



# When Large Multimodal Models Confront Evolving Knowledge: Challenges and Pathways

Kailin Jiang<sup>\* 1,2</sup> Yuntao Du<sup>\* 3,2</sup> Yukai Ding<sup>4,2</sup> Yuchen Ren<sup>5</sup> Ning Jiang<sup>6</sup>

Zhi Gao<sup>7,2</sup> Zilong Zheng<sup>2</sup> Lei Liu<sup>1</sup>  Bin Li<sup>1</sup> Qing Li<sup>2</sup> 

<sup>1</sup>University of Science and Technology of China, USTC    <sup>2</sup>State Key Laboratory of General Artificial Intelligence, BIGAI

<sup>3</sup>Shandong University, SDU

<sup>4</sup>Wuhan University, WHU

<sup>5</sup>The University of Sydney, USYD

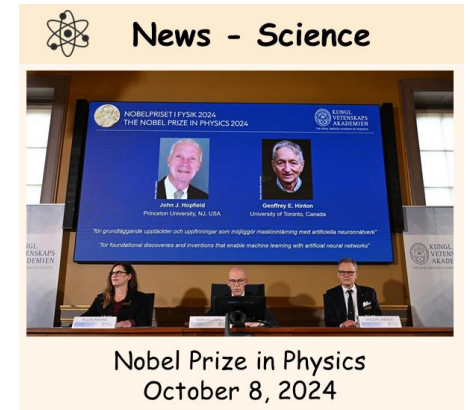
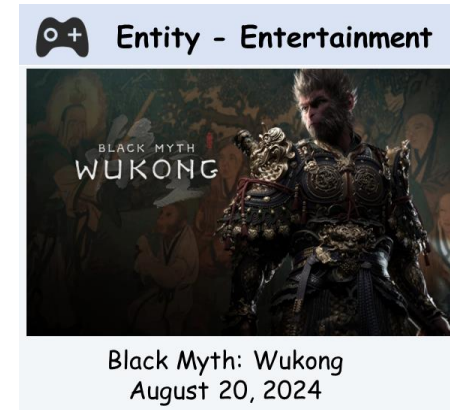
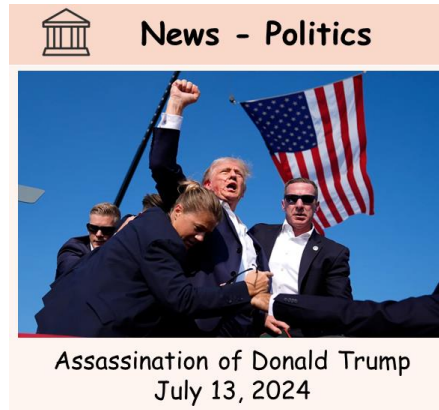
<sup>6</sup>Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education    <sup>7</sup>Peking University, PKU



Machine Learning Lab , BIGAI

# Background

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

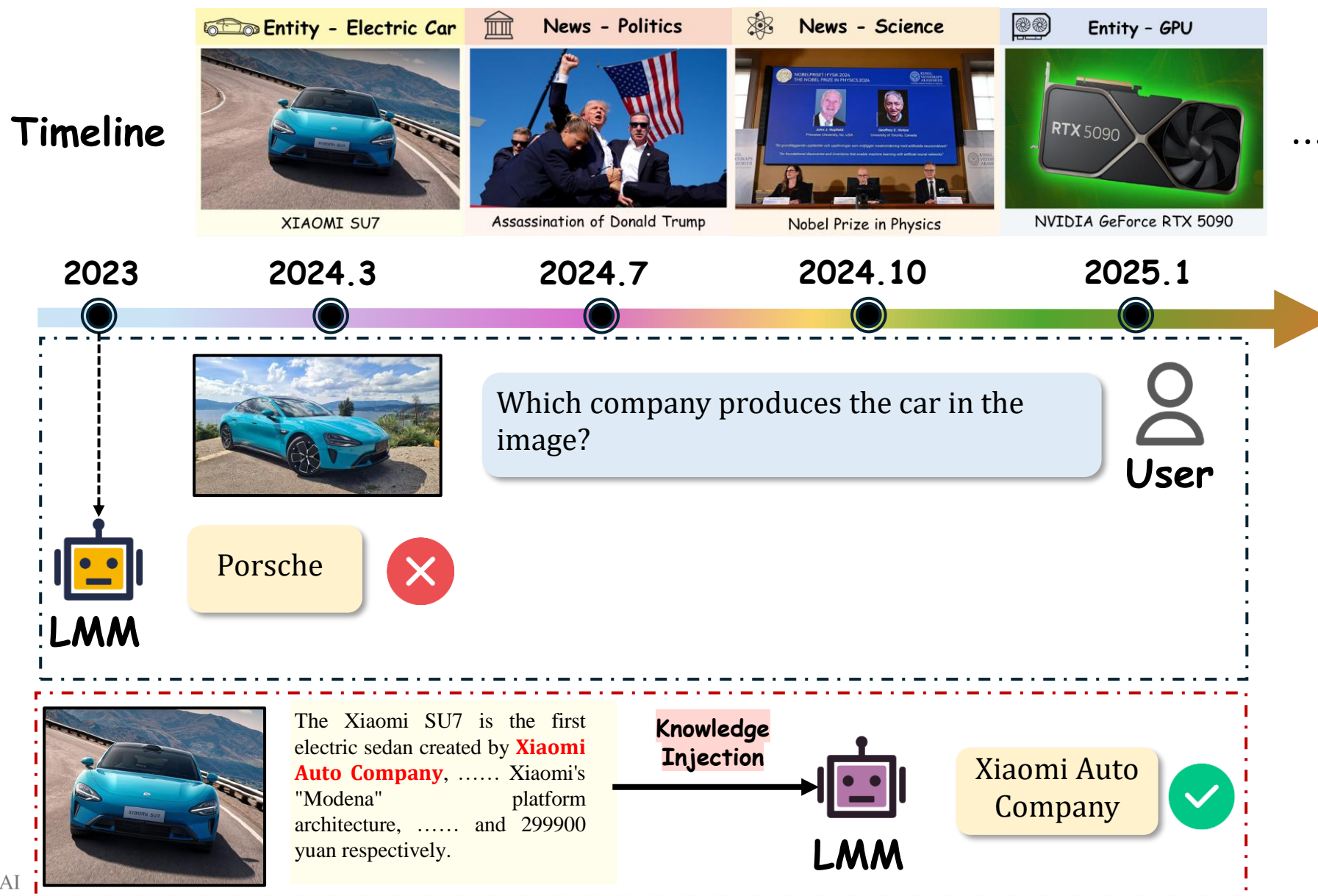


## Timeline







# Knowledge Injection: How LMMs adapt to multimodal evolving knowledge



# 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?



<Knowledge Image>




<Query Image>


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>


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>

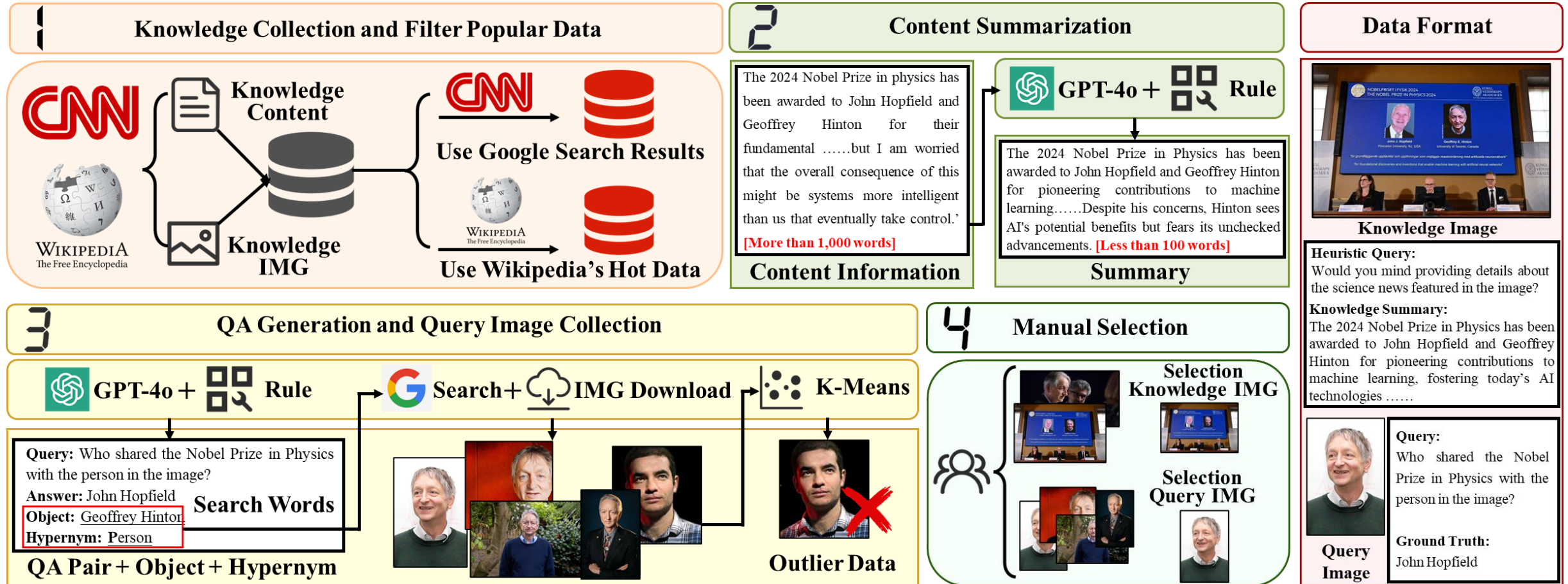
Type: **Science**

**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?



# Evolving Knowledge Benchmark Construction Pipeline



# 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 <sup>Text-Only</sup>	24.05	34.32	37.32	49.39	9.50	17.80
MM-RAG <sup>Image-Only</sup>	25.25	37.11	19.28	26.76	31.80	48.45
MM-RAG <sup>UniIR</sup>	40.68	57.51	40.12	53.21	41.30	62.23
MM-RAG <sup>Gloden Context</sup>	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 <sup>Text-Only</sup>	21.79	31.28	31.51	41.14	11.13	20.47
MM-RAG <sup>Image-Only</sup>	22.31	33.09	17.82	25.15	27.24	41.79
MM-RAG <sup>UniIR</sup>	32.75	46.18	33.26	43.36	32.20	49.28
MM-RAG <sup>Gloden Context</sup>	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 <sup>†</sup>	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.

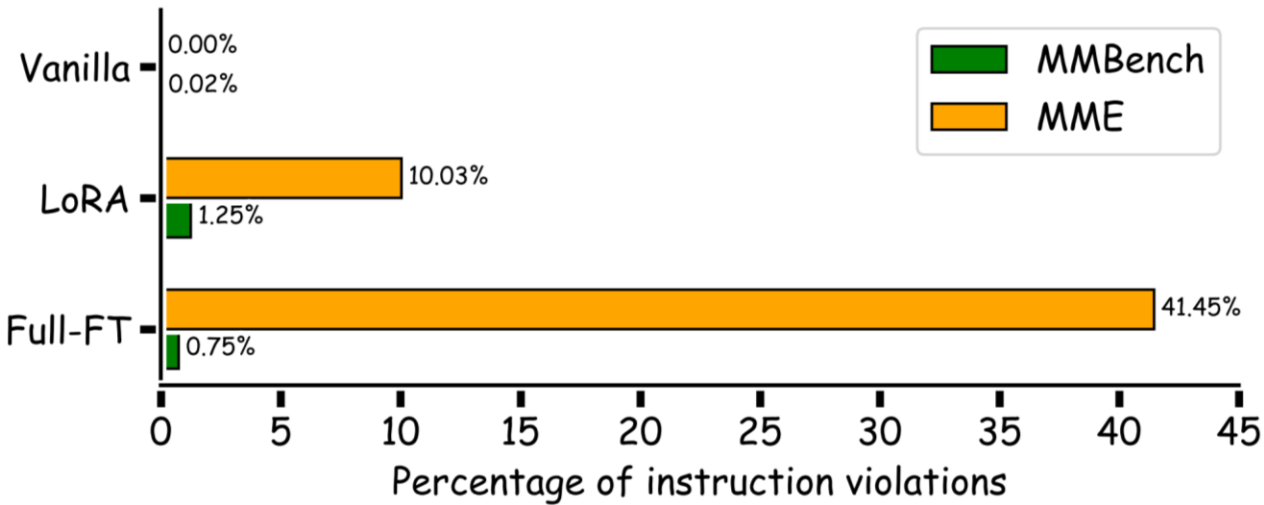
4) Internet Augmented Generation can help LMMs adapt to evolving knowledge.

# 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 <sup>BP</sup> ↑	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 performance among benchmarks of the same type.

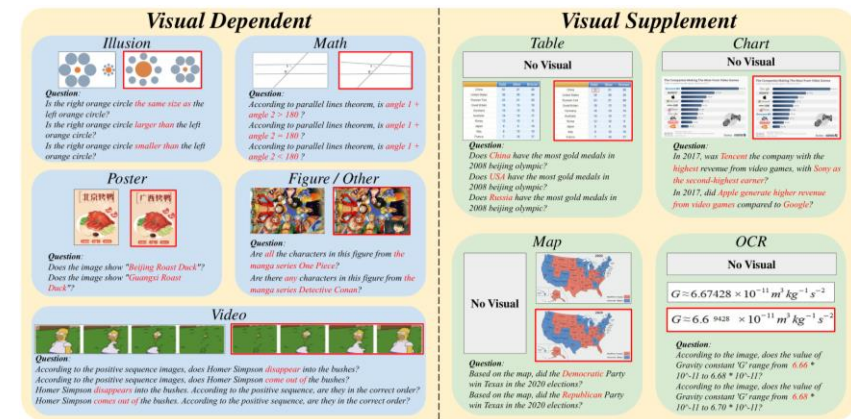
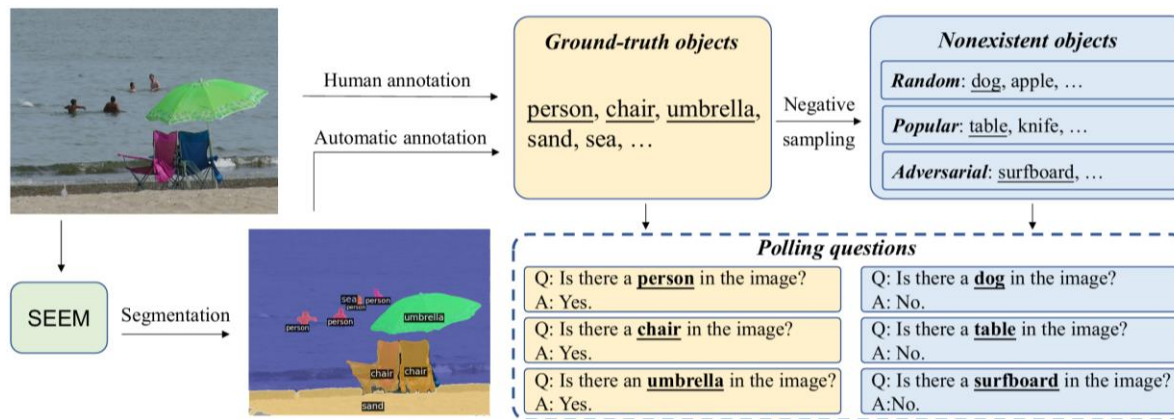


**Reason 1:** The degree of damage to the ability to follow instructions varies

# Challenges of Evolving Knowledge Injection

## 2) Analysis of significant differences in performance among benchmarks of the same type.

Method	Comprehensive		OCR		Multidisciplinary		Instruction	Multi-Round	Mathematical		Hallucination		Ranking
	MME ↑	MMBench ↑	SEED <sup>BP</sup> ↑	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



POPE simple yes/no tasks

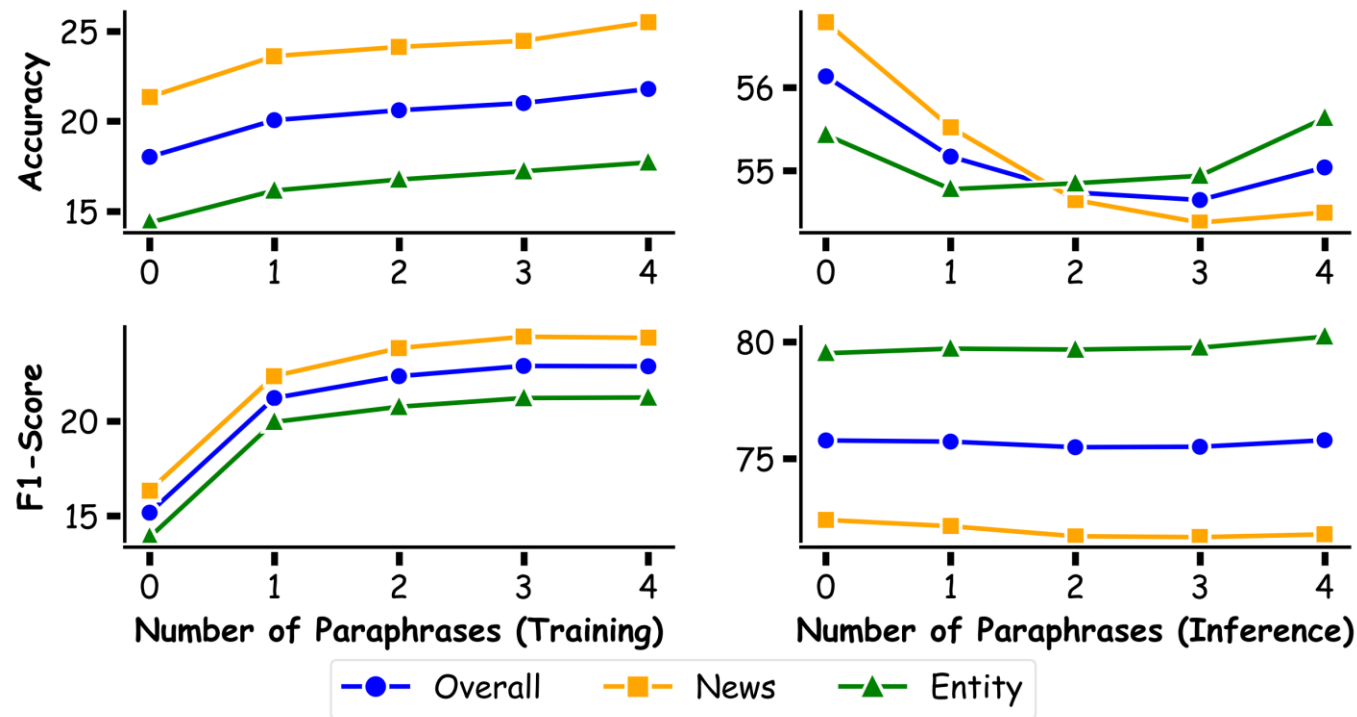
Hallusionbench complex and diverse tasks

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

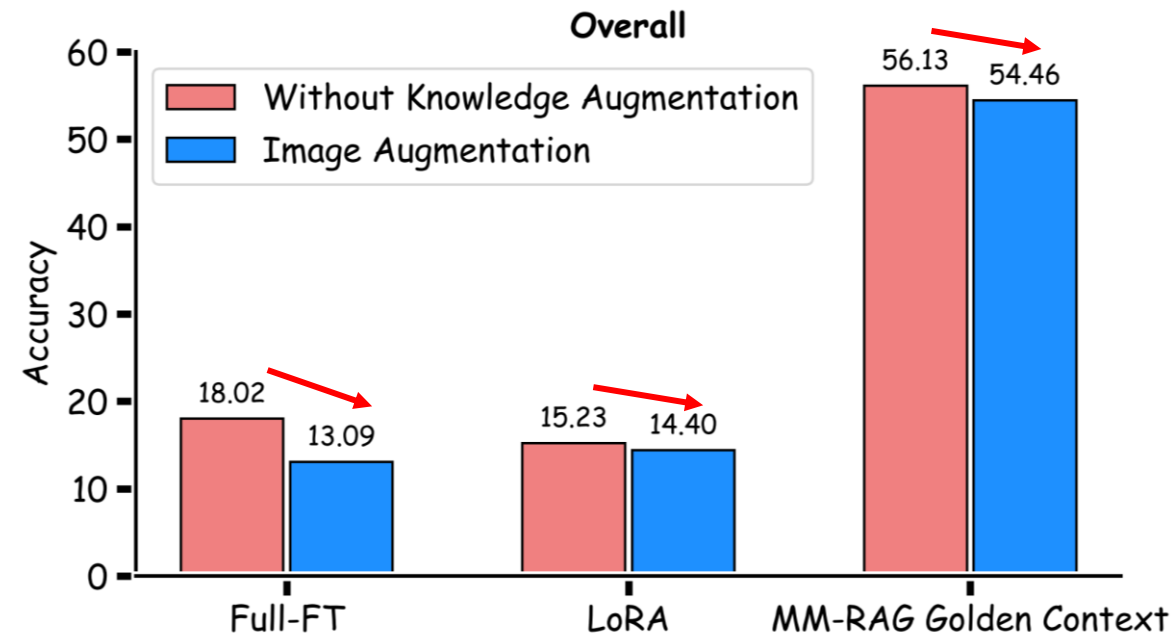


# Pathways of Evolving Knowledge Injection

## Knowledge Augmentation



## Text Augmentation



## Image Augmentation

# Pathways of Evolving Knowledge Injection


## Continual Learning for Mitigating Catastrophic Forgetting

Method	Comprehensive		OCR		Multidisciplinary		Instruction	Multi-Round	Mathematical		Hallucination		Ranking
	MME ↑	MMBench ↑	SEED <sup>BP</sup> ↑	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 <sup>Full-FT</sup> <sub>+10%</sub>	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 <sup>LoRA</sup> <sub>+10%</sub>	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


# 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.



Knowledge Image



Query Image

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