



MissRAG: Addressing the Missing Modality Challenge in Multimodal Large Language Models

V. Pipoli*, A. Saporita*, F. Bolelli, M. Cornia, L. Baraldi, C. Grana,
R. Cucchiara, E. Ficarra

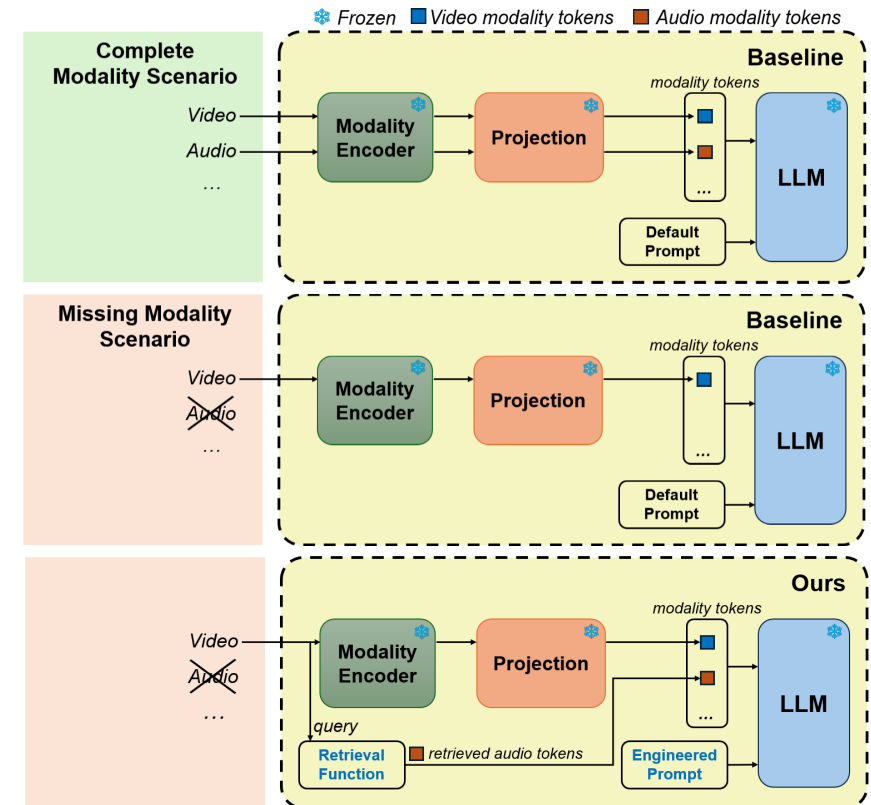
Missing Modality Problem with MLLMs

- In real-world scenarios, data samples often have **one or more absent or incomplete modalities**.

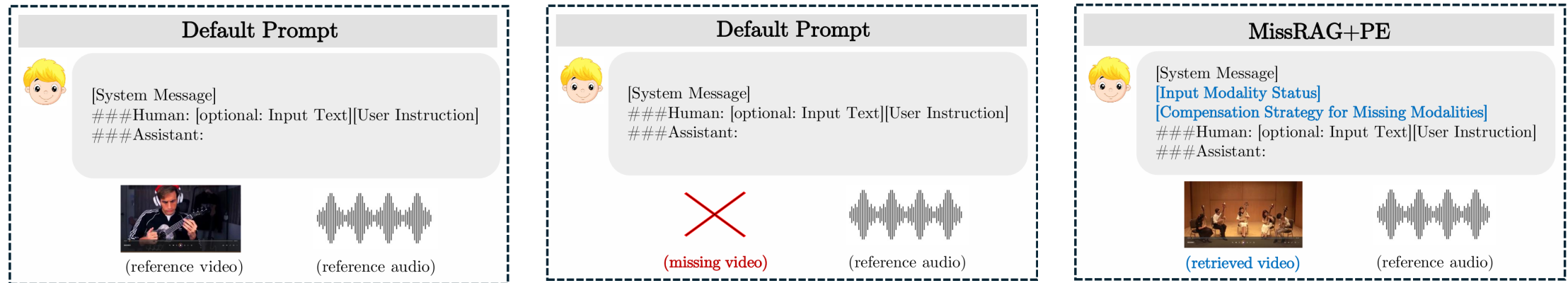
In this work, we are the first to assess the robustness of MLLMs under missing modality conditions.

- To address the missing modality problem, numerous fine-tune the models. However, fine-tuning MLLMs is computationally prohibitive.

We propose the first multimodal RAG framework to address the missing modality problem with MLLMs while keeping their parameters frozen!

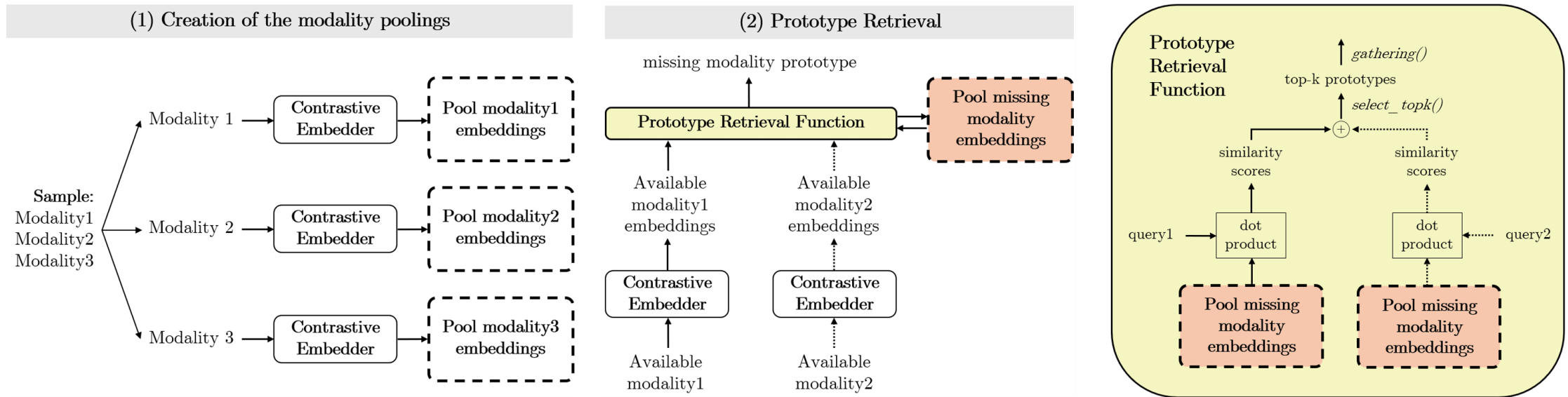


Our Proposal: MissRAG



- Our proposal to mitigate the missing modality problem consists in a multimodal RAG framework for missing modalities, called MISSRAG, enhanced with a prompt engineering strategy.
- The prompt engineering strategy is designed to inform the model about the input modality status and change its behavior with a *Compensation Strategy for Missing Modalities prompt*.

Our Proposal: MissRAG



- Our RAG framework is capable of retrieving complementary audio, visual, and textual data for any given combination of these input modalities using the information from the available modalities.

Experimental Results

Task	Audio-Video			Audio-Video						Audio-Video-Text														
	Question Answering				Captioning						Sentiment Analysis													
		Dataset			MUSIC-AVQA			VALOR32K			CharadesEgo			MOSI							MOSEI			
Metric	Accuracy			CIDEr			CIDEr			Accuracy							Accuracy							
Method	MA	MV	C	MA	MV	C	MA	MV	C	MA	MV	MT	MAV	MAT	MVT	C	MA	MV	MT	MAV	MAT	MVT	C	
ChatBridge	41.80	27.25	43.87	18.36	4.30	20.22	9.47	1.51	9.96	18.51	29.01	17.49	23.18	15.16	11.37	26.38	40.20	21.76	35.20	28.38	39.02	21.76	36.81	
MissRAG+PE	45.14	43.18	46.10	19.09	10.20	20.66	10.85	4.37	10.25	24.78	30.09	18.22	23.76	16.18	17.93	28.12	41.21	40.95	37.05	40.46	36.19	37.33	38.94	
Improvement	+3.34	+15.93	+2.23	+0.73	+5.90	+0.44	+1.38	+2.86	+0.29	+6.27	+1.08	+0.73	+0.58	+1.02	+6.56	+1.74	+1.01	+19.19	+1.85	+12.08	-2.83	+15.57	+2.13	
OneLLM	48.89	39.48	49.86	20.41	3.34	26.40	2.98	1.20	3.76	48.54	73.91	40.23	75.88	38.34	41.69	56.41	30.82	57.70	40.22	58.75	38.14	46.10	40.09	
MissRAG+PE	50.85	46.66	51.32	27.19	16.57	26.42	4.03	2.50	4.17	78.57	74.20	46.65	76.68	54.08	45.48	75.51	59.18	51.66	45.16	52.35	45.89	47.66	61.56	
Improvement	+1.96	+7.18	+1.46	+6.78	+13.23	+0.02	+1.05	+1.30	+0.41	+30.03	+0.29	+6.42	+0.80	+15.74	+3.79	+19.10	+28.36	-6.04	+4.94	-6.40	+7.75	+1.56	+21.47	
VideoLLaMA 2	77.20	59.57	79.94	19.36	7.02	22.40	10.17	0.69	12.20	46.06	43.59	10.50	74.34	10.20	4.37	25.80	41.36	44.24	28.55	53.17	28.07	22.00	36.62	
MissRAG+PE	77.53	60.11	79.97	21.42	10.30	22.64	11.96	3.29	12.62	33.38	37.76	23.76	40.52	11.02	23.34	34.69	38.63	39.47	29.20	40.14	30.31	27.15	39.17	
Improvement	+0.33	+0.54	+0.03	+2.06	+3.28	+0.24	+1.79	+2.60	+0.42	-12.68	-6.23	+13.26	-33.82	+082	+18.97	+8.89	-2.73	-4.77	+0.65	-13.03	+2.24	+5.15	+2.55	

- MissRAG **effectively mitigates the missing modality problem with MLLM** across a wide range of tasks involving audio-video, and audio-video-text data.



Scan Me!

ICCV
OCT 19-23, 2025



HONOLULU
HAWAII



V. Pipoli



A. Saporita



F. Bolelli



M. Cornia



L. Baraldi



C. Grana



R. Cucchiara



E. Ficarra