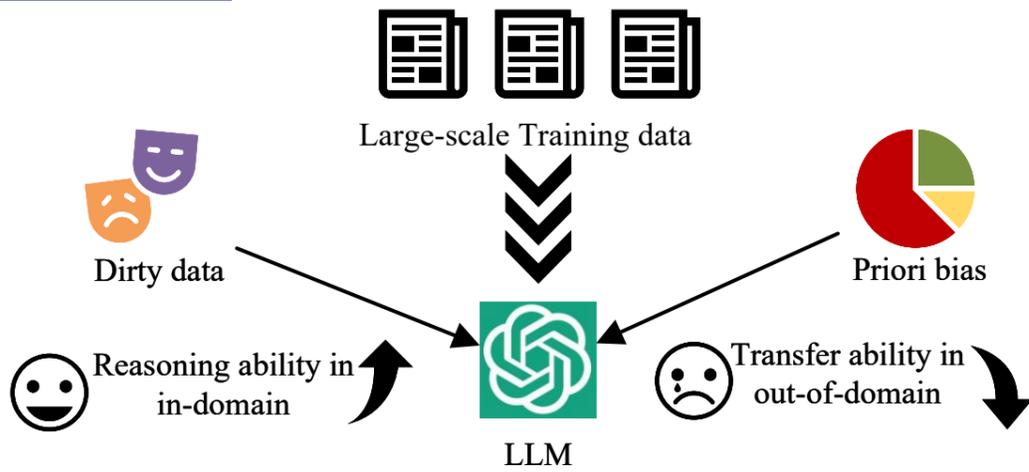


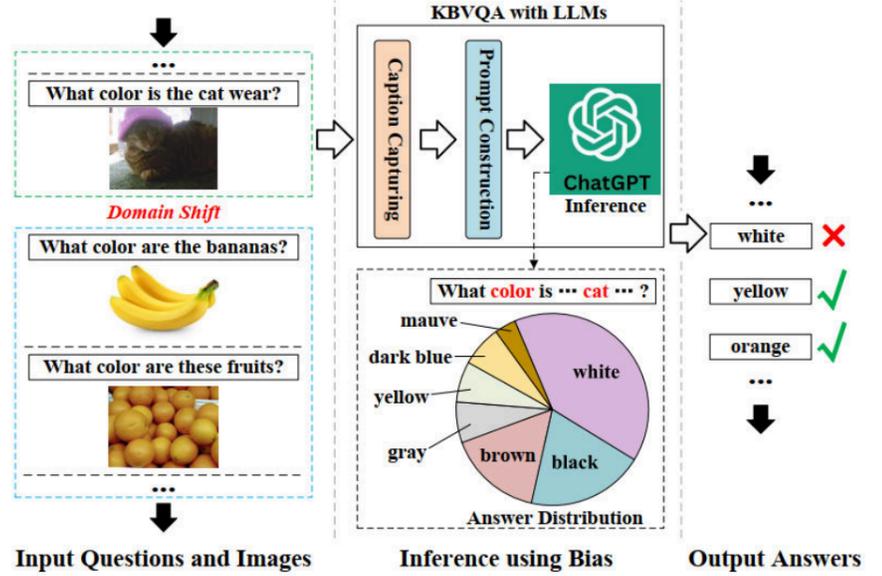
Background



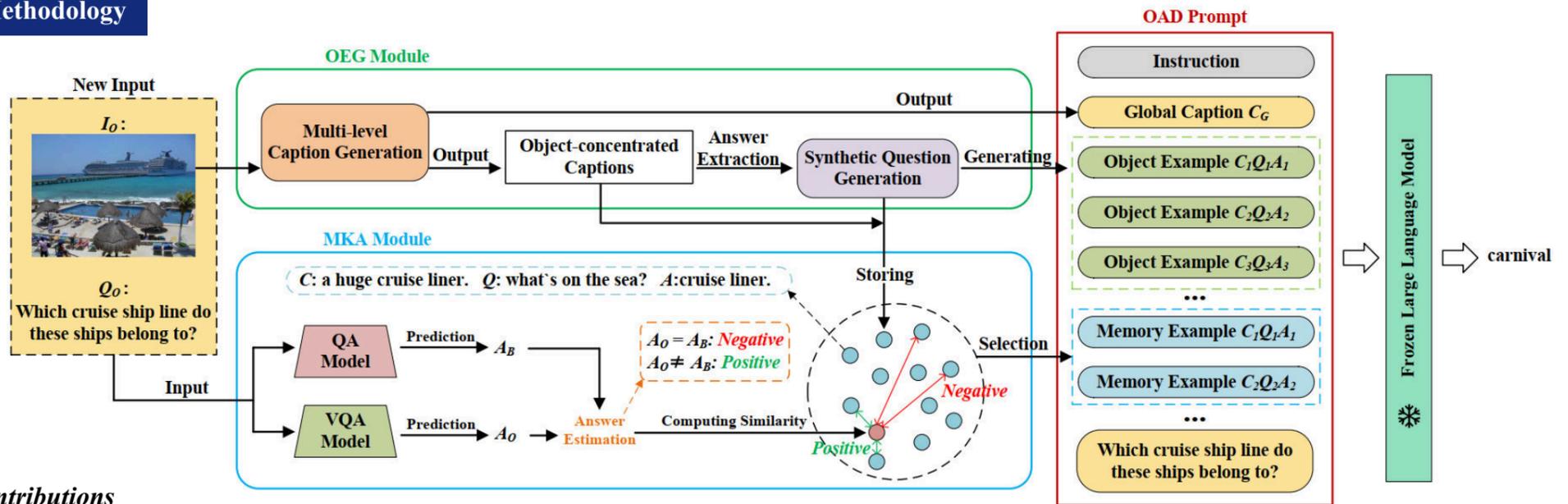
Limitations

- Limited ability in domain shift.
- Unreliable inference accuracy.

Existing Drawback



Methodology



Contributions

- We propose the OEG module, leveraging enhanced visual information to mitigate inherited language biases in LLMs.
- We design the MKA module, improving prediction reliability through memory-augmented knowledge support.
- We develop the OAD Prompt, enhancing LLM robustness in distribution-changing scenarios during inference.
- Extensive experiments demonstrate the effectiveness and generalizability of proposed OAD-Promoter.

Results

Method	Few/Zero-shot	VQAV2		A-OKVQA		OKVQA
		val	test	val	test	test
<i>Methods with Large-scale Multi-modal Pretraining</i>						
VL-T5 (Cho et al. 2021)	Zero-shot	13.50	14.13	-	-	5.73
Frozen (Tsimpoukelli et al. 2021)	Zero-shot	29.43	29.55	-	-	5.90
Flamingo-3B (Alayrac et al. 2022)	Zero-shot	48.24	49.18	-	-	41.17
Flamingo-9B (Alayrac et al. 2022)	Zero-shot	50.77	51.80	-	-	44.64
Flamingo-80B (Alayrac et al. 2022)	Zero-shot	56.08	56.21	-	-	50.57
FewVLM-base (Jin et al. 2022)	Zero-shot	43.28	43.39	-	-	11.52
FewVLM-large (Jin et al. 2022)	Zero-shot	47.68	47.73	-	-	16.50
VLKD-ViT-B/16 (Dai et al. 2022)	Zero-shot	38.60	39.69	-	-	10.50
VLKD-ViT-L/14 (Dai et al. 2022)	Zero-shot	42.55	44.48	-	-	13.24
<i>Methods with Frozen Large Language Models</i>						
PICa (Yang et al. 2022) w/ GPT-3	Few-shot	56.10	56.12	47.08	47.64	48.01
REVIVE (Lin et al. 2022) w/ GPT-3	Few-shot	57.11	57.69	57.33	57.05	58.03
KAT (Gui et al. 2022) w/ GPT-3	Few-shot	56.30	56.48	55.38	55.12	54.41
Prophet (Shao et al. 2023) w/ GPT-3	Few-shot	58.37	59.22	59.27	57.30	61.08
PromptCap (Hu et al. 2023) w/ GPT-3	Few-shot	58.18	58.77	58.86	57.28	60.44
GRACE (Zhang, Jiang, and Zhao 2024) w/ GPT-3	Few-shot	58.07	58.45	58.61	57.15	60.29
OAD-Promoter w/ GPT-3	Few-shot	57.96	58.42	58.50	56.99	60.04
PICa (Yang et al. 2022) w/ GPT-3	Zero-shot	28.67	29.30	23.74	26.88	17.63
PICa (Yang et al. 2022) w/ GPT-3 + RQP (Lan et al. 2023)	Zero-shot	28.70	29.34	28.92	27.73	20.27
Img2LLM (Guo et al. 2023) w/ GPT-3	Zero-shot	58.69	59.22	38.88	43.39	42.80
Img2LLM (Guo et al. 2023) w/ GPT-3 + RQP (Lan et al. 2023)	Zero-shot	58.96	59.35	42.88	<u>43.61</u>	45.57
Img2LLM (Guo et al. 2023) w/ OPT	Zero-shot	60.58	61.83	42.90	40.69	45.58
Img2LLM (Guo et al. 2023) w/ OPT + RQP (Lan et al. 2023)	Zero-shot	60.60	61.85	42.83	40.64	45.52
OAD-Promoter w/ OPT	Zero-shot	60.62	61.93	43.03	40.68	45.58
OAD-Promoter w/ GPT-3	Zero-shot	60.64	61.98	<u>43.09</u>	41.71	<u>45.61</u>

Table 1: Performance comparison on VQAV2, A-OKVQA, and OKVQA TEST SET. The bold fonts indicate the best results in the entire table, and the underlined fonts denote the best results among methods with frozen LLMs in the zero-shot setting. Note that we use GPT-3 in this experiment for the sake of a fair comparison.