KORE

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

Kailin Jiang [1,2] Hongbo Jiang [3] Ning Jiang[4] Zhi Gao[5,2] Jinhe Bi[6]
Yuchen Ren[7] Bin Li[1] Yuntao Du[8] Lei Liu[1] Qing Li[2]

[1]University of Science and Technology of China, USTC [2]State Key Laboratory of General Artificial Intelligence, BIGAI

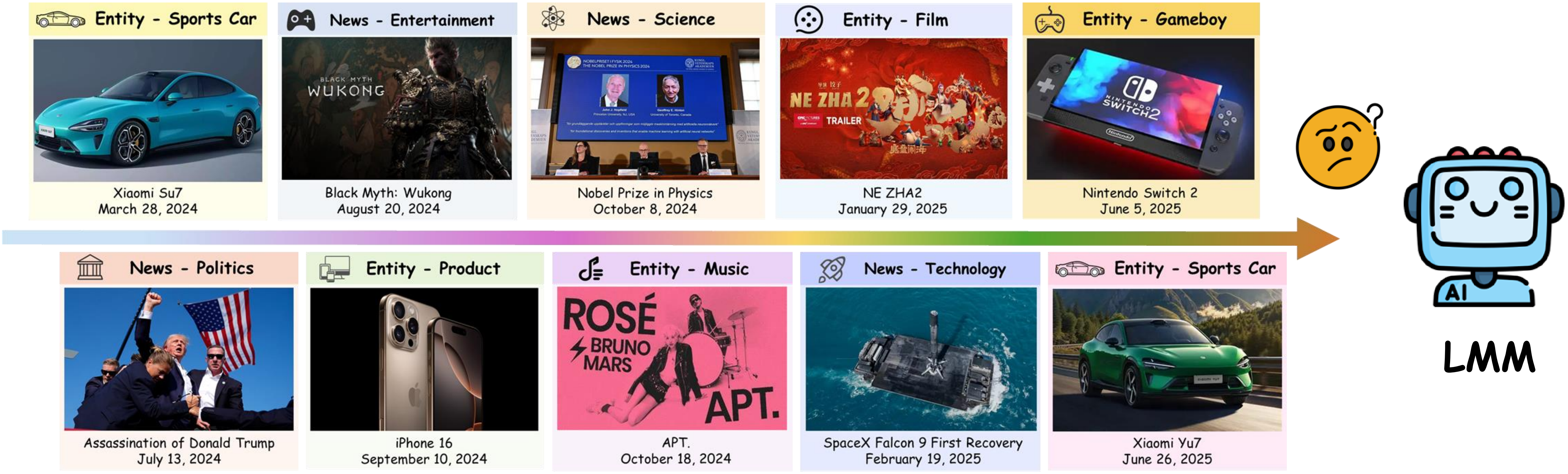
[3]Xiamen University, XMU [4] Northeast Forestry University, NEFU [5]Beijing Institute of Technology, BIT

[6]Ludwig Maximilian University of Munich, LMU [7]The University of Sydney, USYD [8]Shandong University, SDU

Machine Learning Lab , BIGAI

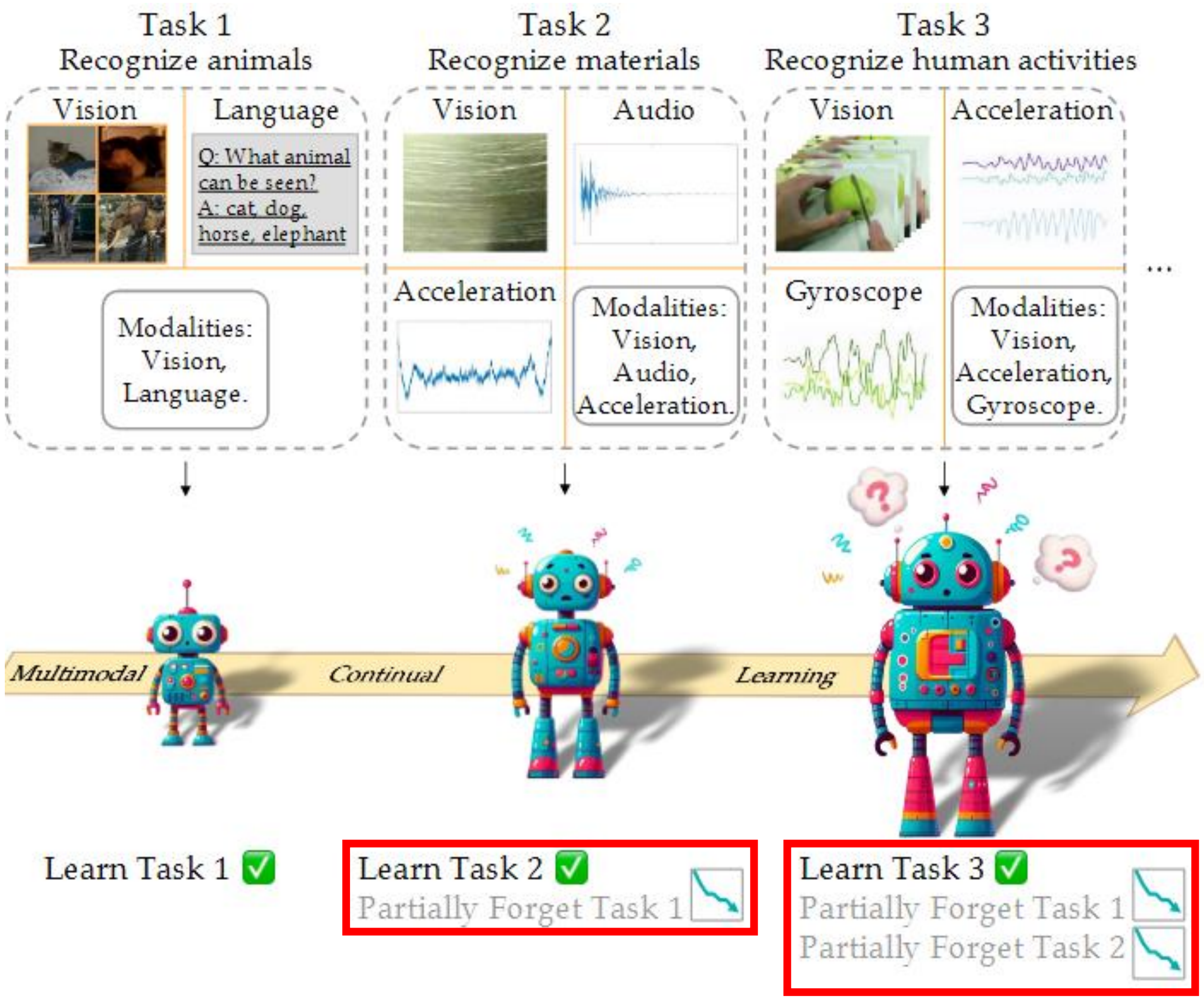
Background: Knowledge Adaptation

Released LMMs can't keep pace with evolving knowledge.

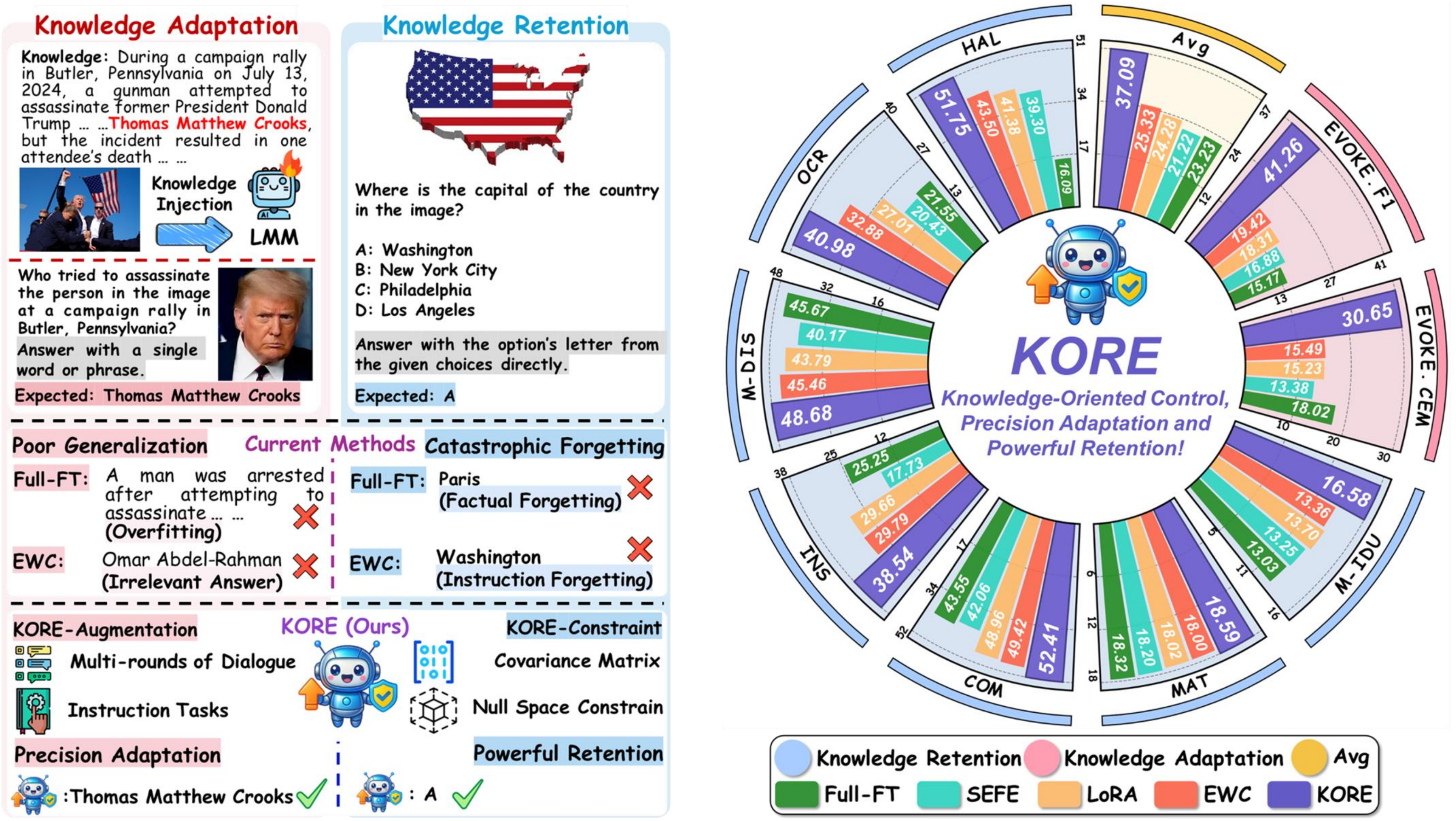


Background: Knowledge Retention

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

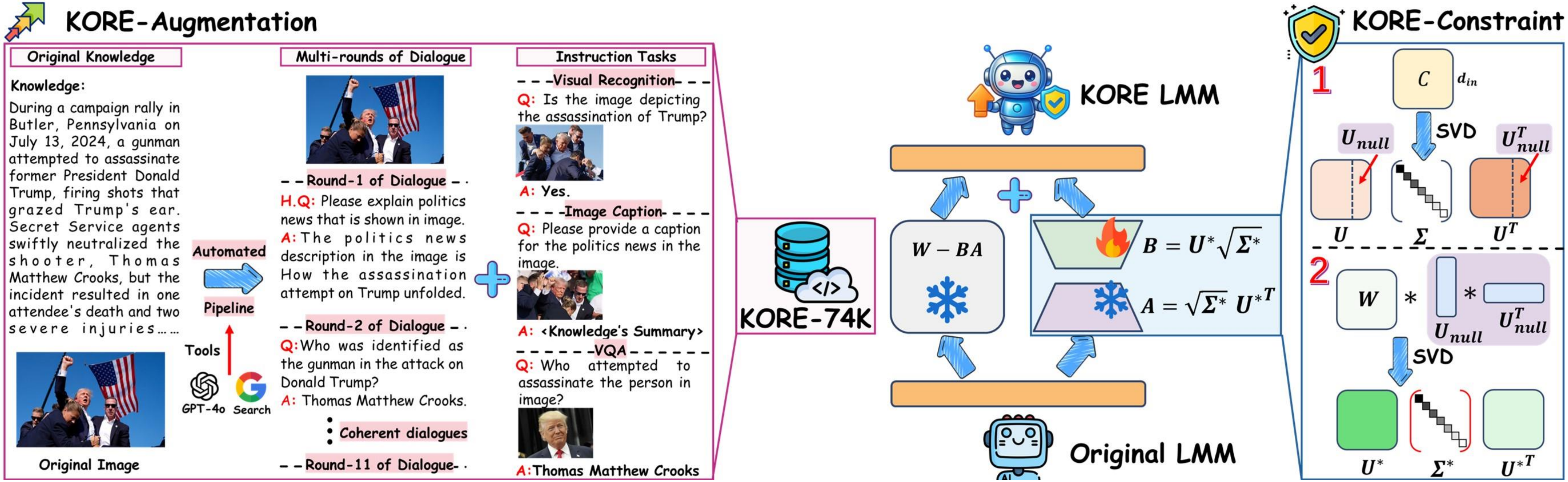


Teaser: Accurate Adaptation and Powerful Retention



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 ana dapter with null space that stores covariance matrix of previous knowledge.

KORE-Augmentation

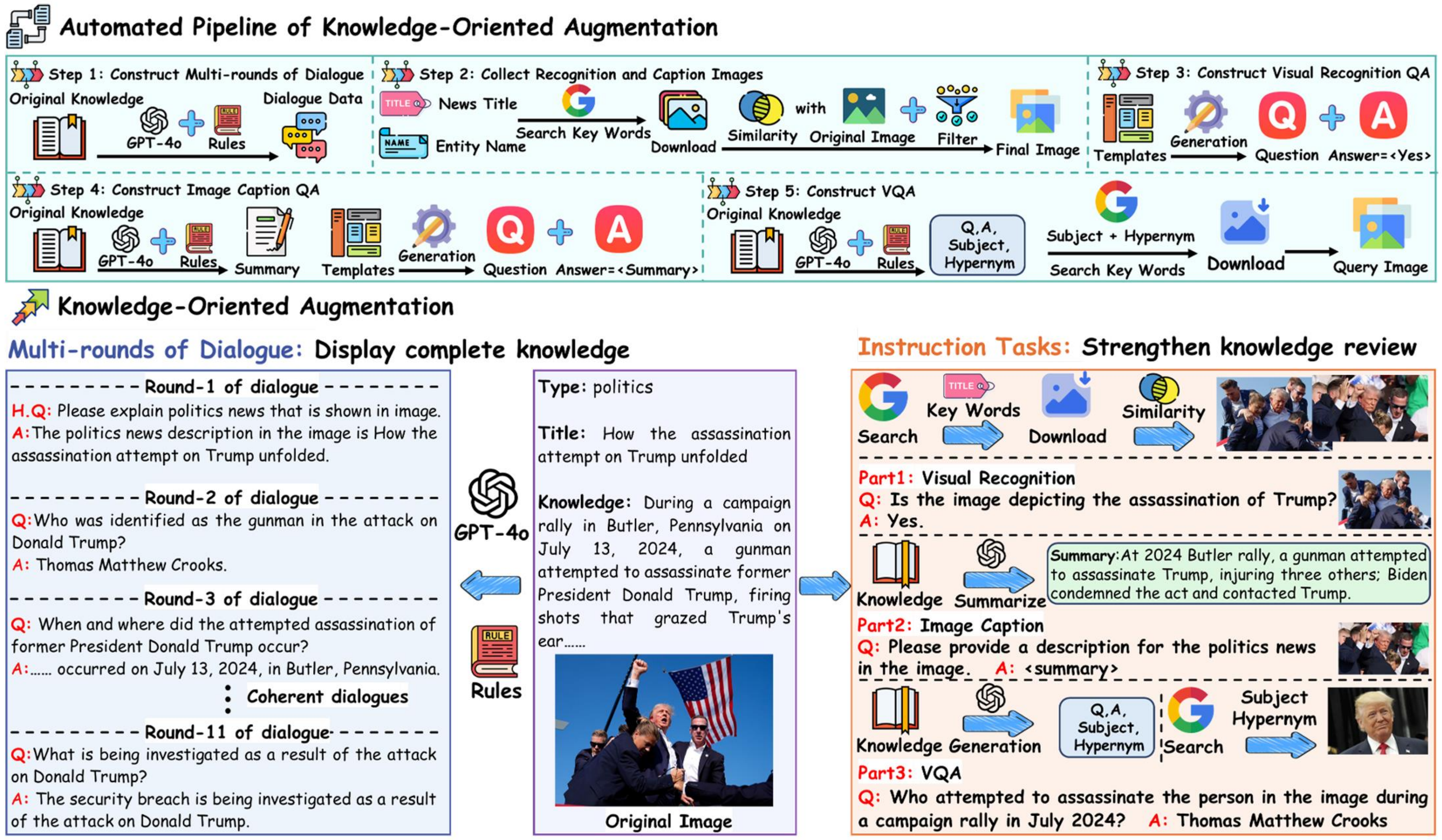


Figure 13: **Overview of construction pipeline for KORE-74K.** The entire data construction process is automated, with only the question templates being manually crafted.

Step 1: Constructing Multi-rounds of Dialogue.

Step 2: Collecting Recognition and Caption Images.

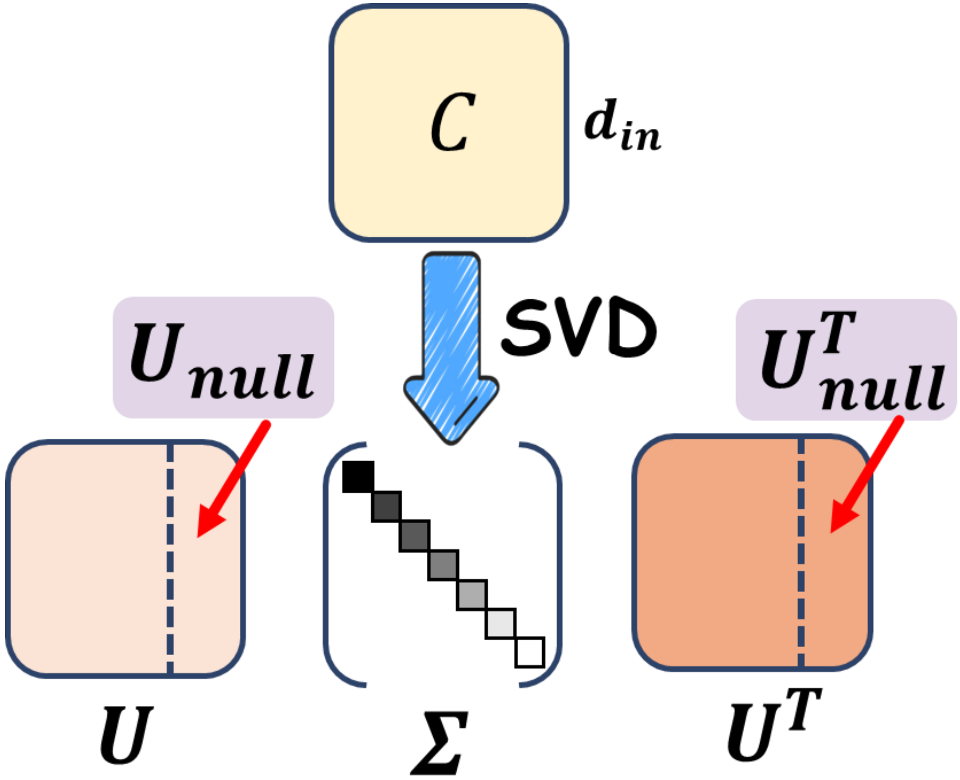
Step 3: Constructing Visual Recognition QA.

Step 4: Constructing Image Caption QA.

Step 5: Constructing VQA.

KORE-Constraint

Capture knowledge into covariance matrix and decompose it



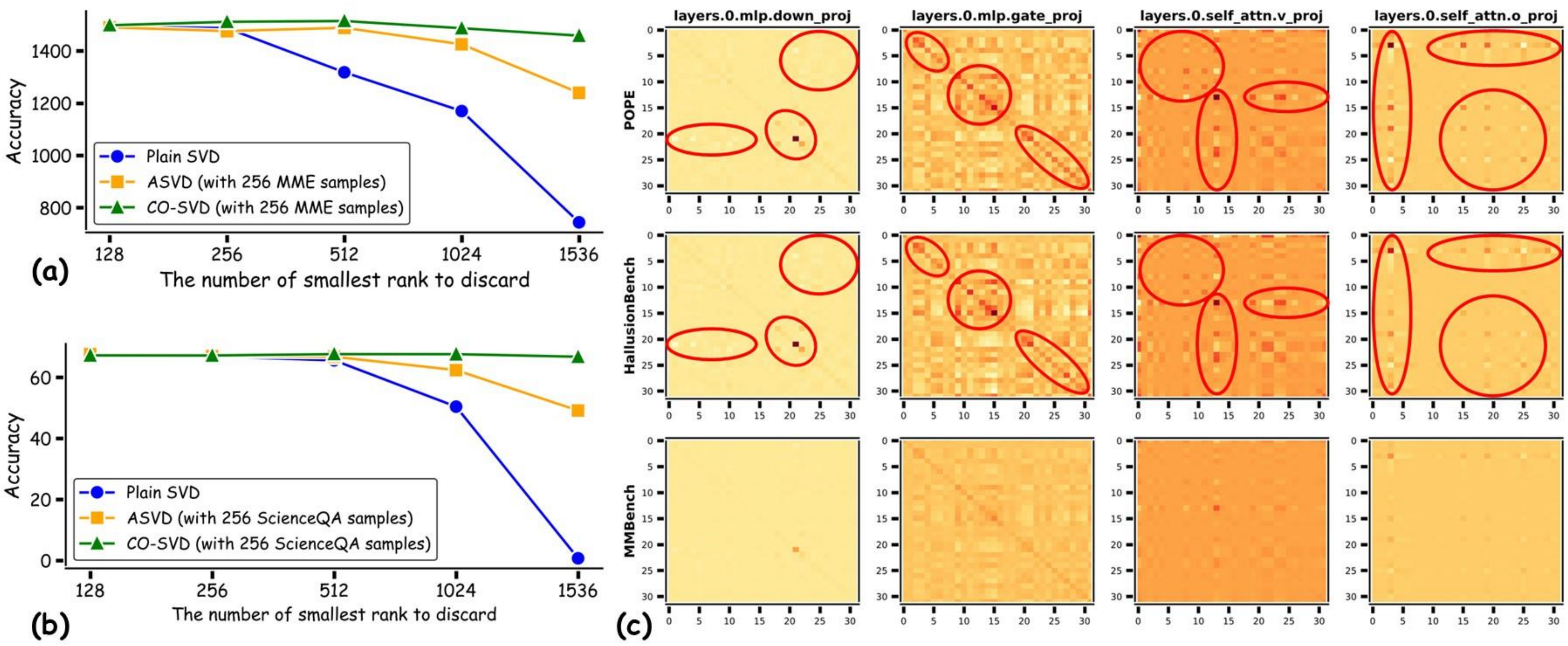
Covariance matrix

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

Decompose covariance matrix

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

KORE-Constraint

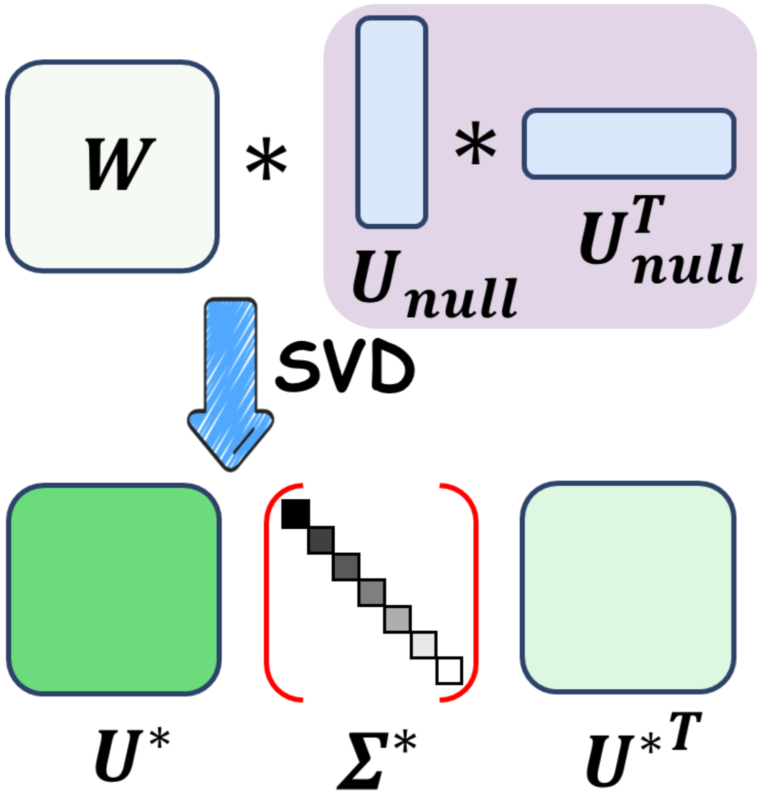


Findings 1: Multimodal knowledge can be effectively captured and stored in covariance matrix.

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

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

Approximate null space

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

Define fine-tuning direction

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

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

Main results

Method	#Params	EVOKE		COM ↑	OCR ↑	M-DIS ↑	INS ↑	M-IDU ↑	MAT ↑	HAL ↑	Avg ↑
		CEM ↑	F1↑								
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>	**62.33**	**19.31**	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	**61.47**	<u>40.91</u>	48.07	34.85	17.28	**19.87**	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
KORE (r=235)	340M	**30.65**	**41.26**	52.41	**40.98**	**48.68**	<u>38.54</u>	16.58	18.59	**51.75**	**37.09**
KORE (r=256)	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

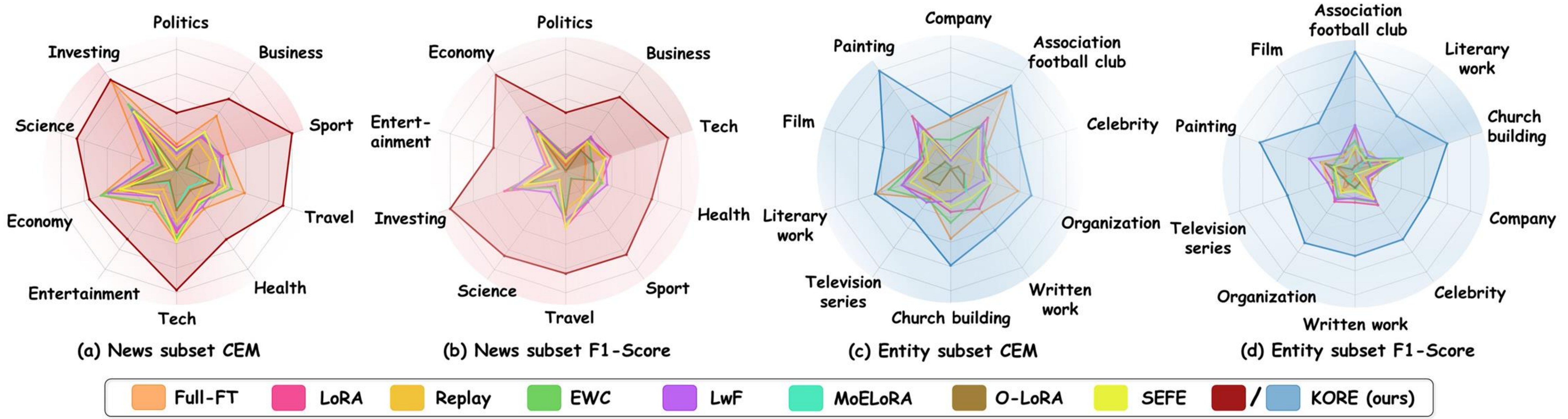


Figure 5: Comparison between KORE and baseline methods on fine-grained knowledge types.

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 ↑	MM^{B} ↑	$SEED^{B2P}$ ↑	OCR^{VQA} ↑	SQA ↑	MMMU ↑	MIA^{B} ↑	MMDU ↑	$Math^{T}$ ↑	$Math^{I}$ ↑	POPE ↑	$Hall^{B}$ ↑	
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	12.83	73.97	8.78	33.47
Replay	58.96	60.48	**38.34**	37.73	68.77	28.50	**62.33**	**19.31**	25.20	**13.13**	**85.44**	17.90	**43.00**
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	**63.32**	37.42	40.17	**69.04**	27.50	35.03	17.85	27.80	11.78	80.70	19.29	40.51
O-LoRA	**60.30**	62.63	37.90	43.91	68.84	27.30	34.85	17.28	**28.20**	11.55	81.46	20.78	41.25
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
KORE (r=235)	49.84	54.98	37.73	**44.24**	68.06	**29.30**	38.54	16.58	25.10	12.09	80.99	**22.51**	40.00
KORE (r=256)	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 ↑	K.R ↑	Avg ↑
KORE	**35.96**	38.22	37.09
KORE_{MME}	34.46	**43.16**	<u>38.81</u>
$\text{KORE}_{\text{OCRVQA}}$	34.85	42.21	38.53
$\text{KORE}_{\text{MathT}}$	<u>35.20</u>	<u>42.87</u>	**39.03**
$\text{KORE}_{\text{HallB}}$	34.96	42.09	38.52

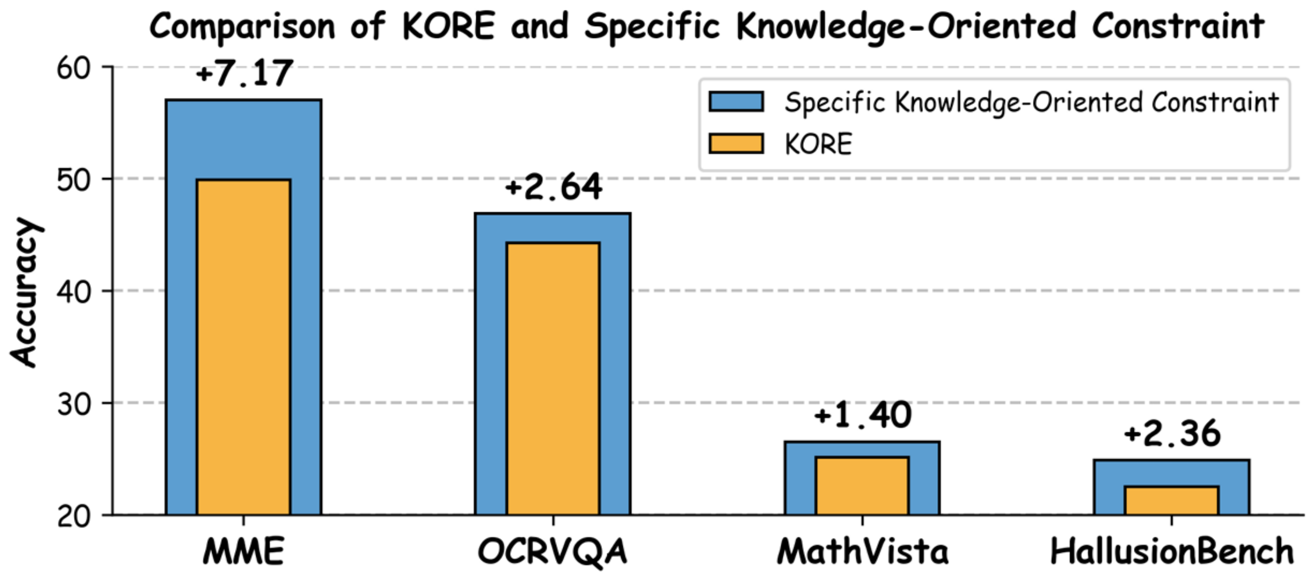


Figure 6: Performance comparison of corresponding tasks under specific knowledge-oriented constraints.

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

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

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

Ablation experiments

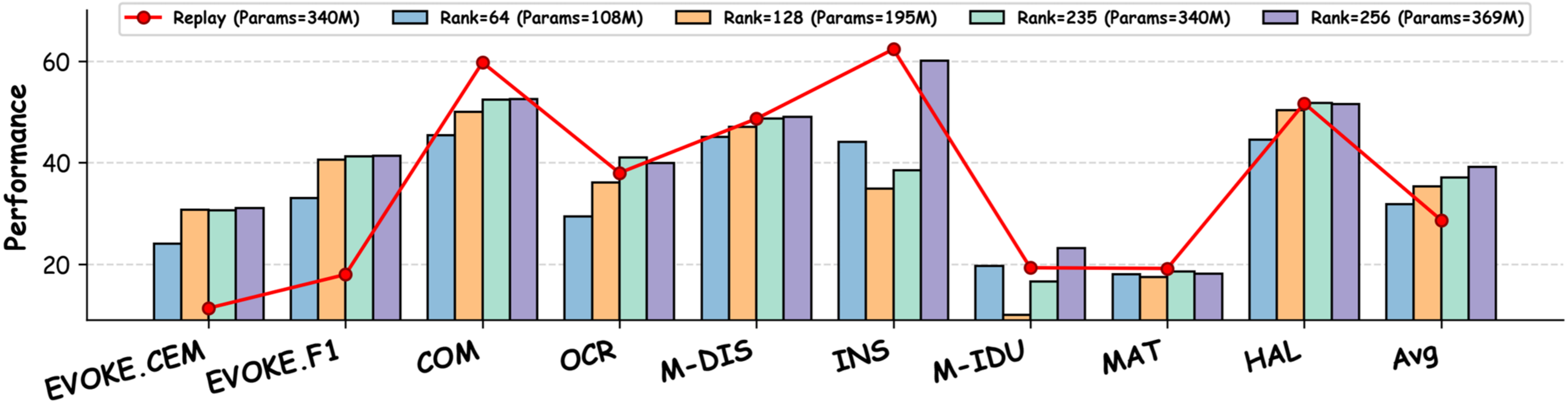


Figure 7: Comparison of different ranks for KORE with LLaVA-v1.5 (7B).

Obs 9: Larger rank enhance KORE's performance.

Ablation experiments

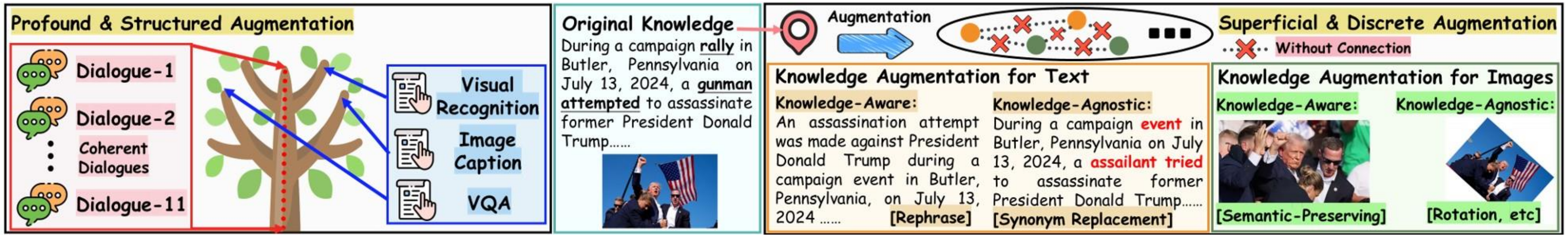
Table 5: Comparison of ablation experiment results of KORE on LLaVA-v1.5 (7B).

Setting	EVOKE		COM ↑	OCR ↑	M-DIS ↑	INS ↑	M-IDU ↑	MAT ↑	HAL ↑	Avg ↑
	CEM ↑	F1↑								
KORE	30.65	41.26	<u>52.41</u>	**40.98**	**48.68**	**38.54**	**16.58**	18.59	**51.75**	**37.09**
W/o Augmentation	10.83	18.31	**59.96**	<u>40.42</u>	47.13	32.53	16.00	**19.71**	49.50	26.23
W/o Constraint	**33.93**	**43.71**	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



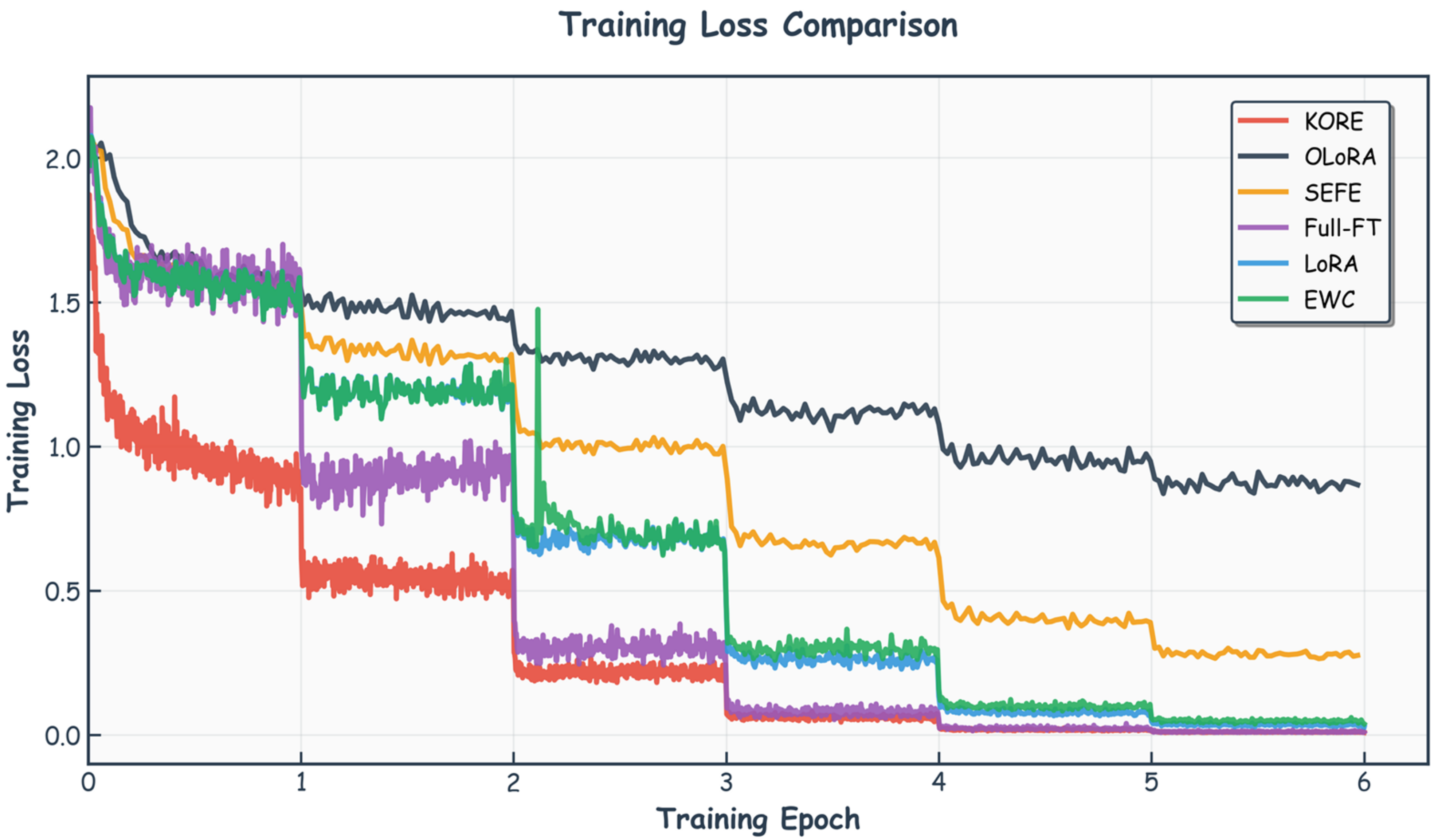
Results

Table 6: Performance comparison of different augmentation methods.

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

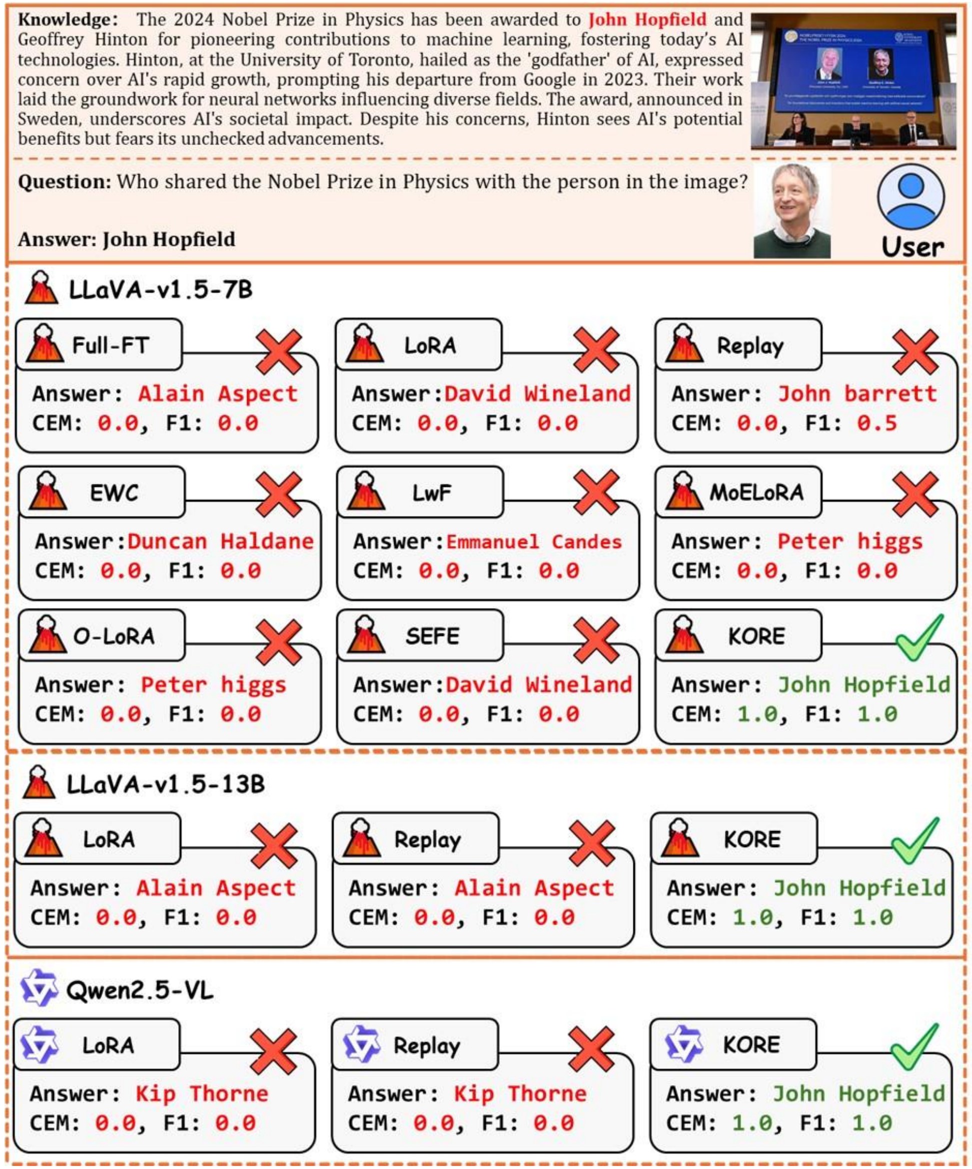
Obs 11: KORE-Augmentation is superior to general augmentation methods.

Loss curves

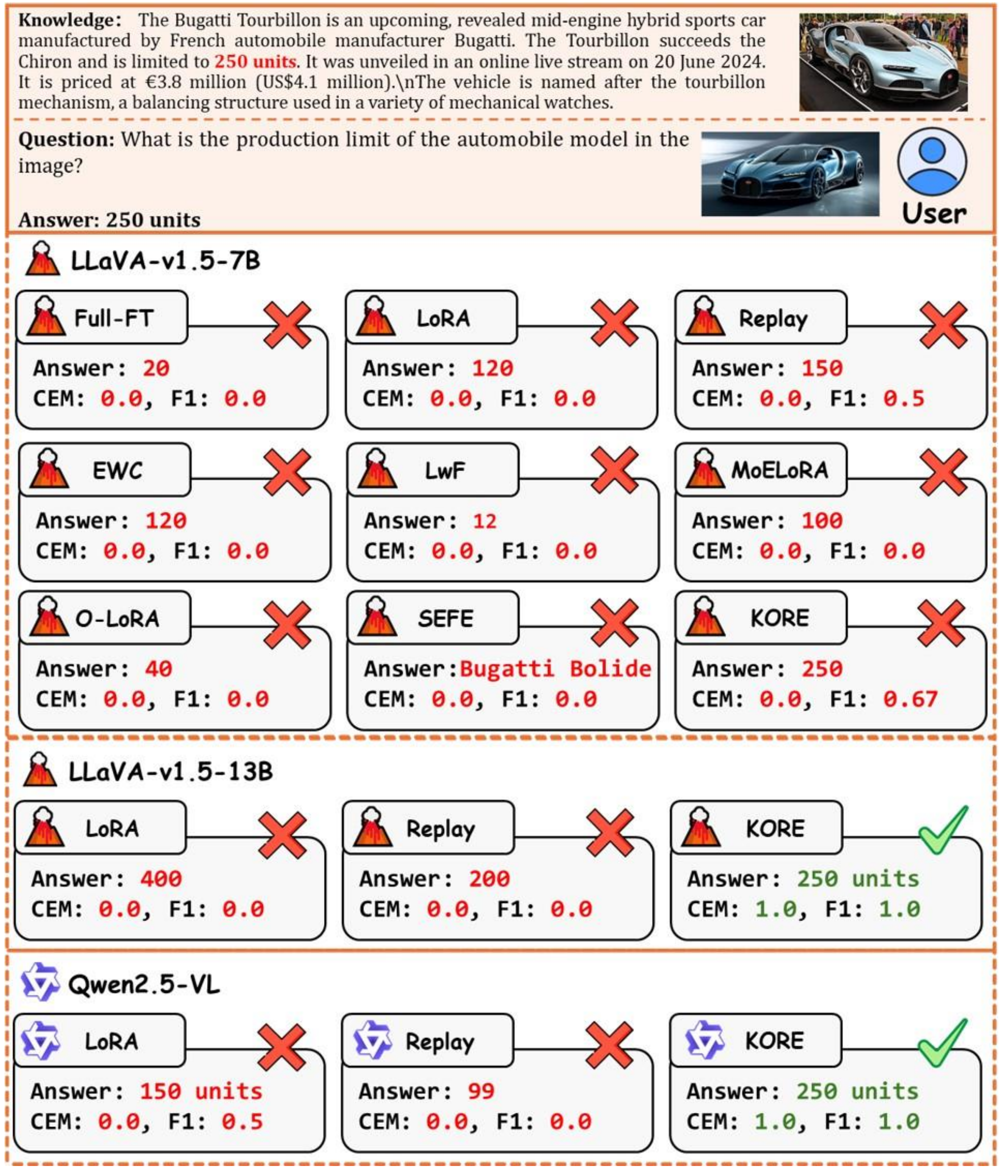


Training Loss Comparison
KORE
OLoRA
SEFE
Full-FT
LoRA
EWC
Training Loss
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
Training Epoch

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
User
LLaVA-v1.5-7B
Full-FT
Answer: Alain Aspect
CEM: 0.0, F1: 0.0
LoRA
Answer:David Wineland
CEM: 0.0, F1: 0.0
Replay
Answer: John barrett
CEM: 0.0, F1: 0.5
EWC
Answer:Duncan Haldane
CEM: 0.0, F1: 0.0
LwF
Answer:Emmanuel Candes
CEM: 0.0, F1: 0.0
MoELoRA
Answer: Peter higgs
CEM: 0.0, F1: 0.0
O-LoRA
Answer: Peter higgs
CEM: 0.0, F1: 0.0
SEFE
Answer:David Wineland
CEM: 0.0, F1: 0.0
KORE
Answer: John Hopfield
CEM: 1.0, F1: 1.0
LLaVA-v1.5-13B
LoRA
Answer: Alain Aspect
CEM: 0.0, F1: 0.0
Replay
Answer: Alain Aspect
CEM: 0.0, F1: 0.0
KORE
Answer: John Hopfield
CEM: 1.0, F1: 1.0
Qwen2.5-VL
LoRA
Answer: Kip Thorne
CEM: 0.0, F1: 0.0
Replay
Answer: Kip Thorne
CEM: 0.0, F1: 0.0
KORE
Answer: 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).\nThe 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
User
LLaVA-v1.5-7B
Full-FT
Answer: 20
CEM: 0.0, F1: 0.0
LoRA
Answer: 120
CEM: 0.0, F1: 0.0
Replay
Answer: 150
CEM: 0.0, F1: 0.5
EWC
Answer: 120
CEM: 0.0, F1: 0.0
LwF
Answer: 12
CEM: 0.0, F1: 0.0
MoELoRA
Answer: 100
CEM: 0.0, F1: 0.0
O-LoRA
Answer: 40
CEM: 0.0, F1: 0.0
SEFE
Answer:Bugatti Bolide
CEM: 0.0, F1: 0.0
KORE
Answer: 250
CEM: 0.0, F1: 0.67
LLaVA-v1.5-13B
LoRA
Answer: 400
CEM: 0.0, F1: 0.0
Replay
Answer: 200
CEM: 0.0, F1: 0.0
KORE
Answer: 250 units
CEM: 1.0, F1: 1.0
Qwen2.5-VL
LoRA
Answer: 150 units
CEM: 0.0, F1: 0.5
Replay
Answer: 99
CEM: 0.0, F1: 0.0
KORE
Answer: 250 units
CEM: 1.0, F1: 1.0