



# KORE

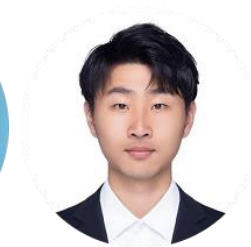
## Enhancing Knowledge Injection for Large Multimodal Models via Knowledge-Oriented Augmentations and Constraints

Kailin Jiang<sup>1,2</sup> Hongbo Jiang<sup>3</sup> Ning Jiang<sup>4</sup> Zhi Gao<sup>5,2</sup> Jinhe Bi<sup>6</sup>  
Yuchen Ren<sup>7</sup> Bin Li<sup>1</sup> Yuntao Du<sup>8</sup> Lei Liu<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>Xiamen University, XMU    <sup>4</sup>Northeast Forestry University, NEFU    <sup>5</sup>Beijing Institute of Technology, BIT

<sup>6</sup>Ludwig Maximilian University of Munich, LMU    <sup>7</sup>The University of Sydney, USYD    <sup>8</sup>Shandong University, SDU




# Background: Knowledge Adaptation


Released LMMs can't keep pace with evolving knowledge.

 Entity - Sports Car



Xiaomi Su7  
March 28, 2024

 News - Entertainment



Black Myth: Wukong  
August 20, 2024

 News - Science



Nobel Prize in Physics  
October 8, 2024

 Entity - Film



NE ZHA2  
January 29, 2025

 Entity - Gameboy



Nintendo Switch 2  
June 5, 2025

 News - Politics



Assassination of Donald Trump  
July 13, 2024

 Entity - Product



iPhone 16  
September 10, 2024

 Entity - Music



APT.  
October 18, 2024

 News - Technology

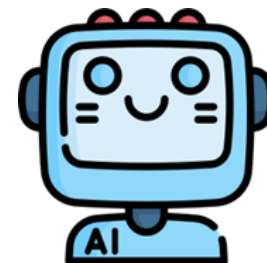


SpaceX Falcon 9 First Recovery  
February 19, 2025

 Entity - Sports Car



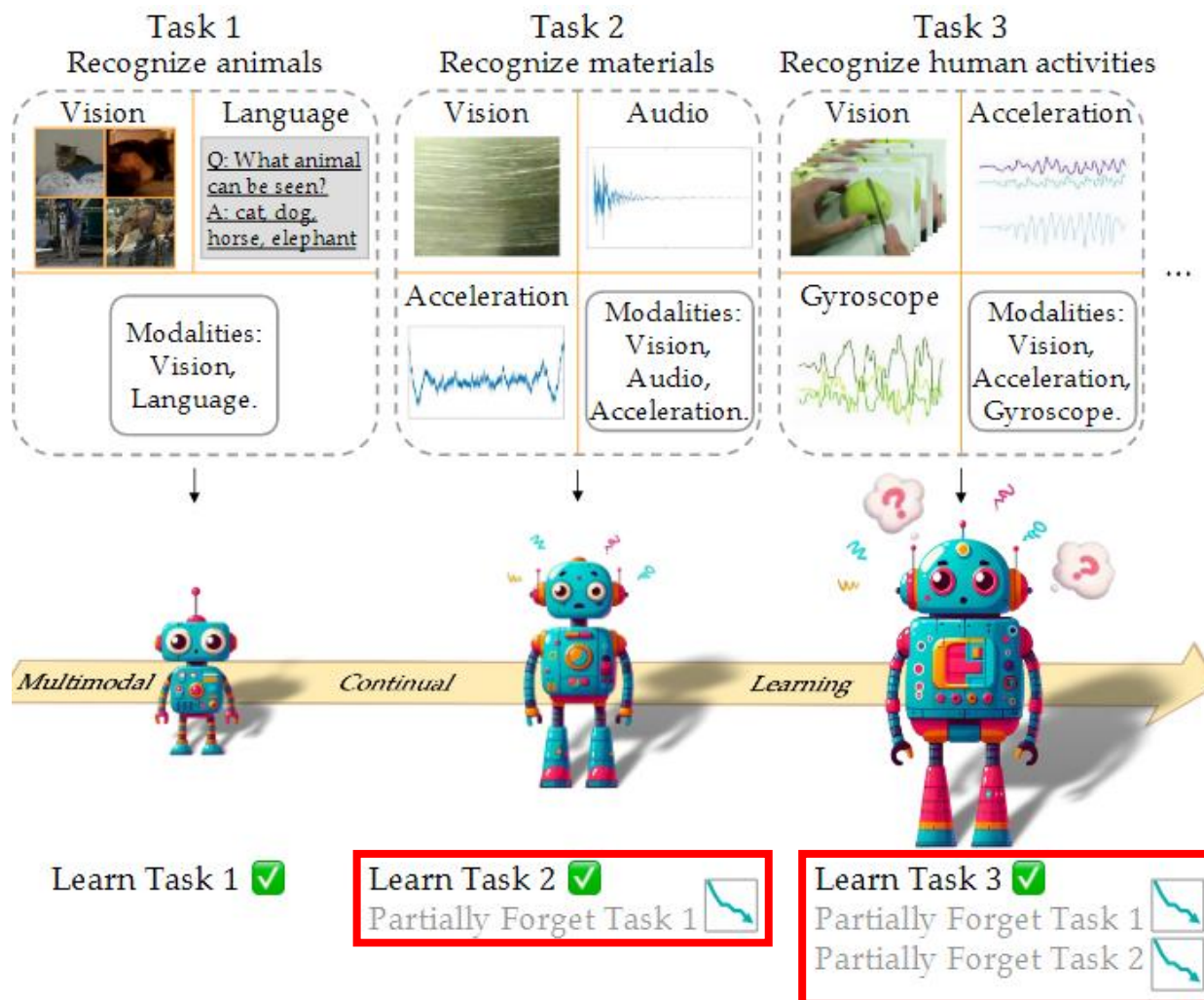
Xiaomi Yu7  
June 26, 2025



LMM

# Background: Knowledge Retention

Injecting new knowledge leads to catastrophic forgetting, causing model to forget its previous abilities and knowledge



# Teaser: Accurate Adaptation and Powerful Retention

## Knowledge Adaptation

**Knowledge:** During a campaign rally in Butler, Pennsylvania on July 13, 2024, a gunman attempted to assassinate former President Donald Trump ... **Thomas Matthew Crooks**, but the incident resulted in one attendee's death ...



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



Answer with a single word or phrase.

Expected: Thomas Matthew Crooks

## Knowledge Retention



Where is the capital of the country in the image?

- A: Washington
- B: New York City
- C: Philadelphia
- D: Los Angeles

Answer with the option's letter from the given choices directly.

Expected: A

### Poor Generalization

**Full-FT:** A man was arrested after attempting to assassinate ... (Overfitting) ❌

**EWC:** Omar Abdel-Rahman (Irrelevant Answer) ❌

### Current Methods

**Full-FT:** Paris (Factual Forgetting) ❌

**EWC:** Washington (Instruction Forgetting) ❌

### Catastrophic Forgetting

### KORE-Augmentation

- Multi-rounds of Dialogue
- Instruction Tasks

### Precision Adaptation

: Thomas Matthew Crooks ✓

### KORE (Ours)

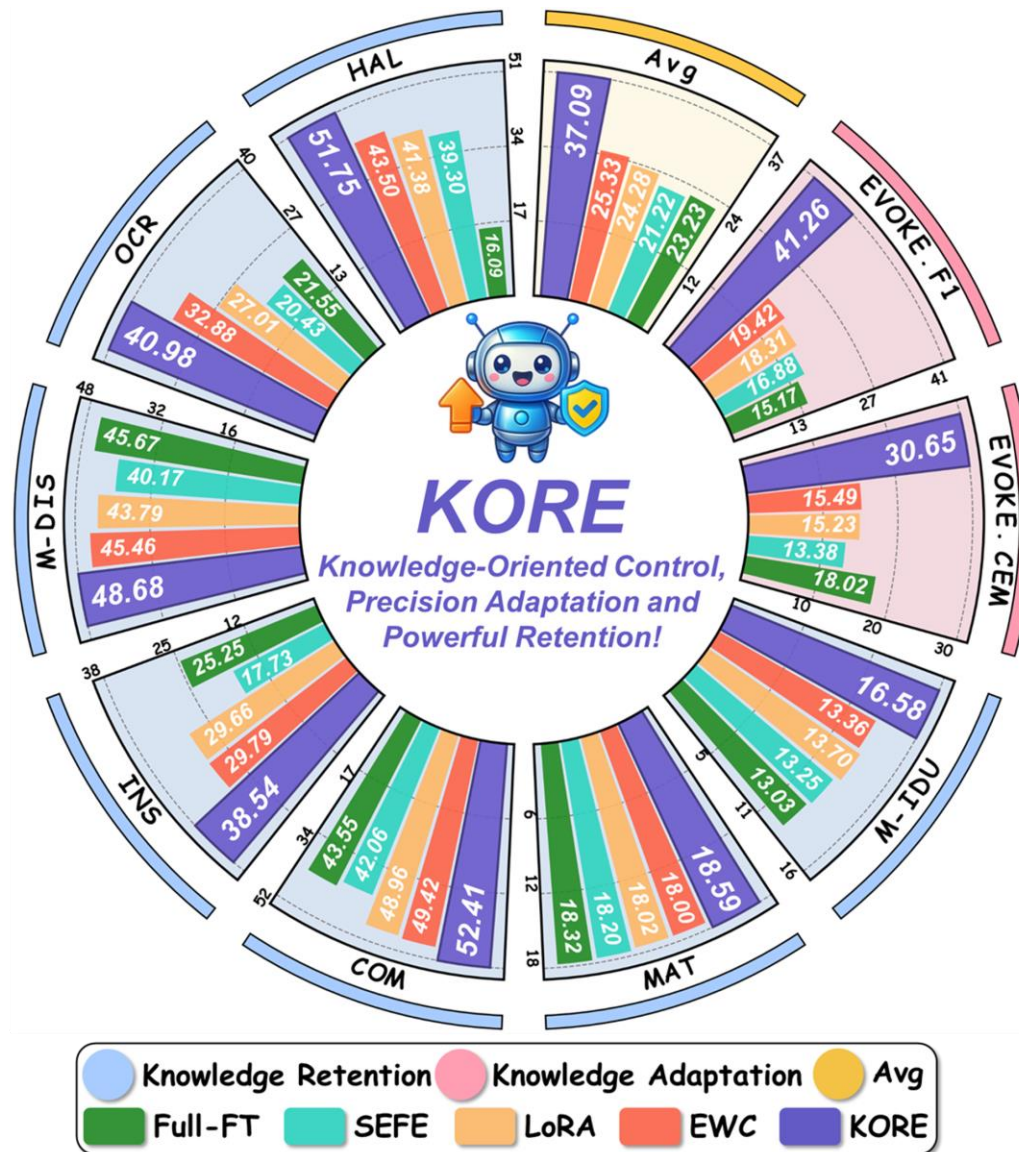


### KORE-Constraint

- Covariance Matrix
- Null Space Constraint

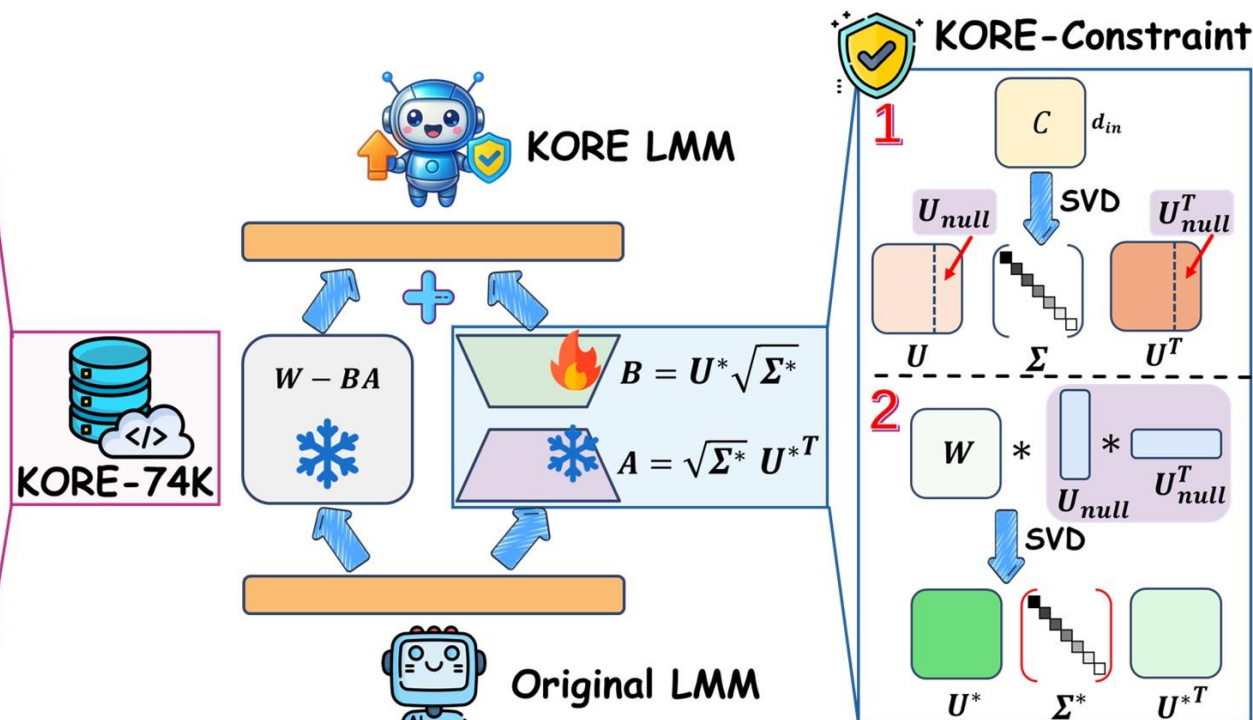
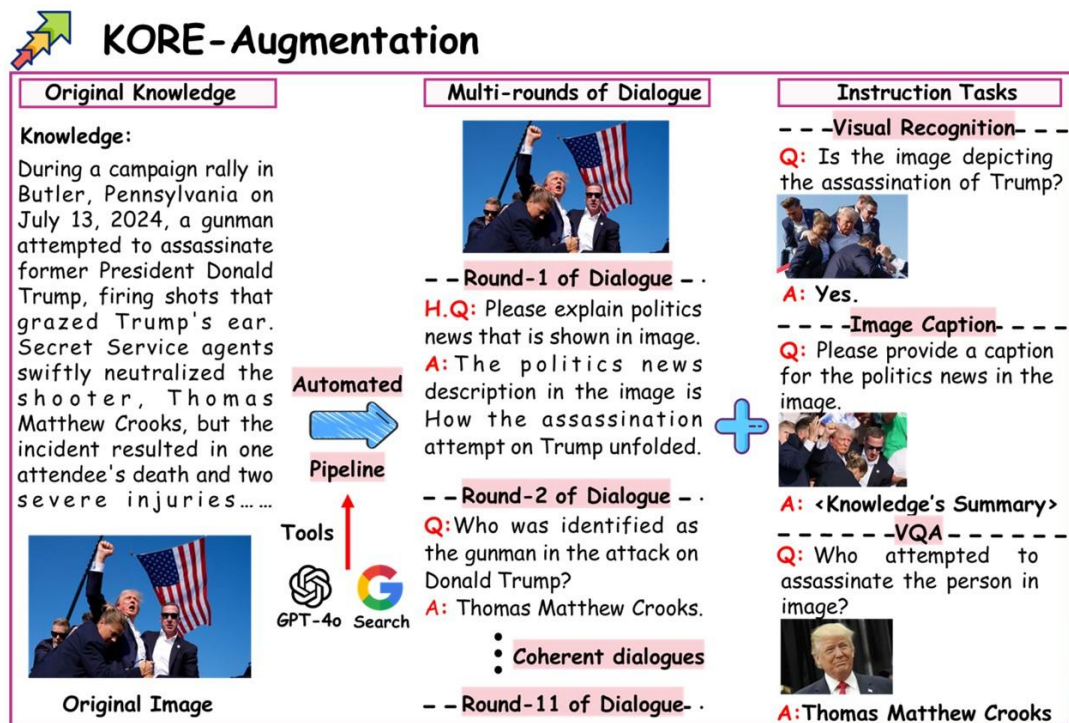
### Powerful Retention

: A ✓



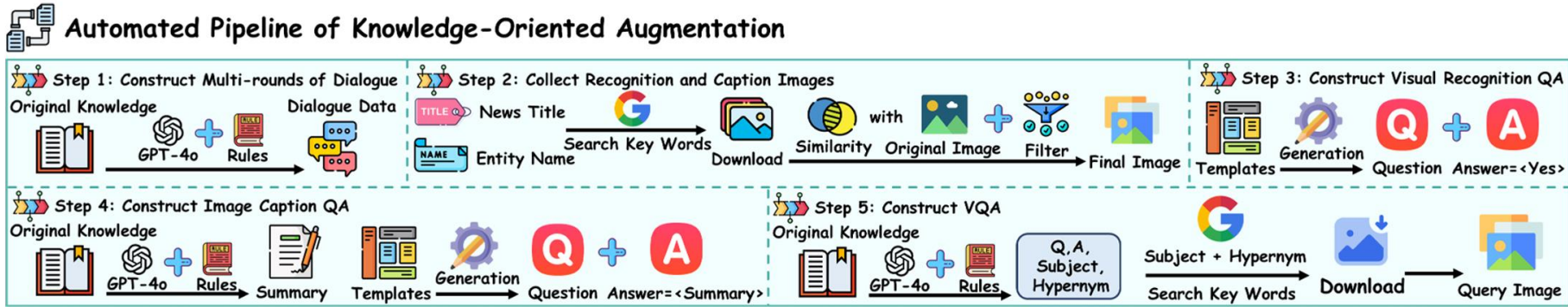
# KORE: KnOWledge-oRientEd Augmentations and Constraints

KORE-Augmentation automatically converts each piece of knowledge into profound and structured knowledge.



KORE-Constraint minimizes interference with previous knowledge by initializing a adapter with null space that stores covariance matrix of previous knowledge.

# KORE-Augmentation



**Step 1:**  
Constructing Multi-rounds of Dialogue.

**Step 2:**  
Collecting Recognition and Caption Images.

**Step 3:**  
Constructing Visual Recognition QA.

**Step 4:**  
Constructing Image Caption QA.

**Step 5:**  
Constructing VQA.

## Knowledge-Oriented Augmentation

### Multi-rounds of Dialogue: Display complete knowledge

----- Round-1 of dialogue -----  
**H.Q:** Please explain politics news that is shown in image.  
**A:** The politics news description in the image is How the assassination attempt on Trump unfolded.

----- Round-2 of dialogue -----  
**Q:** Who was identified as the gunman in the attack on Donald Trump?  
**A:** Thomas Matthew Crooks.

----- Round-3 of dialogue -----  
**Q:** When and where did the attempted assassination of former President Donald Trump occur?  
**A:** ..... occurred on July 13, 2024, in Butler, Pennsylvania.


⋮ Coherent dialogues

----- Round-11 of dialogue -----  
**Q:** What is being investigated as a result of the attack on Donald Trump?  
**A:** The security breach is being investigated as a result of the attack on Donald Trump.

Type: politics

Title: How the assassination attempt on Trump unfolded

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.....



Original Image

### Instruction Tasks: Strengthen knowledge review

Key Words → Download → Similarity → Image

**Part1: Visual Recognition**  
**Q:** Is the image depicting the assassination of Trump?  
**A:** Yes.

**Knowledge Summarize**  
**Q:** Please provide a description for the politics news in the image. **A:** <summary>

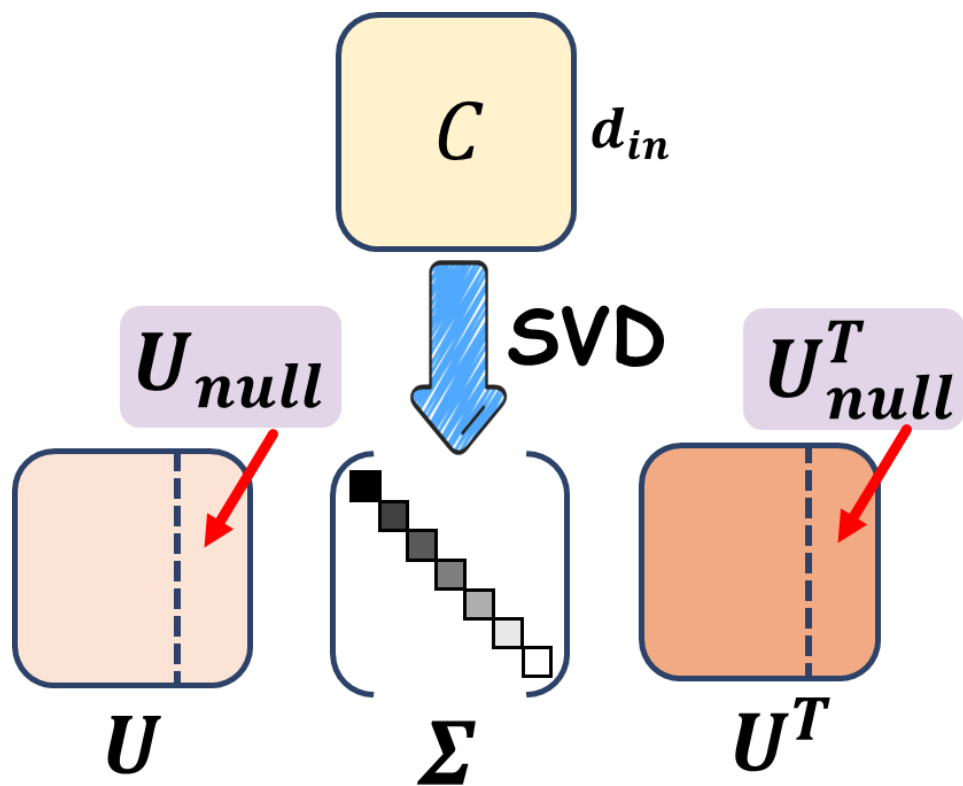
**Knowledge Generation**  
**Q, A, Subject, Hypernym** → Search → Subject Hypernym → Image

**Part3: VQA**  
**Q:** Who attempted to assassinate the person in the image during a campaign rally in July 2024? **A:** Thomas Matthew Crooks

Figure 13: **Overview of construction pipeline for KORE-74K.** The entire data construction process is automated, with only the question templates being manually crafted.

# KORE-Constraint

Capture knowledge into covariance matrix and decompose it



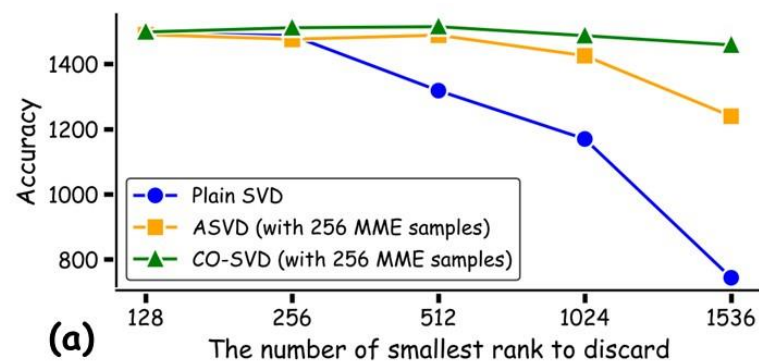
Covariance matrix

$$C = \mathbf{X} \mathbf{X}^T \in \mathbb{R}^{d_{in} \times d_{in}}$$

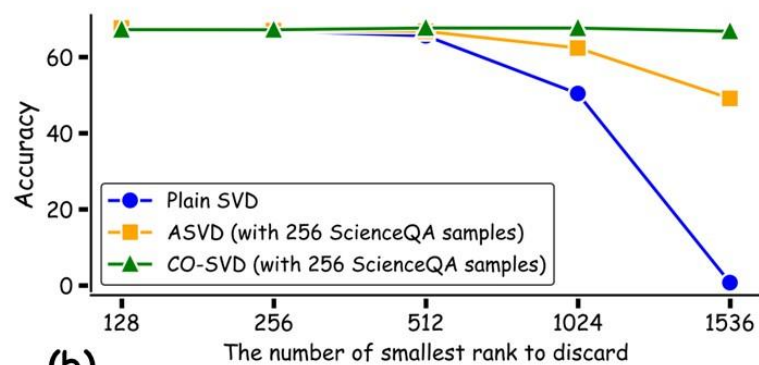
Decompose covariance matrix

$$\text{SVD} (\mathbf{X} (\mathbf{X})^T) = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{u}_i^T$$

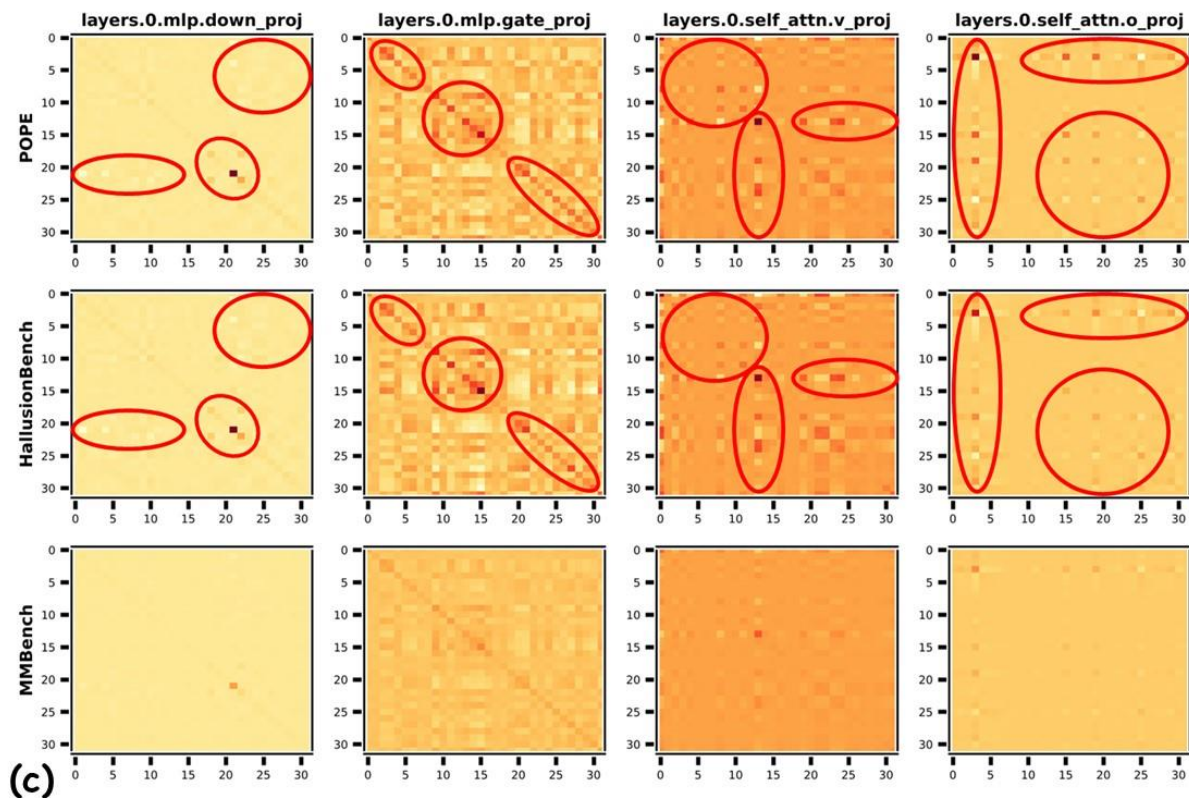
# KORE-Constraint



(a)



(b)



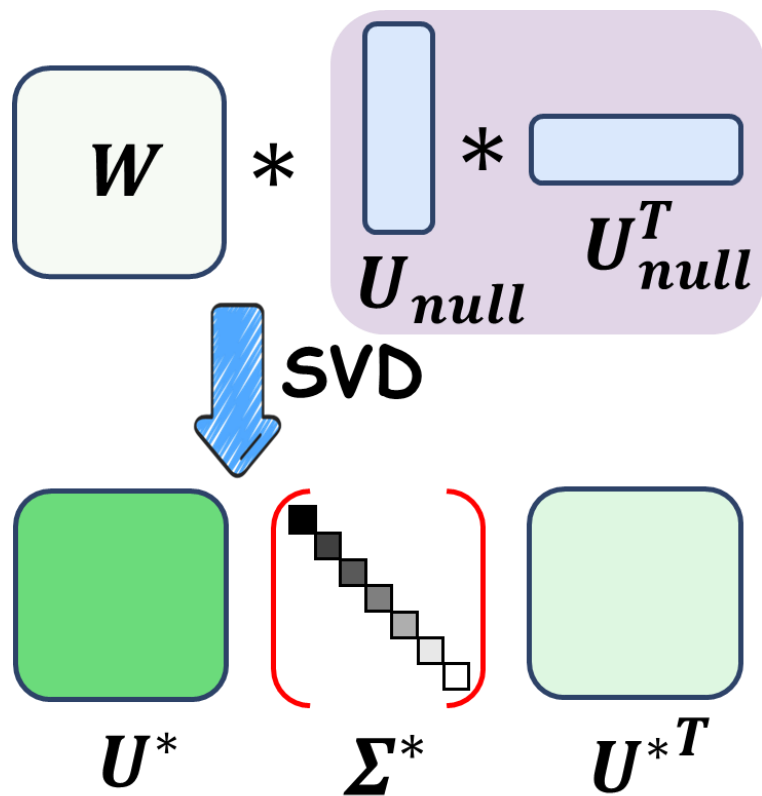
**Findings 1:** Multimodal knowledge can be effectively captured and stored in covariance matrix.

**Findings 2:** Distinct tasks exhibit different outlier distributions in the covariance matrix.



# KORE-Constraint

Define the fine-tuning direction that minimizes interference with previous knowledge



Satisfy null space

$$U_{null}^T C = \mathbf{0}$$

Approximate null space

$$\hat{U} \in \mathbb{R}^{d_{in} \times r} \quad P = \hat{U}\hat{U}^T$$

Define fine-tuning direction

$$\text{SVD}(W_0 P) = \{U^*, \Lambda^*, (U^*)^T\}$$

$$B = U^* \sqrt{\Sigma^*}, \quad A = \sqrt{\Sigma^*} V^{*T}$$

# Main results

Method	#Params	EVOKE		COM $\uparrow$	OCR $\uparrow$	M-DIS $\uparrow$	INS $\uparrow$	M-IDU $\uparrow$	MAT $\uparrow$	HAL $\uparrow$	Avg $\uparrow$
		CEM $\uparrow$	F1 $\uparrow$								
LLaVA-v1.5 (7B)	—	—	—	65.61	45.59	49.22	66.33	26.37	19.33	54.32	—
Full-FT	6,759M	<u>18.02</u>	15.17	43.55	21.55	45.67	25.25	13.03	18.32	16.09	23.23
LoRA	340M	15.23	18.31	48.96	27.01	43.79	29.66	13.70	18.02	41.38	24.28
Replay	340M	11.36	17.98	59.72	37.98	<u>48.64</u>	<b>62.33</b>	<b>19.31</b>	19.17	<u>51.67</u>	<u>28.68</u>
EWC	340M	15.49	19.42	49.42	32.88	45.46	29.79	13.36	18.00	43.50	25.33
LwF	340M	14.58	<u>19.99</u>	53.14	28.77	43.41	36.19	13.68	18.22	44.18	25.61
MoELoRA	340M	6.45	12.20	<u>60.79</u>	38.79	48.27	35.03	<u>17.85</u>	<u>19.79</u>	49.99	23.98
O-LoRA	340M	6.44	12.08	<b>61.47</b>	<u>40.91</u>	48.07	34.85	17.28	<b>19.87</b>	51.12	24.17
SEFE	340M	13.38	16.88	42.06	20.43	40.17	17.73	13.25	18.20	39.30	22.54
<b>KORE (r=235)</b>	340M	<b>30.65</b>	<b>41.26</b>	52.41	<b>40.98</b>	<b>48.68</b>	<u>38.54</u>	16.58	18.59	<b>51.75</b>	<b>37.09</b>
<b>KORE (r=256)</b>	369M	31.05	41.32	52.48	39.96	48.96	60.02	23.18	18.09	51.50	39.11

**Obs 1:** KORE enables accurate adaptation for effectively injecting new knowledge.

**Obs 2:** KORE enables powerful retention for effectively preserving old knowledge.

**Obs 3:** KORE achieves remarkable holistic performance by harmonizing the dual objectives of knowledge injection.

# Knowledge adaptation and retention's Detailed Results

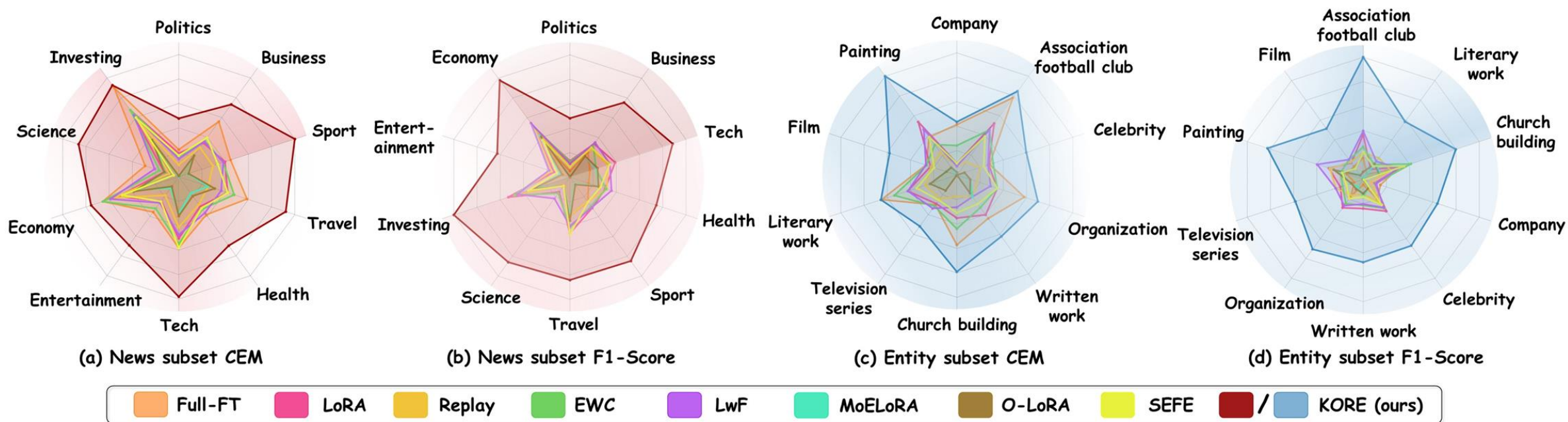


Figure 5: Comparison between KORE and baseline methods on fine-grained knowledge types.

**Obs 4:** KORE demonstrates superior performance across a wide spectrum of fine-grained knowledge.

# Knowledge adaptation and retention's Detailed Results

Table 2: Performance comparison between KORE and baseline methods on fine-grained knowledge retention evaluations with LLaVA-v1.5 (7B).  $MM^B$ : MMBench;  $SEED^{B2P}$ : SEEDBench2\_Plus;  $Math^T$ : MathVista ;  $Math^I$ : MathVision;  $Hall^B$ : HallusionBench. The score of MME is normalized.

Method	COM		OCR		M-DIS		INS	M-IDU	MAT		HAL		Avg
	MME $\uparrow$	$MM^B \uparrow$	$SEED^{B2P} \uparrow$	$OCR^{VQA} \uparrow$	SQA $\uparrow$	MMM U $\uparrow$	$MIA^B \uparrow$	MMDU $\uparrow$	$Math^T \uparrow$	$Math^I \uparrow$	POPE $\uparrow$	$Hall^B \uparrow$	
LLaVA-v1.5 (7B)	66.63	64.60	38.78	52.41	69.83	28.60	66.33	26.37	25.50	13.16	86.87	21.76	46.74
Full-FT	34.17	52.92	31.44	11.65	67.13	24.20	25.25	13.03	24.70	11.94	74.22	9.27	31.66
LoRA	44.06	53.87	30.22	23.80	66.18	21.40	29.66	13.70	23.20	<u>12.83</u>	73.97	8.78	33.47
Replay	<u>58.96</u>	60.48	<b>38.34</b>	37.73	68.77	28.50	<b>62.33</b>	<b>19.31</b>	25.20	<b>13.13</b>	<b>85.44</b>	17.90	<b>43.00</b>
EWC	48.57	50.26	33.60	32.16	65.71	25.20	29.79	13.36	23.30	12.76	76.22	10.77	35.14
LwF	50.87	55.41	32.02	25.52	66.21	20.60	36.19	13.68	24.40	12.04	79.23	9.13	35.44
MoELoRA	58.26	<b>63.32</b>	37.42	40.17	<b>69.04</b>	27.50	35.03	<u>17.85</u>	<u>27.80</u>	11.78	80.70	19.29	40.51
O-LoRA	<b>60.30</b>	<u>62.63</u>	<u>37.90</u>	<u>43.91</u>	<u>68.84</u>	27.30	34.85	17.28	<b>28.20</b>	11.55	<u>81.46</u>	<u>20.78</u>	<u>41.25</u>
SEFE	36.10	48.02	22.79	18.07	65.03	15.30	17.73	13.25	26.00	10.39	72.81	5.79	29.27
<b>KORE (r=235)</b>	49.84	54.98	37.73	<b>44.24</b>	68.06	<b>29.30</b>	<u>38.54</u>	16.58	25.10	12.09	80.99	<b>22.51</b>	40.00
<b>KORE (r=256)</b>	50.06	54.90	36.89	43.03	68.51	29.40	60.02	23.18	24.70	11.48	80.77	22.23	42.10

**Obs 5: KORE achieves competitive knowledge retention.**

# Knowledge adaptation and retention's Detailed Results

Table 3: Performance of knowledge adaptation (K.A) and retention (K.R) under specific knowledge-oriented constraints.

Method	K.A $\uparrow$	K.R $\uparrow$	Avg $\uparrow$
KORE	<b>35.96</b>	38.22	37.09
KORE <sub>MME</sub>	34.46	<b>43.16</b>	<u>38.81</u>
KORE <sub>OCR<sup>VQA</sup></sub>	34.85	42.21	38.53
KORE <sub>Math<sup>T</sup></sub>	<u>35.20</u>	<u>42.87</u>	<b>39.03</b>
KORE <sub>Hall<sup>B</sup></sub>	34.96	42.09	38.52

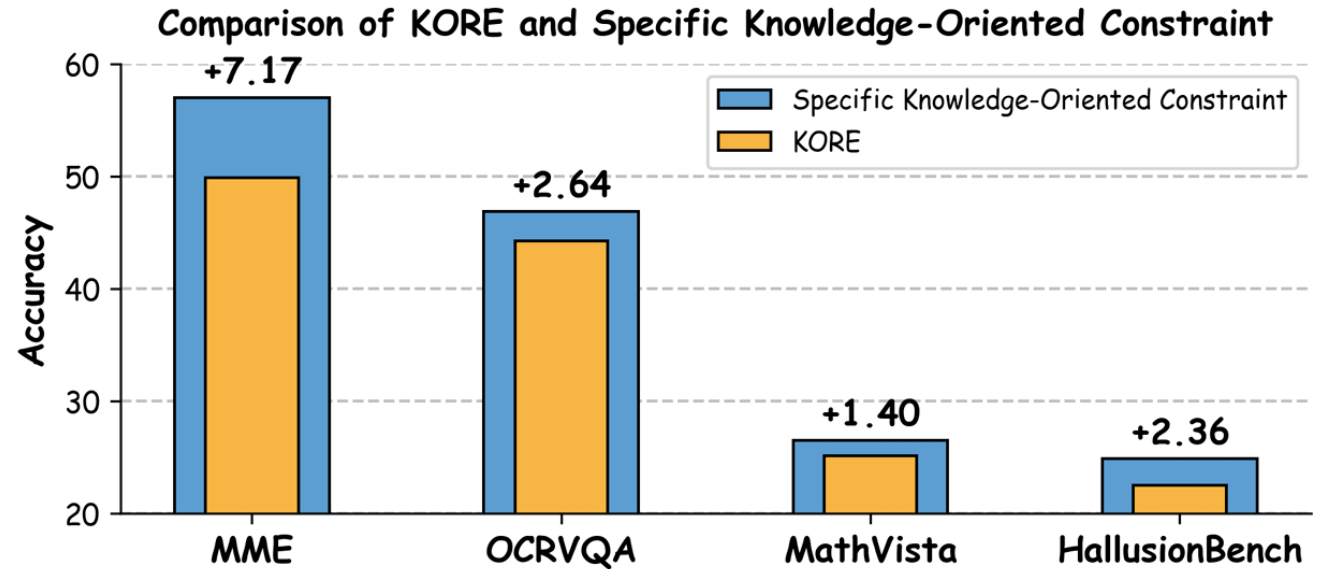


Figure 6: Performance comparison of corresponding tasks under specific knowledge-oriented constraints.

**Obs 6:** Specific constraints enhance knowledge retention and overall performance.

# Various LMM scales and architectures

Table 4: Performance comparison between KORE and baseline methods on knowledge adaptation and retention with various LMMs scales and architectures.

Methods	EVOKE		COM ↑	OCR ↑	M-DIS ↑	INS ↑	M-IDU ↑	MAT ↑	HAL ↑	Avg ↑
	CEM ↑	F1 ↑								
<i>LLaVA-v1.5 (13B)</i>										
Vanilla	—	—	66.86	51.12	52.70	66.04	33.93	19.64	56.77	—
LoRA	<u>16.26</u>	<u>22.83</u>	<u>60.57</u>	32.58	43.72	23.26	17.43	15.82	38.08	25.21
Replay	12.05	20.21	<b>65.81</b>	<b>47.51</b>	<u>48.42</u>	<u>61.04</u>	<u>24.62</u>	<u>19.55</u>	<b>54.16</b>	<u>30.70</u>
<b>KORE</b>	<b>32.89</b>	<b>44.47</b>	59.35	<u>45.96</u>	<b>51.39</b>	<b>65.10</b>	<b>26.84</b>	<b>20.31</b>	<u>40.52</u>	<b>41.44</b>
<i>Qwen2.5-VL (7B)</i>										
Vanilla	—	—	81.18	70.32	65.35	78.46	61.25	47.69	66.96	—
LoRA	<u>14.56</u>	14.01	52.54	64.54	22.35	21.39	23.25	13.52	41.38	24.21
Replay	11.73	<u>18.51</u>	<b>78.54</b>	<b>69.17</b>	65.26	70.20	<b>50.72</b>	42.74	<b>67.48</b>	<u>39.28</u>
<b>KORE</b>	<b>22.91</b>	<b>31.36</b>	<u>56.60</u>	<u>67.74</u>	<b>65.48</b>	<b>70.51</b>	<u>45.02</u>	<b>43.72</b>	<u>58.57</u>	<b>42.68</b>

**Obs 7:** KORE shows enhanced superiority on a larger-scale LMM.

**Obs 8:** KORE's effectiveness is not architecture-specific.

# Ablation experiments

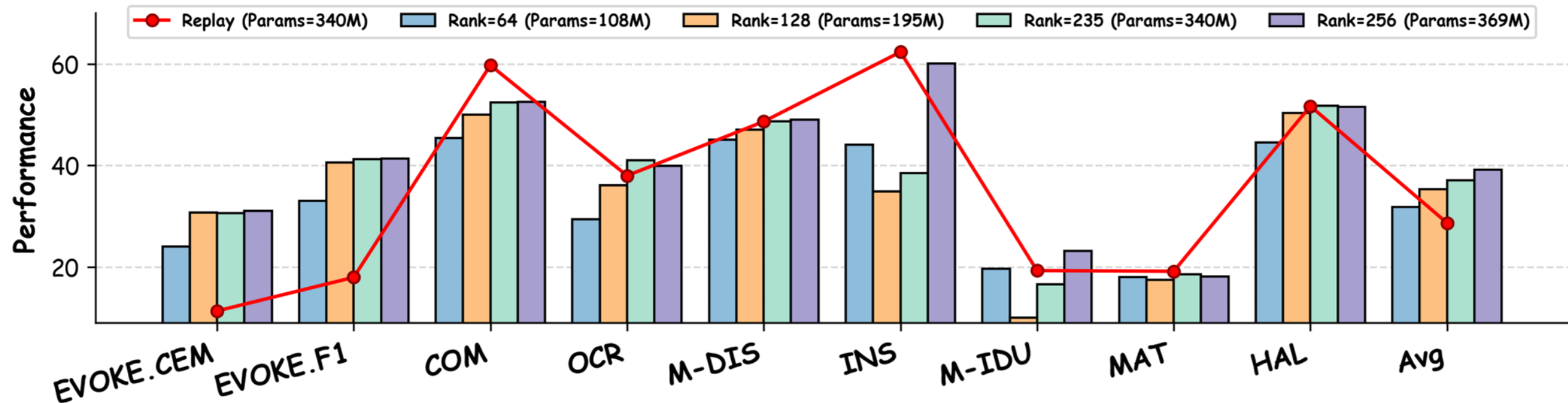


Figure 7: Comparison of different ranks for KORE with LLaVA-v1.5 (7B).

**Obs 9:** Larger rank enhance KORE's performance.

# Ablation experiments

Table 5: Comparison of ablation experiment results of KORE on LLaVA-v1.5 (7B).

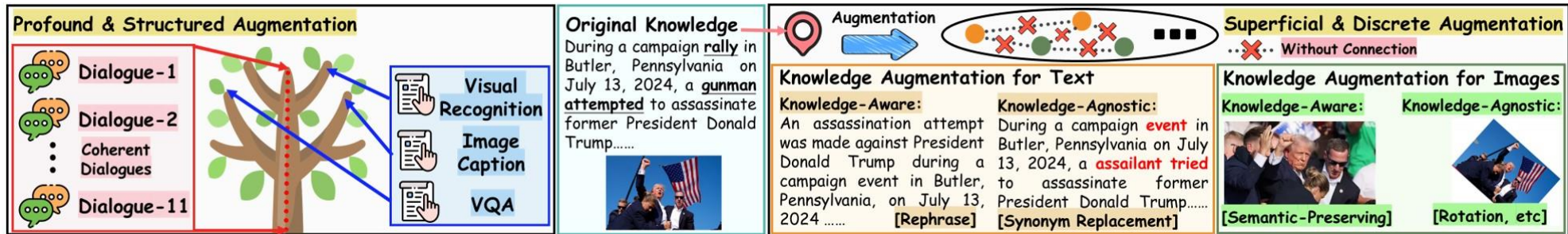
Setting	EVOKE		COM $\uparrow$	OCR $\uparrow$	M-DIS $\uparrow$	INS $\uparrow$	M-IDU $\uparrow$	MAT $\uparrow$	HAL $\uparrow$	Avg $\uparrow$
	CEM $\uparrow$	F1 $\uparrow$								
<b>KORE</b>	30.65	41.26	<u>52.41</u>	<b>40.98</b>	<b>48.68</b>	<b>38.54</b>	<b>16.58</b>	18.59	<b>51.75</b>	<b>37.09</b>
W/o Augmentation	10.83	18.31	<b>59.96</b>	<u>40.42</u>	47.13	32.53	16.00	<b>19.71</b>	49.50	26.23
W/o Constraint	<b>33.93</b>	<b>43.71</b>	46.39	32.38	46.31	32.70	15.38	<u>19.12</u>	46.47	36.46
W/o Frozen Matrix $A$	<u>31.97</u>	<u>41.72</u>	50.73	39.56	<u>48.37</u>	<u>35.30</u>	<u>16.44</u>	19.07	<u>49.91</u>	<u>36.95</u>

**Obs 10:** Ablation studies reveals the effectiveness of KORE's design.



# Comparison with general augmentation methods

Display



Results

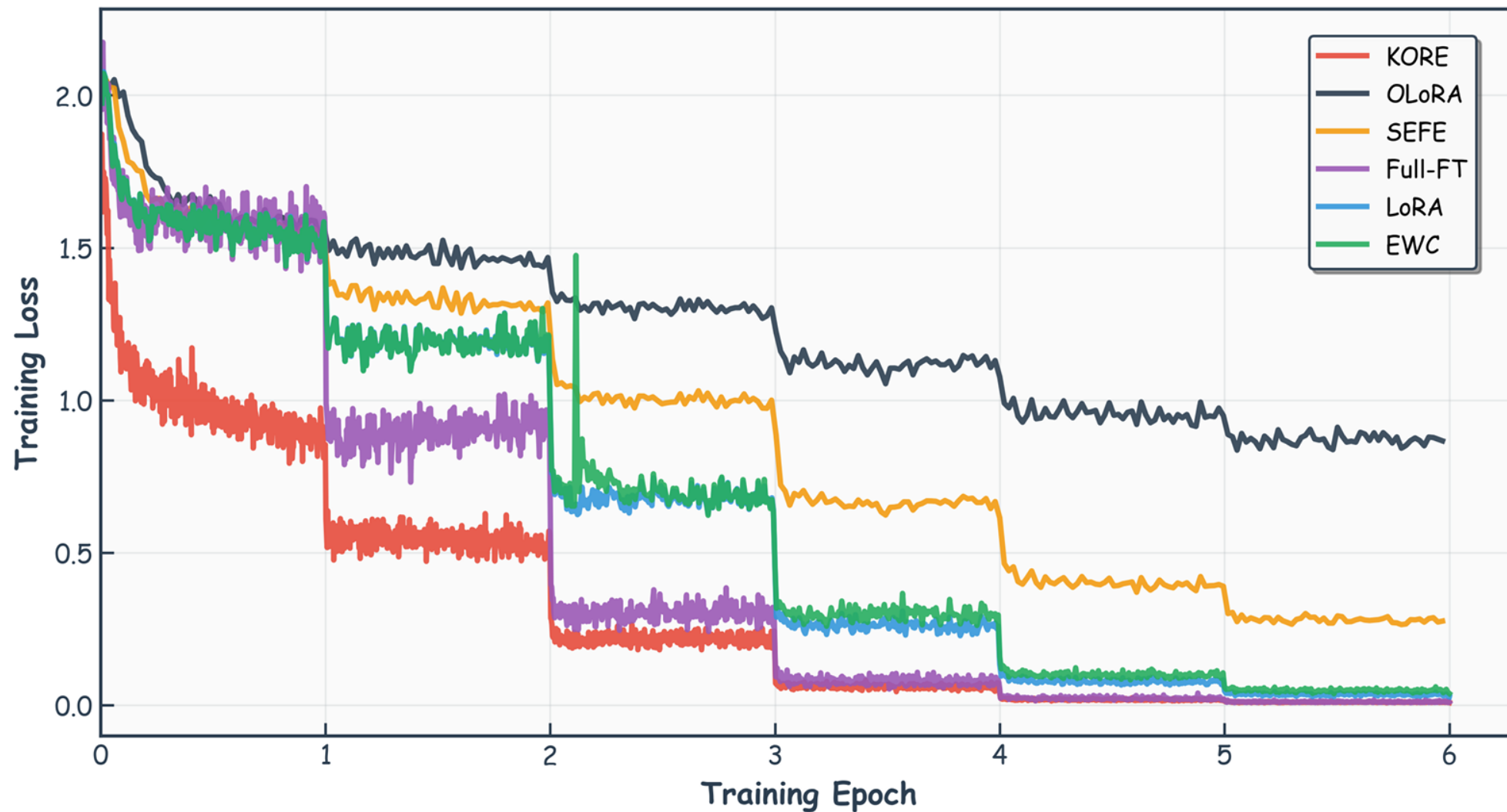
Table 6: Performance comparison of different augmentation methods.

Method	K.A ↑	K.R ↑	Avg ↑
<b>KORE-AUGMENTATION</b>	<b>38.82</b>	<b>35.78</b>	<b>36.46</b>
<i>Augmentation for Text</i>			
Knowledge-Aware	20.29	34.86	27.38
Knowledge-Agnostic	15.60	35.71	25.49
<i>Augmentation for Images</i>			
Knowledge-Aware	18.33	34.02	25.86
Knowledge-Agnostic	18.33	32.09	25.25

**Obs 11:** KORE-Augmentation is superior to general augmentation methods.

# Loss curves

Training Loss Comparison



# Case study

**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.



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



**Answer:** John Hopfield

## LLaVA-v1.5-7B

Full-FT <b>Answer:</b> Alain Aspect CEM: 0.0, F1: 0.0	LoRA <b>Answer:</b> David Wineland CEM: 0.0, F1: 0.0	Replay <b>Answer:</b> John barrett CEM: 0.0, F1: 0.5
EWC <b>Answer:</b> Duncan Haldane CEM: 0.0, F1: 0.0	LwF <b>Answer:</b> Emmanuel Candes CEM: 0.0, F1: 0.0	MoELoRA <b>Answer:</b> Peter higgs CEM: 0.0, F1: 0.0
O-LoRA <b>Answer:</b> Peter higgs CEM: 0.0, F1: 0.0	SEFE <b>Answer:</b> David Wineland CEM: 0.0, F1: 0.0	KORE <b>Answer:</b> John Hopfield CEM: 1.0, F1: 1.0

## LLaVA-v1.5-13B

LoRA <b>Answer:</b> Alain Aspect CEM: 0.0, F1: 0.0	Replay <b>Answer:</b> Alain Aspect CEM: 0.0, F1: 0.0	KORE <b>Answer:</b> John Hopfield CEM: 1.0, F1: 1.0
--	--	---

## Qwen2.5-VL

LoRA <b>Answer:</b> Kip Thorne CEM: 0.0, F1: 0.0	Replay <b>Answer:</b> Kip Thorne CEM: 0.0, F1: 0.0	KORE <b>Answer:</b> John Hopfield CEM: 1.0, F1: 1.0
--	--	---

**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). The vehicle is named after the tourbillon mechanism, a balancing structure used in a variety of mechanical watches.



**Question:** What is the production limit of the automobile model in the image?



**Answer:** 250 units

## LLaVA-v1.5-7B

Full-FT <b>Answer:</b> 20 CEM: 0.0, F1: 0.0	LoRA <b>Answer:</b> 120 CEM: 0.0, F1: 0.0	Replay <b>Answer:</b> 150 CEM: 0.0, F1: 0.5
EWC <b>Answer:</b> 120 CEM: 0.0, F1: 0.0	LwF <b>Answer:</b> 12 CEM: 0.0, F1: 0.0	MoELoRA <b>Answer:</b> 100 CEM: 0.0, F1: 0.0
O-LoRA <b>Answer:</b> 40 CEM: 0.0, F1: 0.0	SEFE <b>Answer:</b> Bugatti Bolide CEM: 0.0, F1: 0.0	KORE <b>Answer:</b> 250 CEM: 0.0, F1: 0.67

## LLaVA-v1.5-13B

LoRA <b>Answer:</b> 400 CEM: 0.0, F1: 0.0	Replay <b>Answer:</b> 200 CEM: 0.0, F1: 0.0	KORE <b>Answer:</b> 250 units CEM: 1.0, F1: 1.0
---	---	---

## Qwen2.5-VL

LoRA <b>Answer:</b> 150 units CEM: 0.0, F1: 0.5	Replay <b>Answer:</b> 99 CEM: 0.0, F1: 0.0	KORE <b>Answer:</b> 250 units CEM: 1.0, F1: 1.0
---	--	---