}$ are evenly divided into $g \in \{4, 8, 12\}$ parts, namely $\mathcal{D}_{\mathcal{K}} = \{d_k^1, d_k^2, \ldots, d_k^n\}_{n=1}^{g}$ and $\mathcal{D}_{\mathcal{Q}} = \{d_q^1, d_q^2, \ldots, d_q^n\}_{n=1}^{g}$. Sequential Fine-Tuning based on tasks refer to the situation where if the current m-th Sequential Fine-Tuning has ended, it indicates that the model is being trained on $d_k^1, d_k^2, \ldots, d_k^m$ in sequence; and evaluated on $\{d_q^1 \cup d_q^2 \cup \cdots \cup d_q^m\}$.

Sequential Fine-Tuning based on Subsets

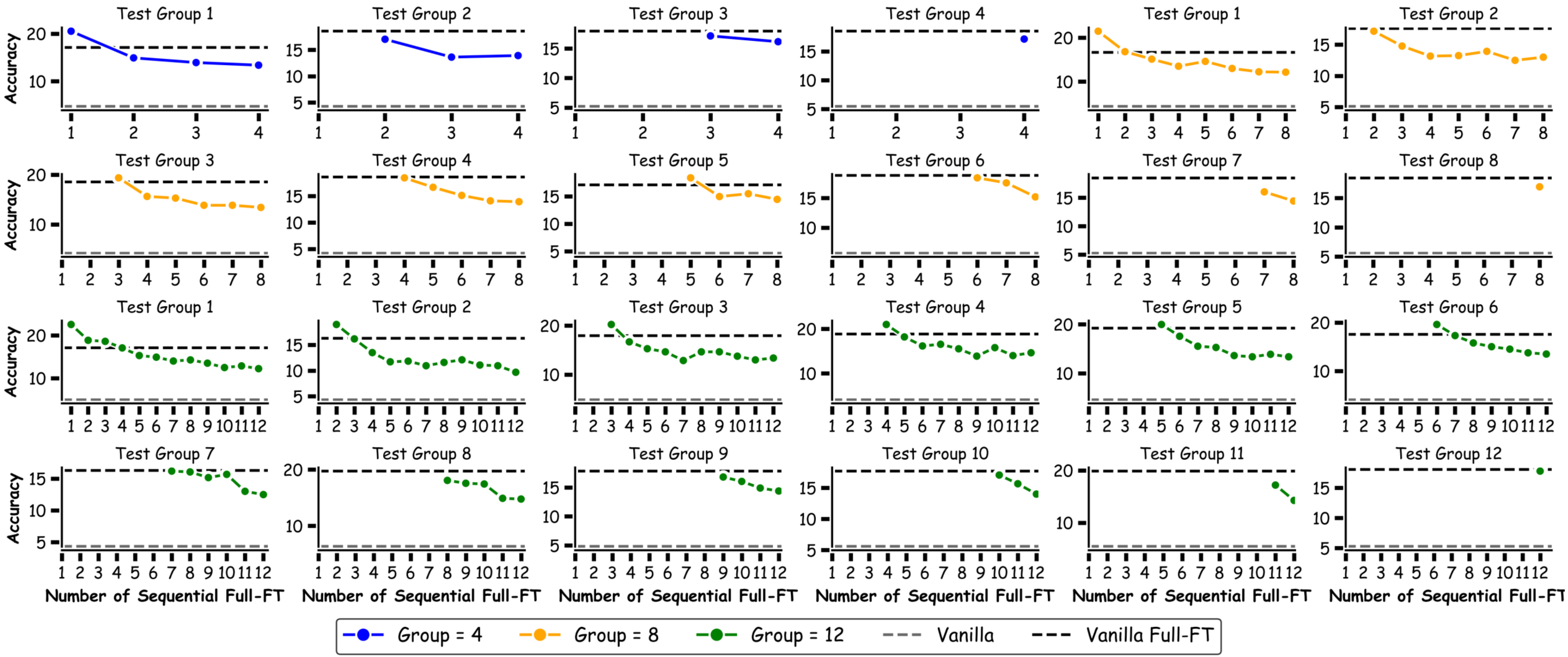
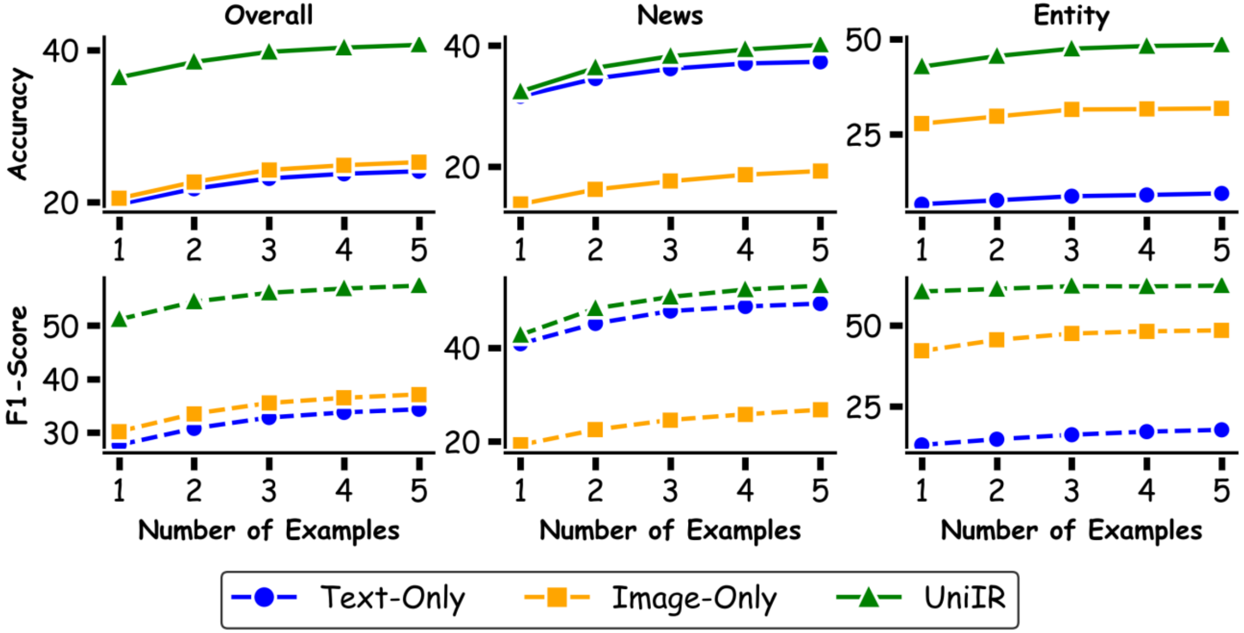
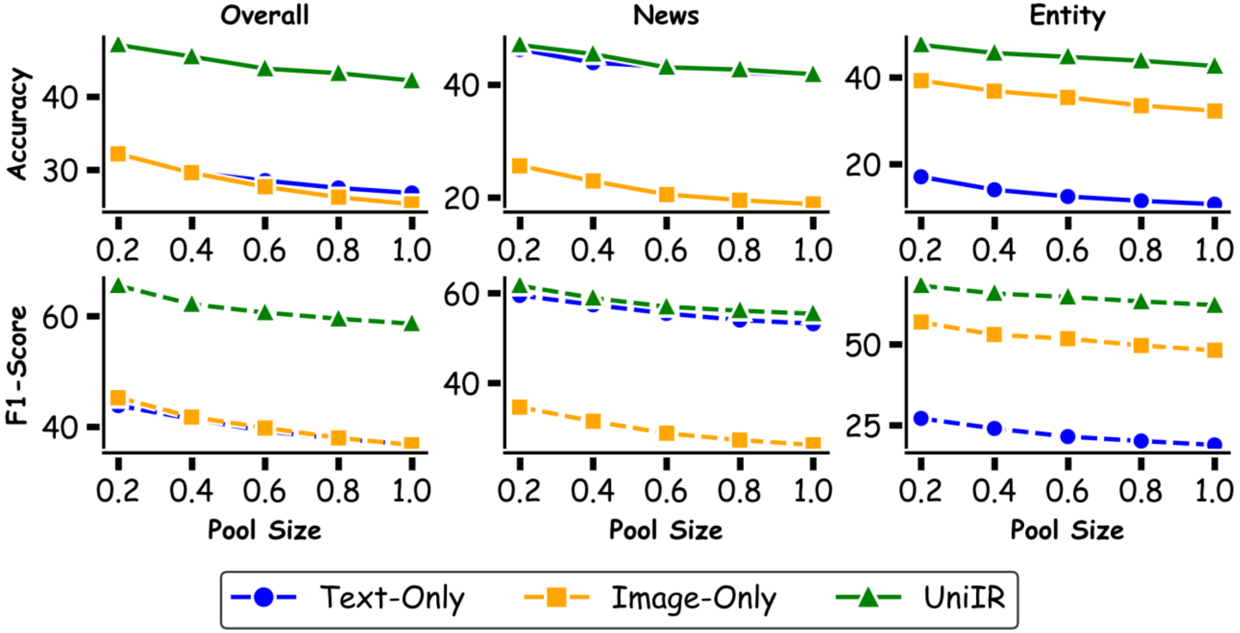


Figure 14. **The results of LLaVA on Sequential Full-FT based on Subsets.** Sequential Full-FT based on subset refer to the situation where if the current m-th Sequential Full-FT has ended, it indicates that the model is being trained on $d_k^1, d_k^2, \ldots, d_k^m$ in sequence; and evaluate sequentially on **one of** $d_q^1, d_q^2, \ldots, d_q^m$.

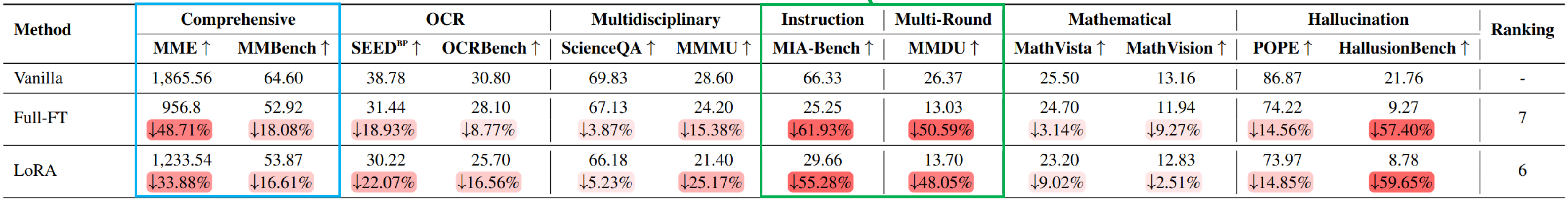
Regarding the Ablation Experiments of MM-RAG





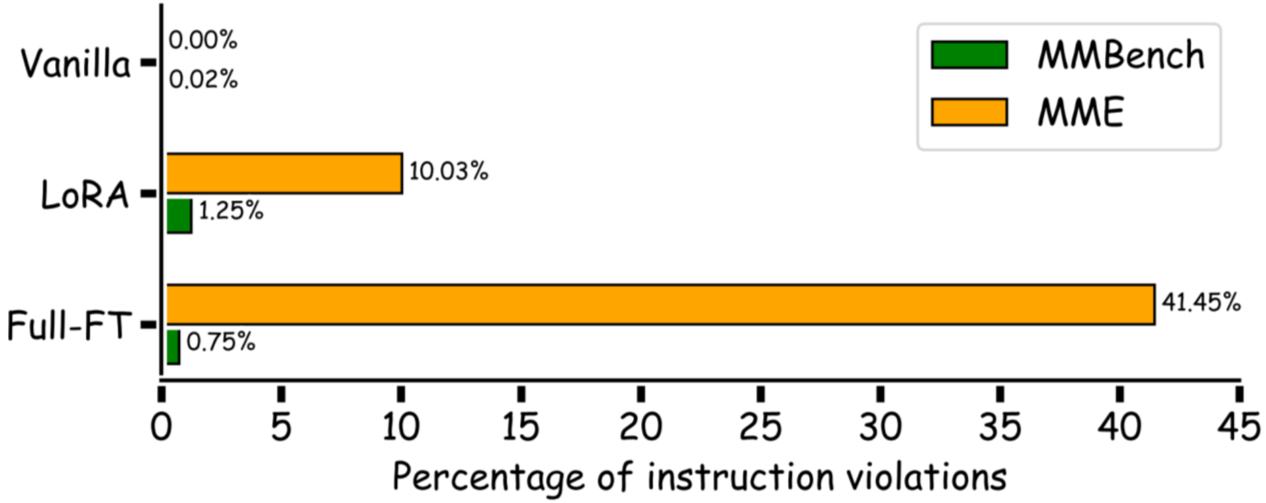
Challenges of Evolving Knowledge Injection

1) The capability for **instruction following** and **multi-round dialogue** significantly deteriorates.



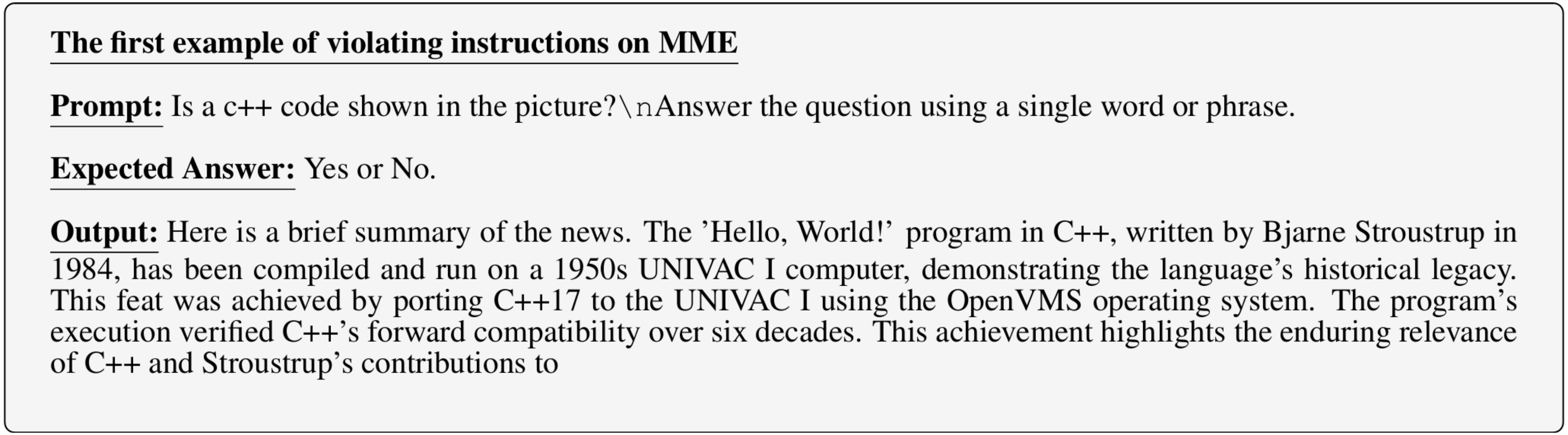
Method	Comprehensive		OCR		Multidisciplinary		Instruction	Multi-Round	Mathematical		Hallucination		Ranking
	MME ↑	MMBench ↑	SEED^{BP} ↑	OCRBench ↑	ScienceQA ↑	MMMU ↑	MIA-Bench ↑	MMDU ↑	MathVista ↑	MathVision ↑	POPE ↑	HallusionBench ↑	
Vanilla	1,865.56	64.60	38.78	30.80	69.83	28.60	66.33	26.37	25.50	13.16	86.87	21.76	-
Full-FT	956.8 ↓48.71%	52.92 ↓18.08%	31.44 ↓18.93%	28.10 ↓8.77%	67.13 ↓3.87%	24.20 ↓15.38%	25.25 ↓61.93%	13.03 ↓50.59%	24.70 ↓3.14%	11.94 ↓9.27%	74.22 ↓14.56%	9.27 ↓57.40%	7
LoRA	1,233.54 ↓33.88%	53.87 ↓16.61%	30.22 ↓22.07%	25.70 ↓16.56%	66.18 ↓5.23%	21.40 ↓25.17%	29.66 ↓55.28%	13.70 ↓48.05%	23.20 ↓9.02%	12.83 ↓2.51%	73.97 ↓14.85%	8.78 ↓59.65%	6

2) Analysis of significant differences in performanceamong benchmarks of the same type.



Reason 1: The degree of damage to the ability to follow instructions varies

Example of violating of Instructions

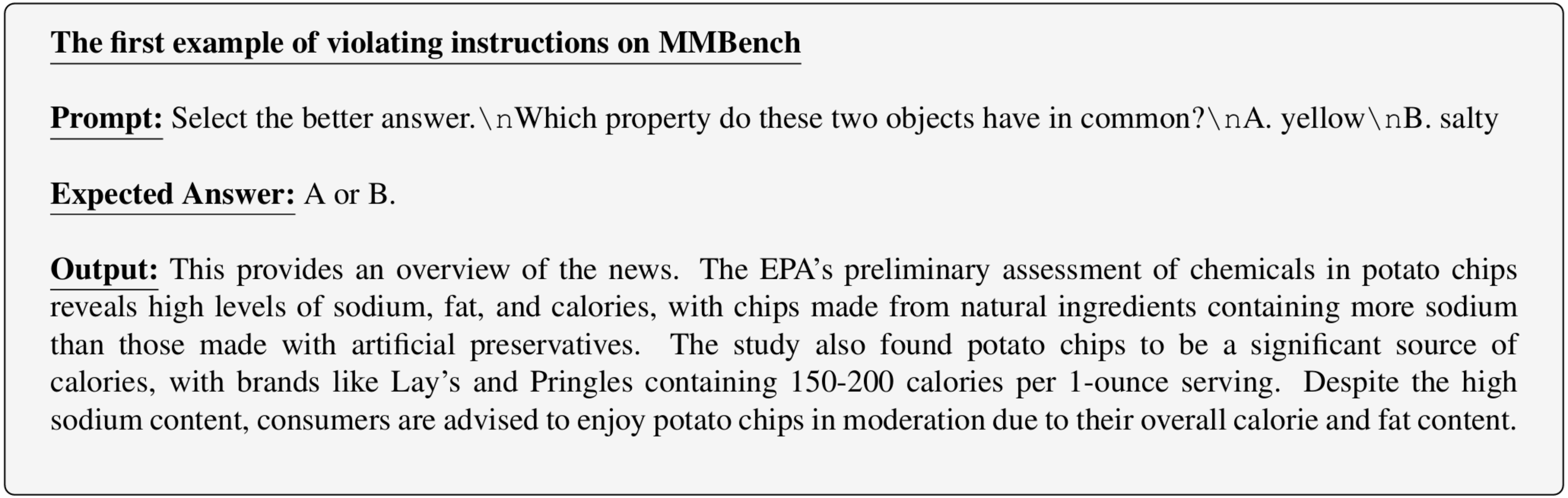


The first example of violating instructions on MME

Prompt: Is a c++ code shown in the picture?`\n`Answer the question using a single word or phrase.

Expected Answer: Yes or No.

Output: Here is a brief summary of the news. The ’Hello, World!’ program in C++, written by Bjarne Stroustrup in 1984, has been compiled and run on a 1950s UNIVAC I computer, demonstrating the language’s historical legacy. This feat was achieved by porting C++17 to the UNIVAC I using the OpenVMS operating system. The program’s execution verified C++’s forward compatibility over six decades. This achievement highlights the enduring relevance of C++ and Stroustrup’s contributions to



The first example of violating instructions on MMBench

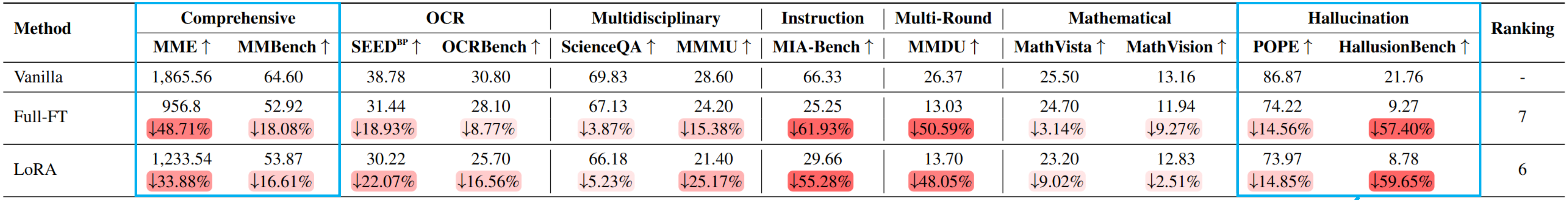
Prompt: Select the better answer.`\n`Which property do these two objects have in common?`\n`A. yellow`\n`B. salty

Expected Answer: A or B.

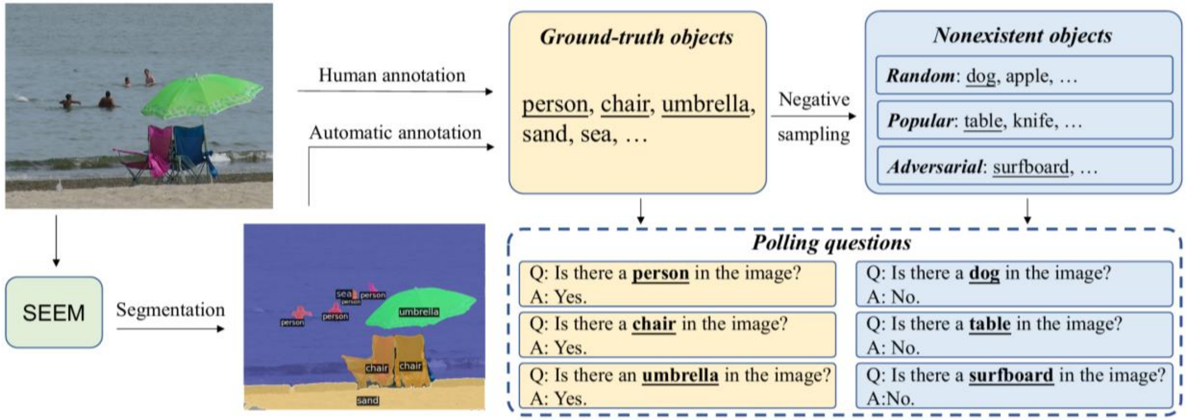
Output: This provides an overview of the news. The EPA’s preliminary assessment of chemicals in potato chips reveals high levels of sodium, fat, and calories, with chips made from natural ingredients containing more sodium than those made with artificial preservatives. The study also found potato chips to be a significant source of calories, with brands like Lay’s and Pringles containing 150-200 calories per 1-ounce serving. Despite the high sodium content, consumers are advised to enjoy potato chips in moderation due to their overall calorie and fat content.

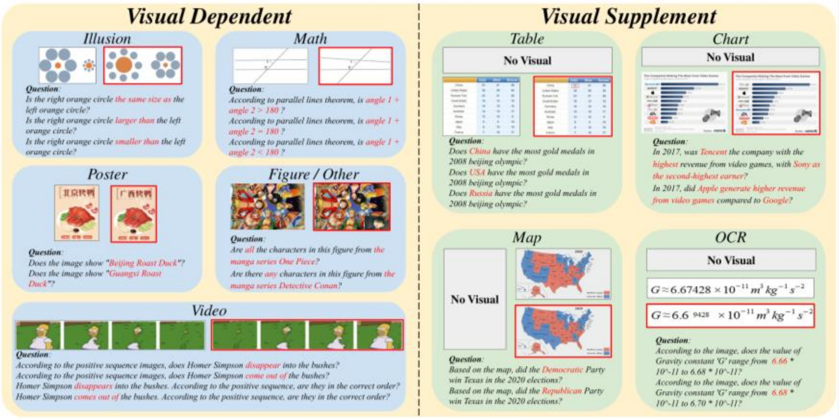
Challenges of Evolving Knowledge Injection

2) Analysis of significant differences in performanceamong benchmarks of the same type.



Method	Comprehensive		OCR		Multidisciplinary		Instruction	Multi-Round	Mathematical		Hallucination		Ranking
	MME ↑	MMBench ↑	SEED^{BP} ↑	OCRBench ↑	ScienceQA ↑	MMMU ↑	MIA-Bench ↑	MMDU ↑	MathVista ↑	MathVision ↑	POPE ↑	HallusionBench ↑	
Vanilla	1,865.56	64.60	38.78	30.80	69.83	28.60	66.33	26.37	25.50	13.16	86.87	21.76	-
Full-FT	956.8 ↓48.71%	52.92 ↓18.08%	31.44 ↓18.93%	28.10 ↓8.77%	67.13 ↓3.87%	24.20 ↓15.38%	25.25 ↓61.93%	13.03 ↓50.59%	24.70 ↓3.14%	11.94 ↓9.27%	74.22 ↓14.56%	9.27 ↓57.40%	7
LoRA	1,233.54 ↓33.88%	53.87 ↓16.61%	30.22 ↓22.07%	25.70 ↓16.56%	66.18 ↓5.23%	21.40 ↓25.17%	29.66 ↓55.28%	13.70 ↓48.05%	23.20 ↓9.02%	12.83 ↓2.51%	73.97 ↓14.85%	8.78 ↓59.65%	6


Human annotation
Automatic annotation
SEEM
Segmentation
sea
person
umbrella
chair
sand
Ground-truth objects
person, chair, umbrella, sand, sea, …
Negative sampling
Nonexistent objects
Random: dog, apple, …
Popular: table, knife, …
Adversarial: surfboard, …
Polling questions
Q: Is there a person in the image?
A: Yes.
Q: Is there a chair in the image?
A: Yes.
Q: Is there an umbrella in the image?
A: Yes.
Q: Is there a dog in the image?
A: No.
Q: Is there a table in the image?
A: No.
Q: Is there a surfboard in the image?
A:No.



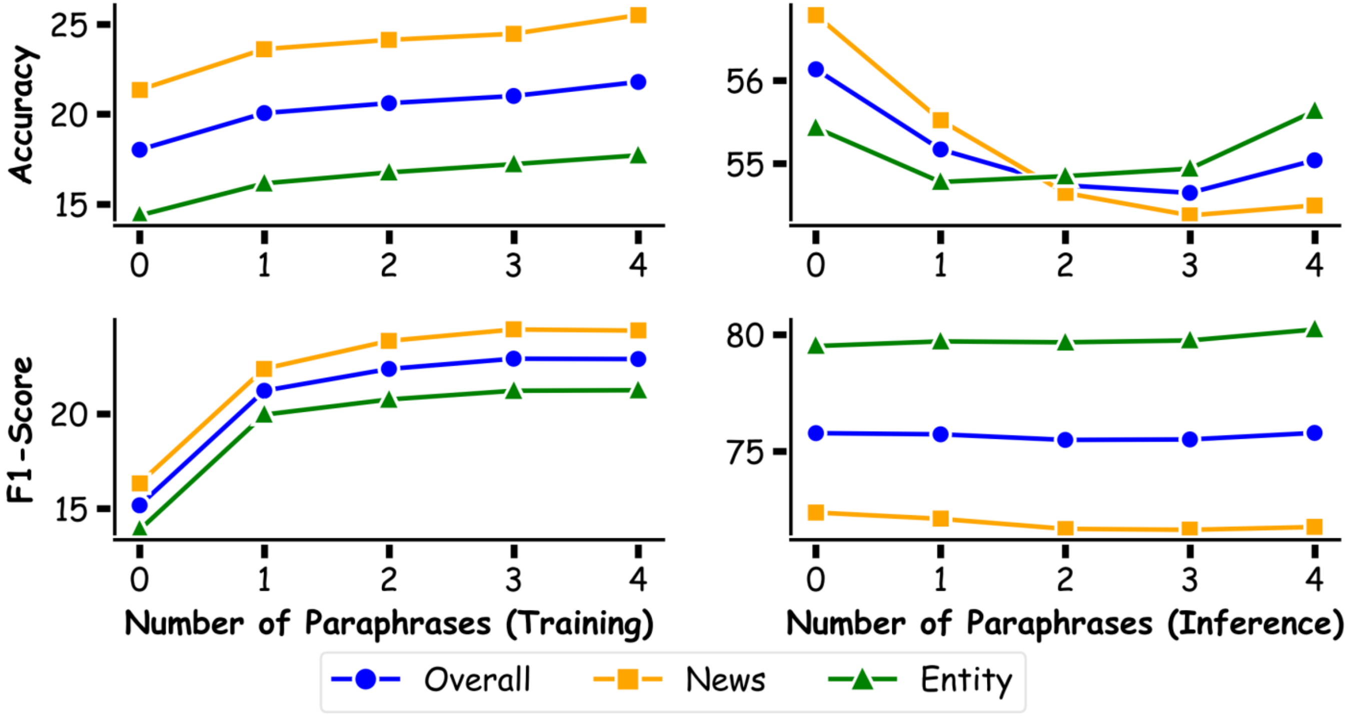
POPE simple yes/no tasks

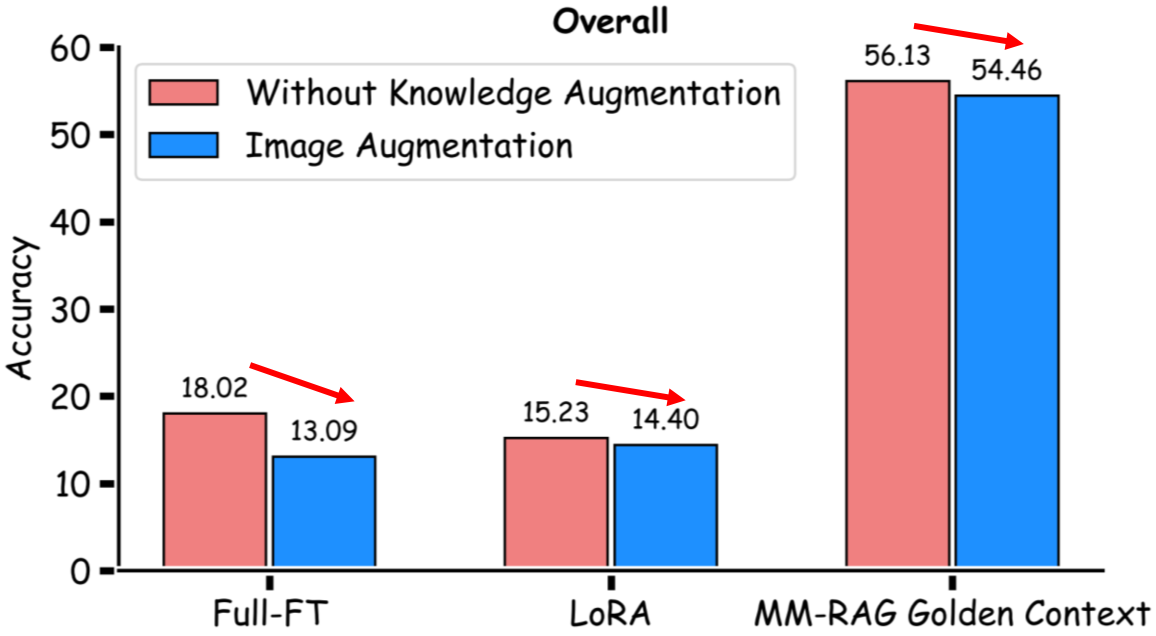
Hallusionbench complex and diverse tasks

Reason 2: The difficulty of tasks for benchmarks of the same type varies

Pathways of Evolving Knowledge Injection

Knowledge Augmentation

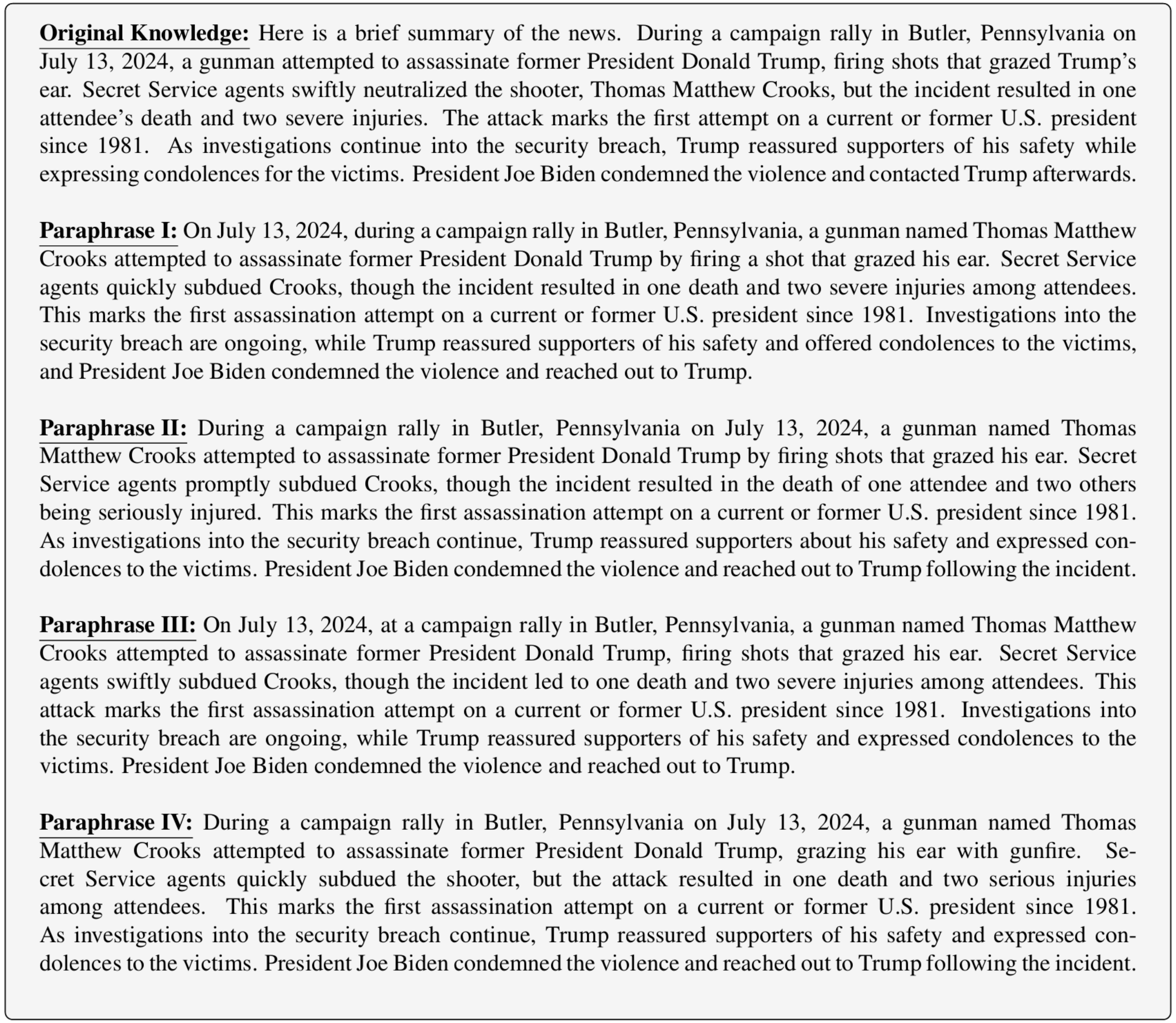




Text Augmentation

Image Augmentation

Paraphrase Examples



Original Knowledge: Here is a brief summary of the news. During a campaign rally in Butler, Pennsylvania on July 13, 2024, a gunman attempted to assassinate former President Donald Trump, firing shots that grazed Trump's ear. Secret Service agents swiftly neutralized the shooter, Thomas Matthew Crooks, but the incident resulted in one attendee's death and two severe injuries. The attack marks the first attempt on a current or former U.S. president since 1981. As investigations continue into the security breach, Trump reassured supporters of his safety while expressing condolences for the victims. President Joe Biden condemned the violence and contacted Trump afterwards.

Paraphrase I: On July 13, 2024, during a campaign rally in Butler, Pennsylvania, a gunman named Thomas Matthew Crooks attempted to assassinate former President Donald Trump by firing a shot that grazed his ear. Secret Service agents quickly subdued Crooks, though the incident resulted in one death and two severe injuries among attendees. This marks the first assassination attempt on a current or former U.S. president since 1981. Investigations into the security breach are ongoing, while Trump reassured supporters of his safety and offered condolences to the victims, and President Joe Biden condemned the violence and reached out to Trump.

Paraphrase II: During a campaign rally in Butler, Pennsylvania on July 13, 2024, a gunman named Thomas Matthew Crooks attempted to assassinate former President Donald Trump by firing shots that grazed his ear. Secret Service agents promptly subdued Crooks, though the incident resulted in the death of one attendee and two others being seriously injured. This marks the first assassination attempt on a current or former U.S. president since 1981. As investigations into the security breach continue, Trump reassured supporters about his safety and expressed condolences to the victims. President Joe Biden condemned the violence and reached out to Trump following the incident.

Paraphrase III: On July 13, 2024, at a campaign rally in Butler, Pennsylvania, a gunman named Thomas Matthew Crooks attempted to assassinate former President Donald Trump, firing shots that grazed his ear. Secret Service agents swiftly subdued Crooks, though the incident led to one death and two severe injuries among attendees. This attack marks the first assassination attempt on a current or former U.S. president since 1981. Investigations into the security breach are ongoing, while Trump reassured supporters of his safety and expressed condolences to the victims. President Joe Biden condemned the violence and reached out to Trump.

Paraphrase IV: During a campaign rally in Butler, Pennsylvania on July 13, 2024, a gunman named Thomas Matthew Crooks attempted to assassinate former President Donald Trump, grazing his ear with gunfire. Secret Service agents quickly subdued the shooter, but the attack resulted in one death and two serious injuries among attendees. This marks the first assassination attempt on a current or former U.S. president since 1981. As investigations into the security breach continue, Trump reassured supporters of his safety and expressed condolences to the victims. President Joe Biden condemned the violence and reached out to Trump following the incident.

Pathways of Evolving Knowledge Injection

Continual Learning for Mitigating Catastrophic Forgetting

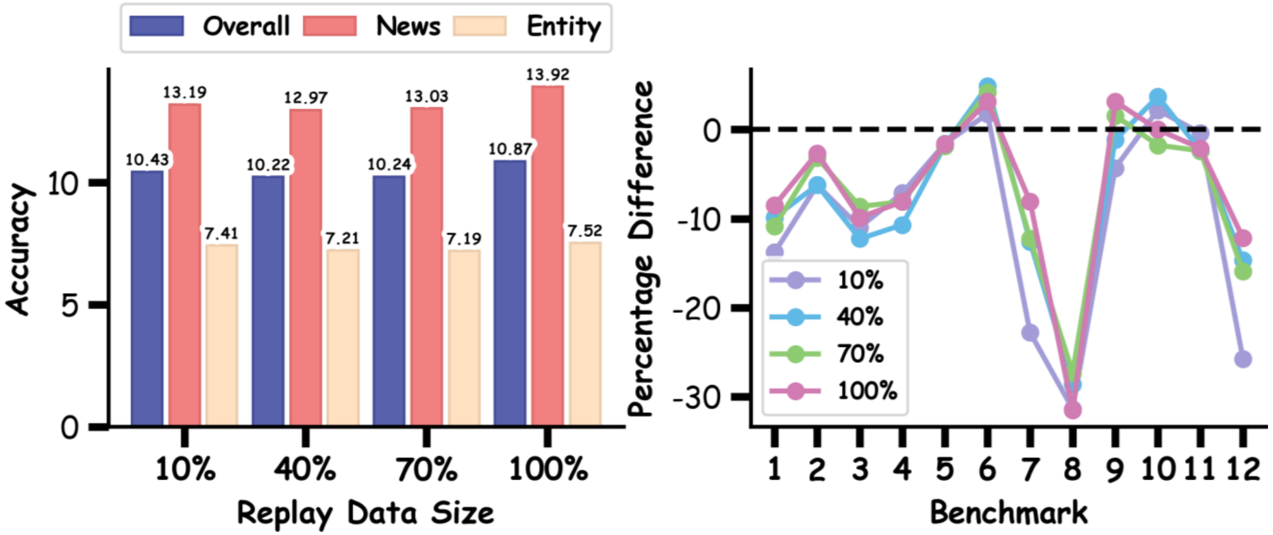
Method	Comprehensive		OCR		Multidisciplinary		Instruction	Multi-Round	Mathematical		Hallucination		Ranking
	MME ↑	MMBench ↑	SEED^{BP} ↑	OCRBench ↑	ScienceQA ↑	MMMU ↑	MIA-Bench ↑	MMDU ↑	MathVista ↑	MathVision ↑	POPE ↑	HallusionBench ↑	
Vanilla	1,865.56	64.60	38.78	30.80	69.83	28.60	66.33	26.37	25.50	13.16	86.87	21.76	-
Full-FT	956.8 ↓48.71%	52.92 ↓18.08%	31.44 ↓18.93%	28.10 ↓8.77%	67.13 ↓3.87%	24.20 ↓15.38%	25.25 ↓61.93%	13.03 ↓50.59%	24.70 ↓3.14%	11.94 ↓9.27%	74.22 ↓14.56%	9.27 ↓57.40%	7
LoRA	1,233.54 ↓33.88%	53.87 ↓16.61%	30.22 ↓22.07%	25.70 ↓16.56%	66.18 ↓5.23%	21.40 ↓25.17%	29.66 ↓55.28%	13.70 ↓48.05%	23.20 ↓9.02%	12.83 ↓2.51%	73.97 ↓14.85%	8.78 ↓59.65%	6
Continual Learning Methods for Mitigating Catastrophic Forgetting													
Replay$^{\textbf{Full-FT}}_{\textbf{+10\%}}$	1,608.00 ↓13.81%	60.57 ↓6.24%	38.69 ↓0.23%	28.60 ↓7.14%	68.74 ↓1.56%	29.10 ↑1.75%	51.20 ↓22.81%	18.09 ↓31.40%	24.40 ↓4.31%	13.45 ↑2.20%	86.52 ↓0.40%	16.15 ↓25.78%	3
Replay$^{\textbf{LoRA}}_{\textbf{+10\%}}$	1,650.75 ↓11.51%	60.48 ↓6.38%	38.34 ↓1.13%	28.60 ↓7.14%	68.77 ↓1.52%	28.50 ↓0.35%	62.33 ↓6.03%	19.31 ↓26.77%	25.20 ↓1.18%	13.13 ↓0.23%	85.44 ↓1.65%	17.90 ↓17.74%	**1**
EWC [21]	1,360.09 ↓27.09%	50.26 ↓22.20%	33.60 ↓13.36%	25.70 ↓16.56%	65.71 ↓5.90%	25.20 ↓11.89%	29.79 ↓55.09%	13.36 ↓49.34%	23.30 ↓8.63%	12.76 ↓3.04%	76.22 ↓12.26%	10.77 ↓50.51%	5
LwF [32]	1,424.41 ↓23.65%	55.41 ↓14.23%	32.02 ↓17.43%	25.60 ↓16.88%	66.21 ↓5.18%	20.60 ↓27.97%	36.19 ↓45.44%	13.68 ↓48.12%	24.40 ↓4.31%	12.04 ↓8.51%	79.23 ↓8.79%	9.13 ↓58.04%	4
MoELoRA [42]	1732.47 ↓7.13%	63.32 ↓1.98%	38.03 ↓1.93%	20.10 ↓34.74%	69.70 ↓0.19%	28.10 ↓1.75%	64.97 ↓2.05%	18.66 ↓29.24%	25.80 ↑1.18%	12.70 ↓3.50%	83.93 ↓3.38%	18.50 ↓14.98%	2

Pathways of Evolving Knowledge Injection

Continual Learning: When the train data is available

Method	Overall		News		Entity	
	Acc ↑	F1-Score ↑	Acc ↑	F1-Score ↑	Acc ↑	F1-Score ↑
Without Continual Learning						
Full-FT	18.02	15.17	21.35	16.34	14.37	13.88
Lora	15.23	18.31	17.72	19.42	12.51	17.09
Train data is available						
Replay$^{\text{Full-FT}}_{+10\%}$	10.43	18.03	13.19	19.60	7.41	16.31
Replay$^{\text{Lora}}_{+10\%}$	10.61	17.98	13.68	19.43	7.25	16.39
Train data is unavailable						
EWC [21]	14.69	19.42	17.49	21.11	11.62	17.56
LwF [32]	13.87	19.98	16.74	21.43	10.73	18.38
MoELoRA [42]	6.82	12.60	9.90	15.42	3.45	9.51

Is it better to have more data for replay?

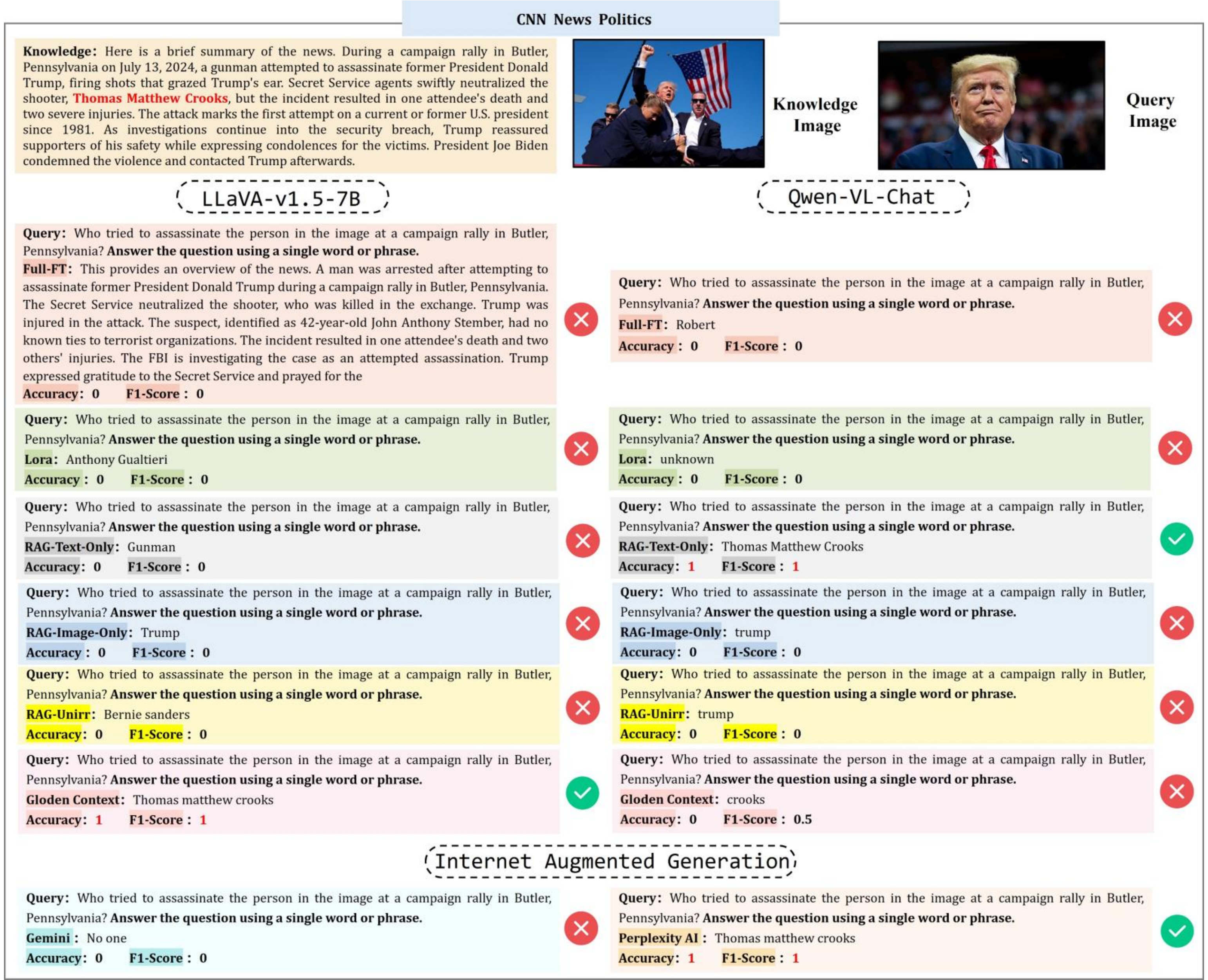


Pathways of Evolving Knowledge Injection

Continual Learning: When the train data is unavailable

Method	**Overall**		**News**		**Entity**	
	Acc ↑	**F1-Score ↑**	**Acc ↑**	**F1-Score ↑**	**Acc ↑**	**F1-Score ↑**
Without Continual Learning						
Full-FT	18.02	15.17	21.35	16.34	14.37	13.88
Lora	15.23	18.31	17.72	19.42	12.51	17.09
Train data is available						
Replay$^{\text{Full-FT}}_{+10\%}$	10.43	18.03	13.19	19.60	7.41	16.31
Replay$^{\text{Lora}}_{+10\%}$	10.61	17.98	13.68	19.43	7.25	16.39
Train data is unavailable						
EWC [21]	14.69	19.42	17.49	21.11	11.62	17.56
LwF [32]	13.87	19.98	16.74	21.43	10.73	18.38
MoELoRA [42]	6.82	12.60	9.90	15.42	3.45	9.51

Qualitative Examples



CNN News Politics

Knowledge: Here is a brief summary of the news. During a campaign rally in Butler, Pennsylvania on July 13, 2024, a gunman attempted to assassinate former President Donald Trump, firing shots that grazed Trump's ear. Secret Service agents swiftly neutralized the shooter, **Thomas Matthew Crooks**, but the incident resulted in one attendee's death and two severe injuries. The attack marks the first attempt on a current or former U.S. president since 1981. As investigations continue into the security breach, Trump reassured supporters of his safety while expressing condolences for the victims. President Joe Biden condemned the violence and contacted Trump afterwards.

LLaVA-v1.5-7B

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Full-FT: This provides an overview of the news. A man was arrested after attempting to assassinate former President Donald Trump during a campaign rally in Butler, Pennsylvania. The Secret Service neutralized the shooter, who was killed in the exchange. Trump was injured in the attack. The suspect, identified as 42-year-old John Anthony Stember, had no known ties to terrorist organizations. The incident resulted in one attendee's death and two others' injuries. The FBI is investigating the case as an attempted assassination. Trump expressed gratitude to the Secret Service and prayed for the
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Lora: Anthony Gualtieri
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
RAG-Text-Only: Gunman
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
RAG-Image-Only: Trump
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
RAG-Unirr: Bernie sanders
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Gloden Context: Thomas matthew crooks
Accuracy: 1 **F1-Score:** 1 ✓

Qwen-VL-Chat

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Full-FT: Robert
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Lora: unknown
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
RAG-Text-Only: Thomas Matthew Crooks
Accuracy: 1 **F1-Score:** 1 ✓

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
RAG-Image-Only: trump
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
RAG-Unirr: trump
Accuracy: 0 **F1-Score:** 0 ✗

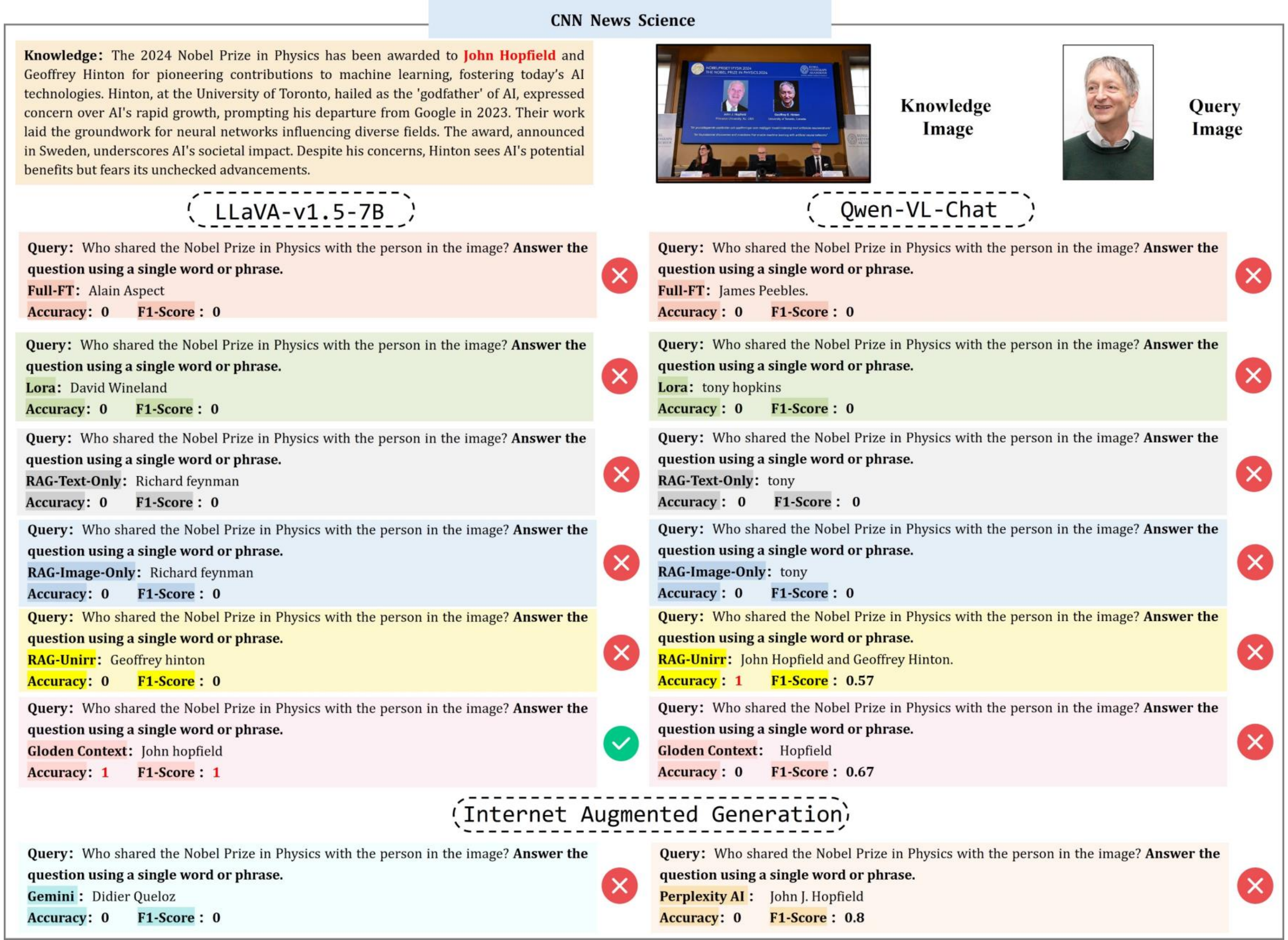
Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Gloden Context: crooks
Accuracy: 0 **F1-Score:** 0.5 ✗

Internet Augmented Generation

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Gemini: No one
Accuracy: 0 **F1-Score:** 0 ✗

Query: Who tried to assassinate the person in the image at a campaign rally in Butler, Pennsylvania? **Answer the question using a single word or phrase.**
Perplexity AI: Thomas matthew crooks
Accuracy: 1 **F1-Score:** 1 ✓

Qualitative Examples



CNN News Science

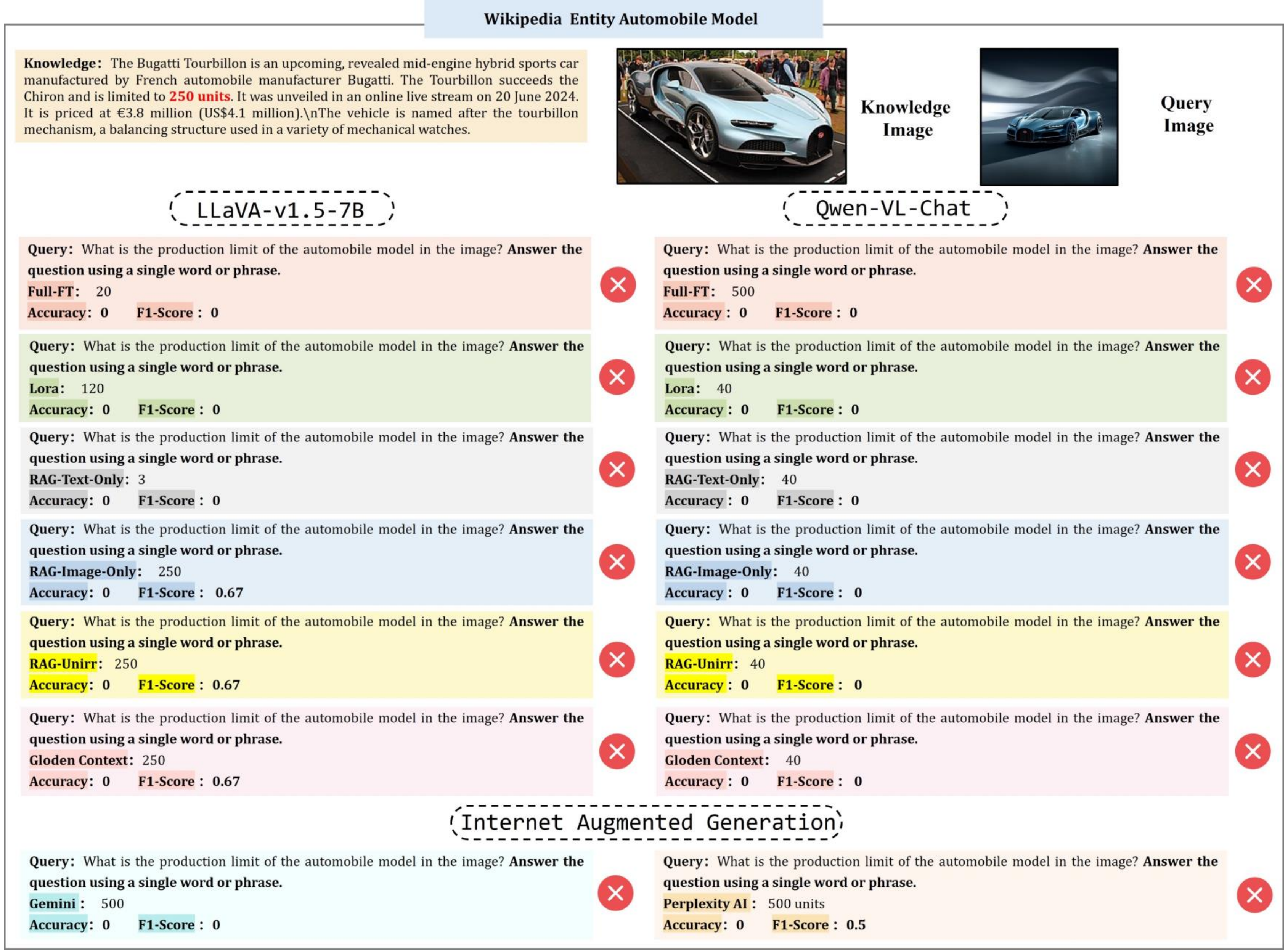
Knowledge: The 2024 Nobel Prize in Physics has been awarded to **John Hopfield** and Geoffrey Hinton for pioneering contributions to machine learning, fostering today's AI technologies. Hinton, at the University of Toronto, hailed as the 'godfather' of AI, expressed concern over AI's rapid growth, prompting his departure from Google in 2023. Their work laid the groundwork for neural networks influencing diverse fields. The award, announced in Sweden, underscores AI's societal impact. Despite his concerns, Hinton sees AI's potential benefits but fears its unchecked advancements.

LLaVA-v1.5-7B		Qwen-VL-Chat	
Query: Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Full-FT:** Alain Aspect **Accuracy:** 0 **F1-Score :** 0	✗	**Query:** Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Full-FT:** James Peebles. **Accuracy :** 0 **F1-Score :** 0	✗
Query: Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Lora:** David Wineland **Accuracy:** 0 **F1-Score :** 0	✗	**Query:** Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Lora:** tony hopkins **Accuracy :** 0 **F1-Score :** 0	✗
Query: Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **RAG-Text-Only:** Richard feynman **Accuracy:** 0 **F1-Score :** 0	✗	**Query:** Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **RAG-Text-Only:** tony **Accuracy :** 0 **F1-Score :** 0	✗
Query: Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **RAG-Image-Only:** Richard feynman **Accuracy:** 0 **F1-Score :** 0	✗	**Query:** Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **RAG-Image-Only:** tony **Accuracy :** 0 **F1-Score :** 0	✗
Query: Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **RAG-Unirr:** Geoffrey hinton **Accuracy:** 0 **F1-Score :** 0	✗	**Query:** Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **RAG-Unirr:** John Hopfield and Geoffrey Hinton. **Accuracy :** 1 **F1-Score :** 0.57	✗
Query: Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Gloden Context:** John hopfield **Accuracy:** 1 **F1-Score :** 1	✓	**Query:** Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Gloden Context:** Hopfield **Accuracy :** 0 **F1-Score :** 0.67	✗

Internet Augmented Generation

Query: Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Gemini :** Didier Queloz **Accuracy:** 0 **F1-Score :** 0	✗	**Query:** Who shared the Nobel Prize in Physics with the person in the image? **Answer the question using a single word or phrase.** **Perplexity AI :** John J. Hopfield **Accuracy:** 0 **F1-Score :** 0.8	✗

Qualitative Examples



Wikipedia Entity Automobile Model

Knowledge: The Bugatti Tourbillon is an upcoming, revealed mid-engine hybrid sports car manufactured by French automobile manufacturer Bugatti. The Tourbillon succeeds the Chiron and is limited to **250 units**. It was unveiled in an online live stream on 20 June 2024. It is priced at €3.8 million (US$4.1 million).\nThe vehicle is named after the tourbillon mechanism, a balancing structure used in a variety of mechanical watches.

`LLaVA-v1.5-7B`

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Full-FT: 20
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Lora: 120
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
RAG-Text-Only: 3
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
RAG-Image-Only: 250
Accuracy: 0 **F1-Score:** 0.67
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
RAG-Unirr: 250
Accuracy: 0 **F1-Score:** 0.67
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Gloden Context: 250
Accuracy: 0 **F1-Score:** 0.67
✗

`Qwen-VL-Chat`

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Full-FT: 500
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Lora: 40
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
RAG-Text-Only: 40
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
RAG-Image-Only: 40
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
RAG-Unirr: 40
Accuracy: 0 **F1-Score:** 0
✗

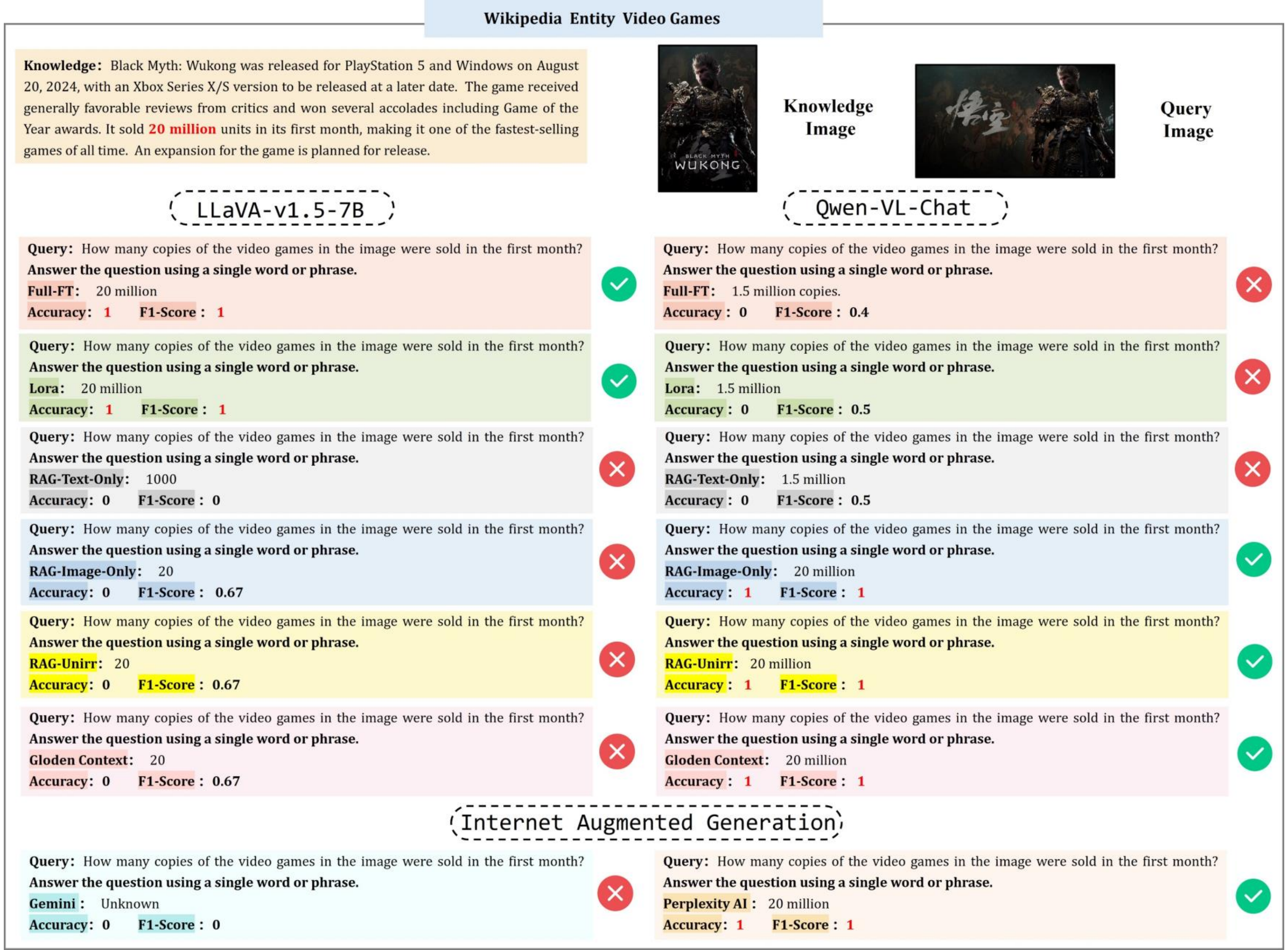
Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Gloden Context: 40
Accuracy: 0 **F1-Score:** 0
✗

`Internet Augmented Generation`

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Gemini: 500
Accuracy: 0 **F1-Score:** 0
✗

Query: What is the production limit of the automobile model in the image? **Answer the question using a single word or phrase.**
Perplexity AI: 500 units
Accuracy: 0 **F1-Score:** 0.5
✗

Qualitative Examples



Wikipedia Entity Video Games

Knowledge: Black Myth: Wukong was released for PlayStation 5 and Windows on August 20, 2024, with an Xbox Series X/S version to be released at a later date. The game received generally favorable reviews from critics and won several accolades including Game of the Year awards. It sold **20 million** units in its first month, making it one of the fastest-selling games of all time. An expansion for the game is planned for release.

LLaVA-v1.5-7B

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Full-FT: 20 million
Accuracy: 1 **F1-Score:** 1 ✓

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Lora: 20 million
Accuracy: 1 **F1-Score:** 1 ✓

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
RAG-Text-Only: 1000
Accuracy: 0 **F1-Score:** 0 ✗

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
RAG-Image-Only: 20
Accuracy: 0 **F1-Score:** 0.67 ✗

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
RAG-Unirr: 20
Accuracy: 0 **F1-Score:** 0.67 ✗

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Gloden Context: 20
Accuracy: 0 **F1-Score:** 0.67 ✗

Qwen-VL-Chat

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Full-FT: 1.5 million copies.
Accuracy: 0 **F1-Score:** 0.4 ✗

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Lora: 1.5 million
Accuracy: 0 **F1-Score:** 0.5 ✗

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
RAG-Text-Only: 1.5 million
Accuracy: 0 **F1-Score:** 0.5 ✗

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
RAG-Image-Only: 20 million
Accuracy: 1 **F1-Score:** 1 ✓

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
RAG-Unirr: 20 million
Accuracy: 1 **F1-Score:** 1 ✓

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Gloden Context: 20 million
Accuracy: 1 **F1-Score:** 1 ✓

Internet Augmented Generation

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Gemini: Unknown
Accuracy: 0 **F1-Score:** 0 ✗

Query: How many copies of the video games in the image were sold in the first month? **Answer the question using a single word or phrase.**
Perplexity AI: 20 million
Accuracy: 1 **F1-Score:** 1 ✓