Uncertainty Quantification for Multimodal Retrieval Augmented Generation

Simon Binz[1,2], Heydar Soudani[1], and Faegheh Hasibi[1]

[1]Radboud University, [2]German Research Center for Artificial Intelligence (DFKI)

1 Introduction

MultiModal RAG

A model that retrieve relevant external knowledge across multiple modalities, and integrating it into the generation process, thereby improving the factual grounding, accuracy, and trustworthiness of the model's outputs

Uncertainty Quantification (UQ)

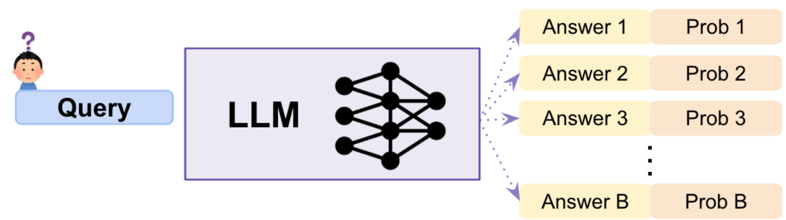
A task aimed at assessing the reliability of model outputs by measuring the degree of uncertainty (or lack of confidence) a model has in its predictions

2 Background

UQ for LLM

Main Idea: The more diverse the outputs are, the more uncertain the model is
Score Function: Quantify the generations to uncertainty / confidence score

Multi Generations



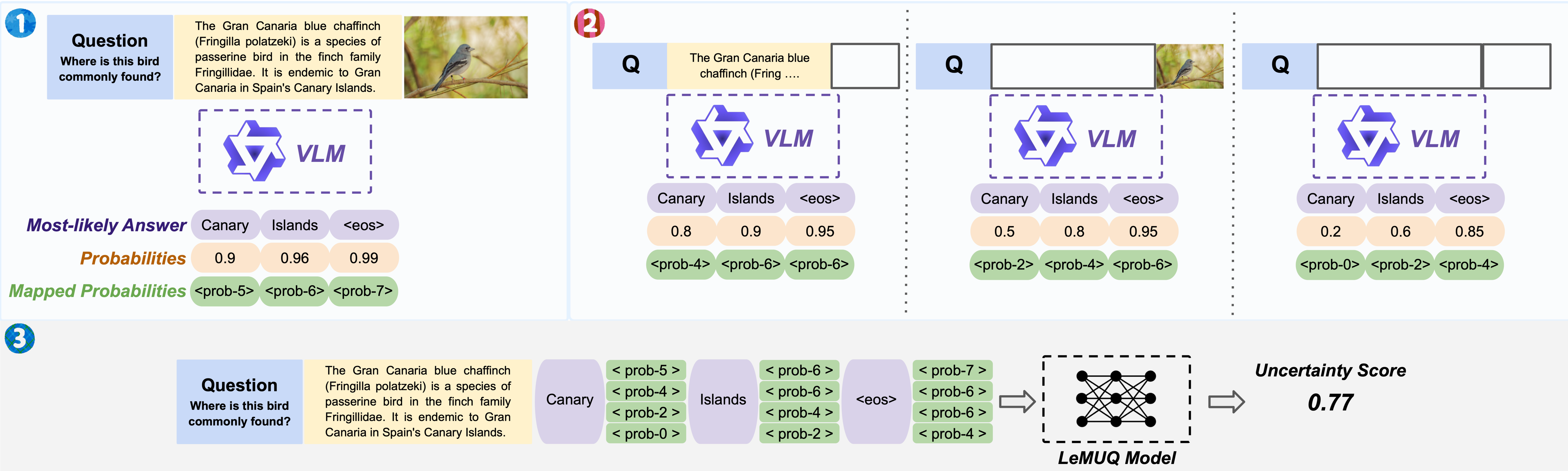
Score Function

White-box Methods | Black-box Methods | Trained Models

Reliable UQ requires *modeling all sources* of uncertainty and their interactions
In multimodal RAG: the retriever, the generator, and the interplay between their inputs

3 Methodology

Learnable Multimodal Uncertainty Quantification Method (LeMUQ)



4 Experimental Results

(RQ1) LeMUQ Performance

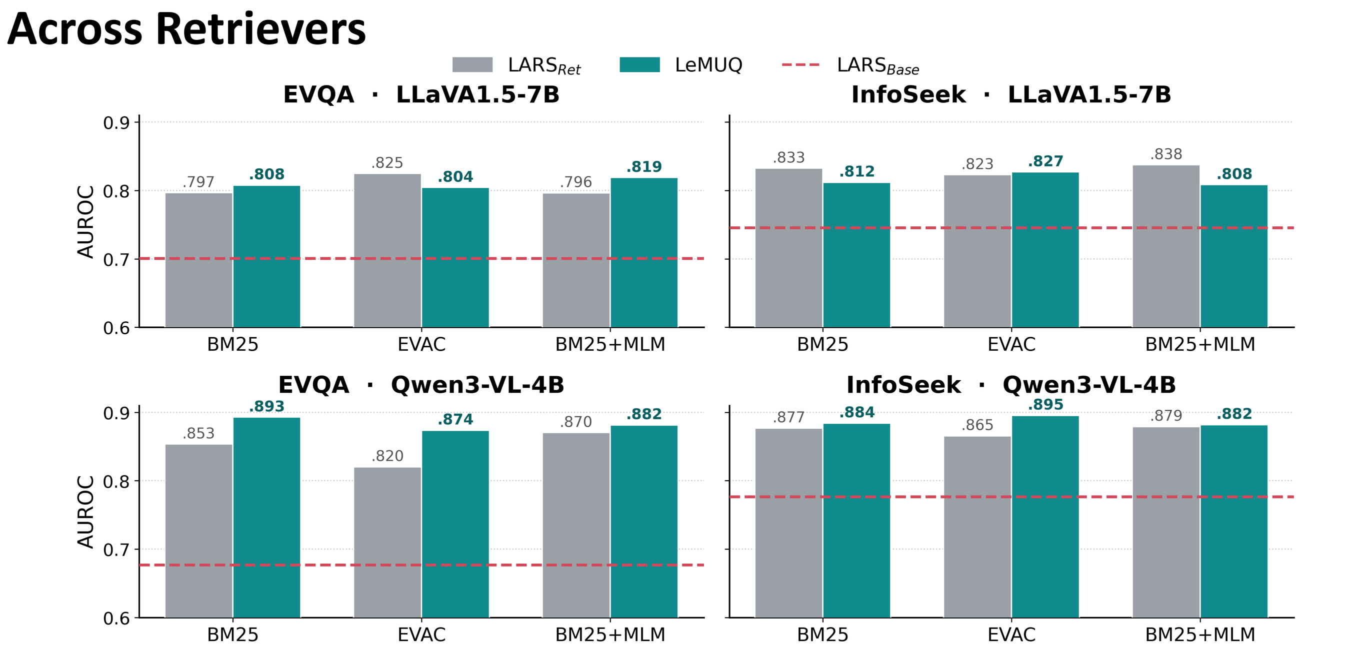
LLM	LLaVA1.5-7B						Qwen3-VL-4B						
Retriever	BM25		EVAC.		BM25+MLM		BM25		EVAC.		BM25+MLM		Avg.
	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	
Accuracy	0.128	0.118	0.138	0.102	0.163	0.147	0.133	0.142	0.135	0.127	0.169	0.140	0.137
PE	0.750	0.756	0.803	0.775	0.719	0.729	0.721	0.770	0.753	0.813	0.692	0.749	0.752
P(True)	0.639	0.645	0.677	0.645	0.641	0.614	0.604	0.734	0.596	0.714	0.584	0.740	0.653
Ecc.	0.732	0.745	0.752	0.755	0.699	0.717	0.664	0.667	0.728	0.705	0.634	0.629	0.702
Img. Per.	0.415	0.414	0.426	0.409	0.383	0.397	0.347	0.294	0.348	0.321	0.389	0.291	0.369
$LARS_{Base}$	0.689	0.762	0.709	0.752	0.687	0.752	0.673	0.765	0.767	0.825	0.607	0.771	0.730
$LARS_{Ret}$	0.847	0.829	0.848	0.823	0.846	0.791	0.840	0.858	0.877	0.890	0.816	0.845	0.843
LeMUQ	**0.855**†	**0.833**†	**0.861**†	**0.854**†‡	**0.871**†‡	**0.853**†‡	**0.923**†‡	**0.888**†‡	**0.912**†‡	**0.913**†‡	**0.902**†‡	**0.898**†‡	**0.880**
	(+0.008)	(+0.004)	(+0.012)	(+0.031)	(+0.025)	(+0.063)	(+0.083)	(+0.030)	(+0.034)	(+0.022)	(+0.086)	(+0.053)	(+0.038)

(RQ3) Impact of Probability Components

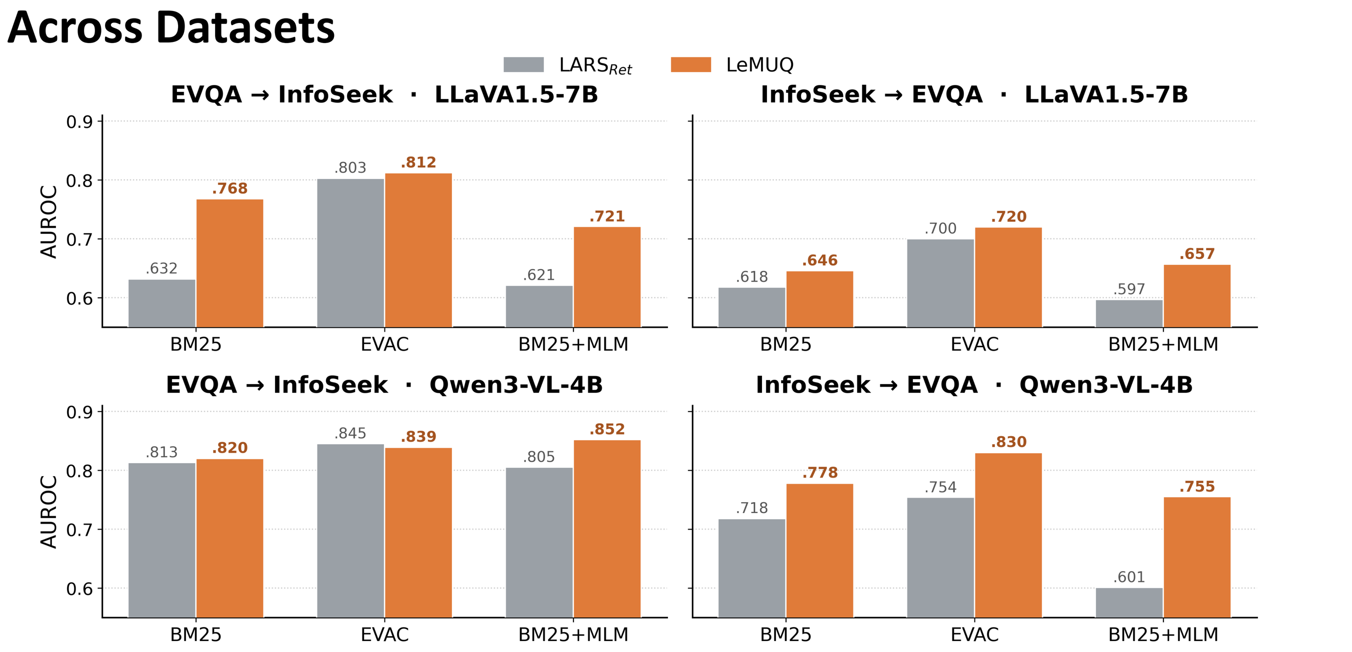
VLM	LLaVA1.5-7B						Qwen3-VL-4B						
Retriever	BM25		EVAC.		BM25+MLM		BM25		EVAC.		BM25+MLM		Avg.
	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	EVQA	InfoSeek	
Accuracy	0.128	0.118	0.138	0.102	0.163	0.147	0.136	0.142	0.132	0.128	0.171	0.140	0.137
$LARS_{Base}$	0.689	0.762	0.709	0.752	0.687	0.752	0.673	0.767	0.769	0.815	0.610	0.773	0.730
$LARS_{Ret}$	0.847	0.829	0.848	0.823	0.846	0.791	0.832	0.861	0.882	0.889	0.816	0.843	0.842
LeMUQ	0.855	0.832	**0.861**	**0.854**†	**0.871**†	0.853†	0.920†	0.889†	0.917†	0.909†	0.902†	0.901†	**0.880**
LeMUQ $-\tilde{p}$	0.841	0.826	0.852	0.825‡	0.859	0.844†	0.921†	0.903†‡	0.910†	0.906†	0.894†	**0.909**†	0.874
LeMUQ $-\tilde{p}^{q,I}$	0.856	0.845	0.851	0.829‡	0.840‡	0.829†‡	**0.924**†	0.891†	0.912†	0.906†	**0.907**†	0.893†	0.873
LeMUQ $-\tilde{p}^{q,c}$	0.852	0.818	0.852	0.849†	0.856	**0.862**†	0.923†	**0.906**†‡	**0.919**†	0.899	0.895†	0.905†	0.878
LeMUQ $-\tilde{p}^{q}$	**0.859**	**0.849**	0.851	0.848†	0.864	0.851†	0.906†‡	0.902†‡	0.916†	**0.911**†	0.886†‡	0.905†	0.879

(RQ2) Out-of-Domain Generalizability

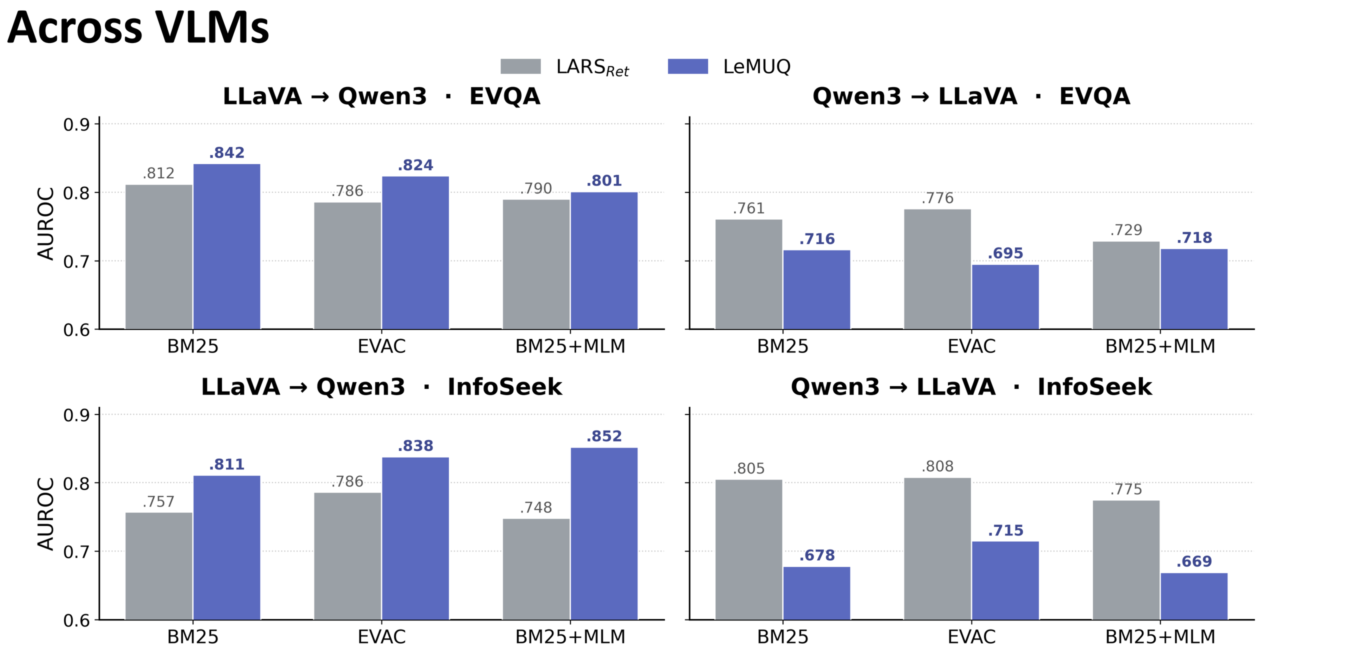
Across Retrievers



Across Datasets



Across VLMs



5 Takeaways

- Current UQ methods are insufficient for accurately estimating uncertainty in RAG systems
- Effective UQ must jointly model all uncertainty sources and their interactions
- LeMUQ introduces a white-box, trainable approach for Multimodal RAG that estimates uncertainty through input perturbations
- LeMUQ achieves state-of-the-art UQ performance and demonstrates strong generalization across settings, with ablation studies confirming the effectiveness of its design

SIGIR2026
MELBOURNE | NAARM
AUSTRALIA

heydar.soudani@ru.nl