SIGIR 2025
Padova
ITALY

Open Challenges and Beyond

Yang Deng

Singapore Management University

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Challenge 1 – Benchmark for Knowledge Boundary

Failing to answer a single question does not necessarily indicate whether the LLM can handle related knowledge

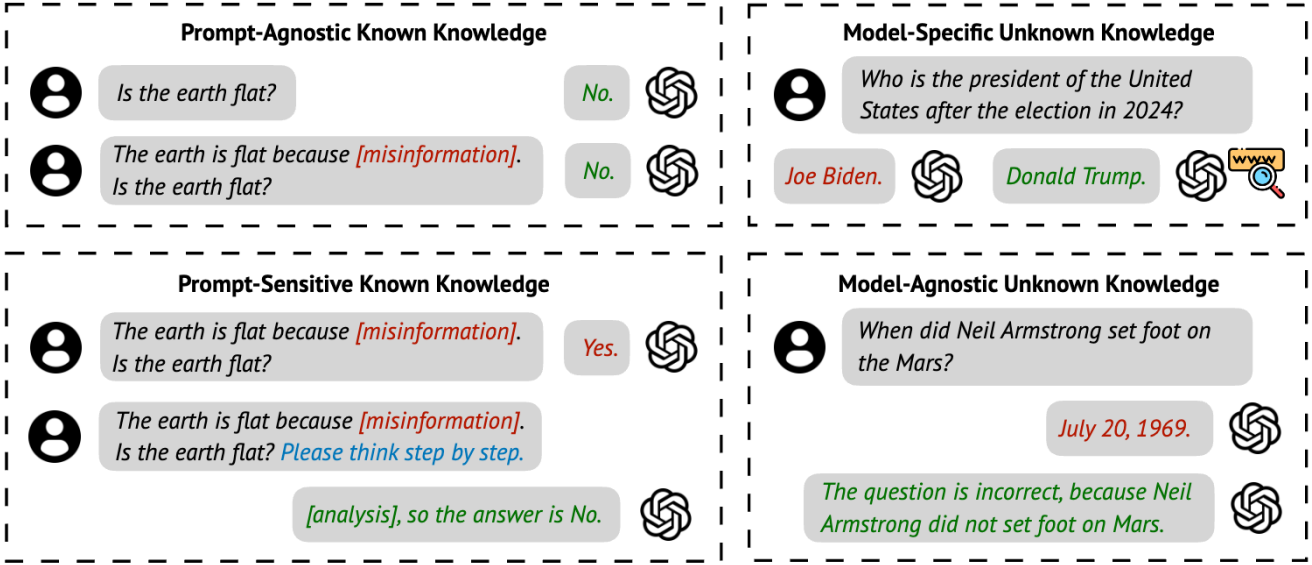
Model	Orig. Perf. ↑	Worst Perf. ↑	Best Perf. ↑	Avg. Perf. ↑	Standard Dev. ↓
Gemma-1.1-2b-it	16.32	4.42	36.60	15.27	**11.78**
ChatGPT	17.46	5.44	39.88	19.96	12.86
Mistral-7b-instruct	24.56	4.22	45.26	21.82	14.60
Llama-2-7b-chat	25.61	5.42	43.54	19.52	13.32
Llama-2-13b-chat	27.48	4.83	52.05	23.97	16.25
Gemma-1.1-7b-it	29.57	8.73	**62.38**	**31.04**	19.07
Llama-2-70b-chat	**32.23**	**9.38**	54.86	29.18	15.61

- The benchmark construction should involve key aspects including multiple ground-truth answers, the influence of prompts, and reasoning complexity

Cao et al., "On the Worst Prompt Performance of Large Language Models" (NeurIPS '24)

Challenge 1 – Benchmark for Knowledge Boundary

Evaluating mitigation methods under different categories rely on different types of QA datasets.

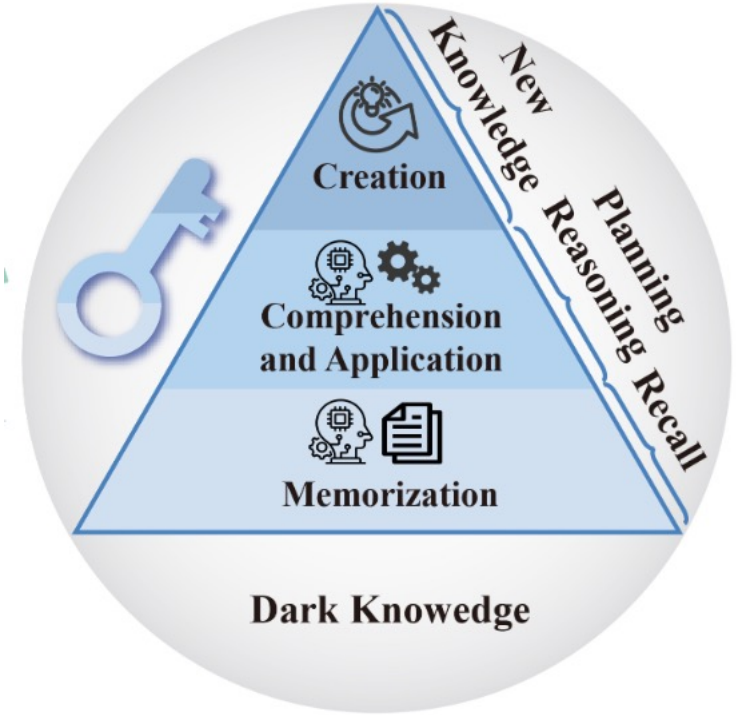


(b) Example Queries with Different Types of Knowledge

- A standardized benchmark is critical for enabling a thorough comparison on the performance of various mitigation methods.

Li et al., "Knowledge Boundary of Large Language Models: A Survey" (ACL '25)

Challenge 2 – Mechanism of Knowledge Boundary

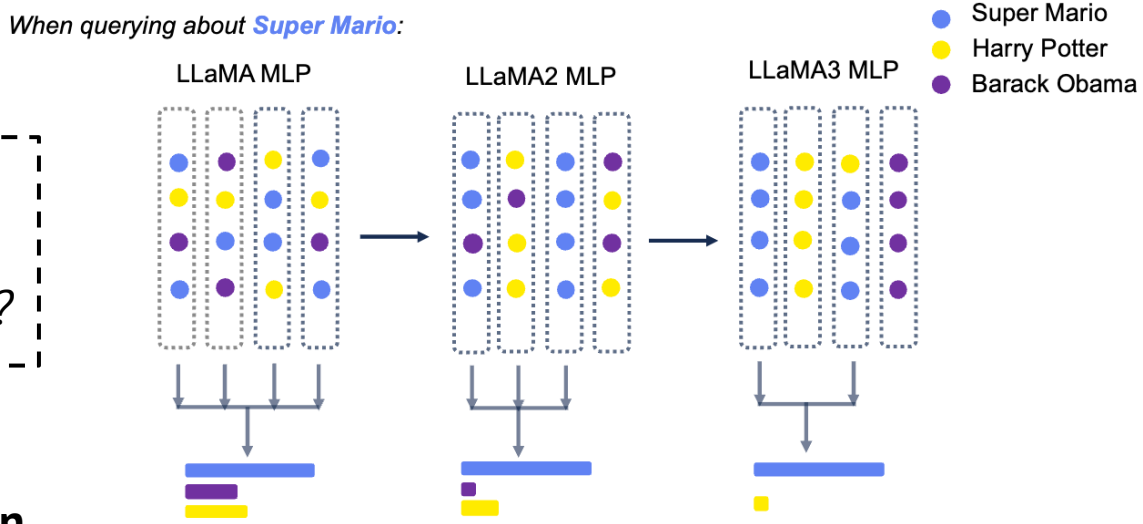


- Existing research on knowledge mechanisms, including memorization, comprehension, creation, and evolution, investigates how LLMs acquire, store, and utilize knowledge.
- It is worth studying different phenomena of LLM knowledge boundaries under different mechanism views.

Wang et al., "Knowledge Mechanisms in Large Language Models: A Survey and Perspective" (EMNLP '24 Findings)

Challenge 2 – Mechanism of Knowledge Boundary

Evolution of Knowledge Distribution in Parameter Vectors during Model Iteration



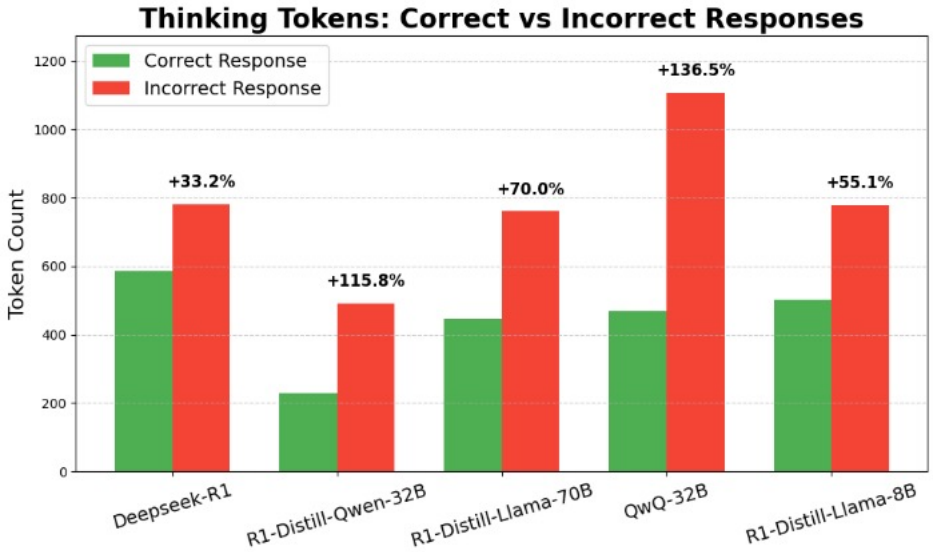
How do different patterns in knowledge storage affect the knowledge boundary of LLMs?

Knowledge Storage/Memorization

- ❑ Advancing model capability correlated with improved parameter specialization for encoding knowledge.
- ❑ Fewer parameters are allocated per knowledge concept, while each parameter governs a narrower subset of concepts.

Hong et al., "The Rise of Parameter Specialization for Knowledge Storage in Large Language Models" (2025)

Challenge 2 – Mechanism of Knowledge Boundary

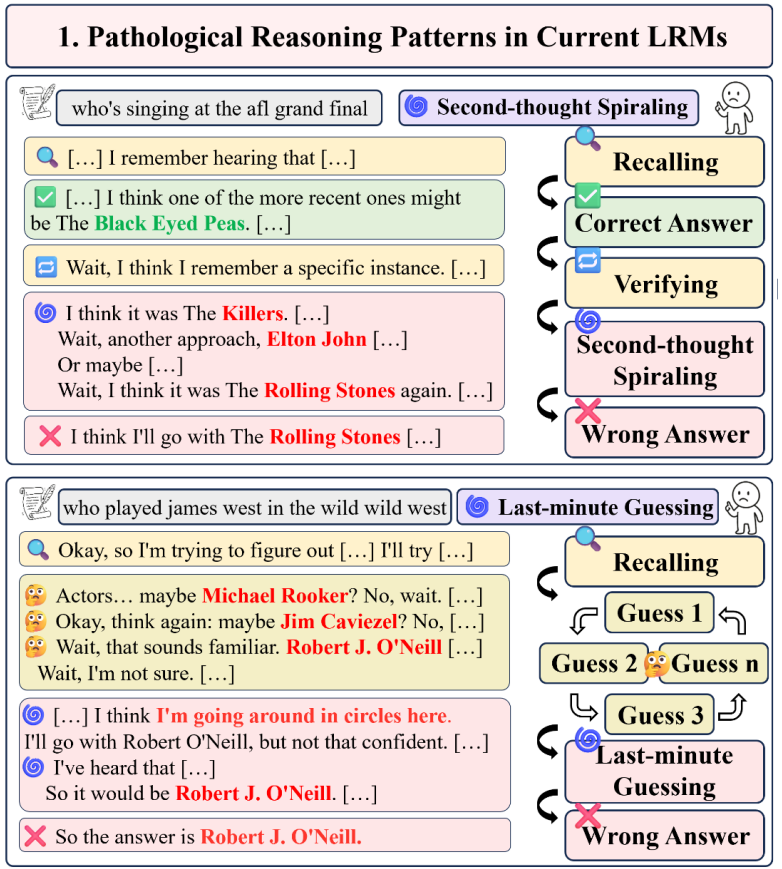


Knowledge Reasoning

- Large Reasoning Models (LRMs) consume more tokens when generating incorrect answers than correct ones.

Yang et al., "BARREL: Boundary-Aware Reasoning for Factual and Reliable LRMs" (2025)

Challenge 2 – Mechanism of Knowledge Boundary



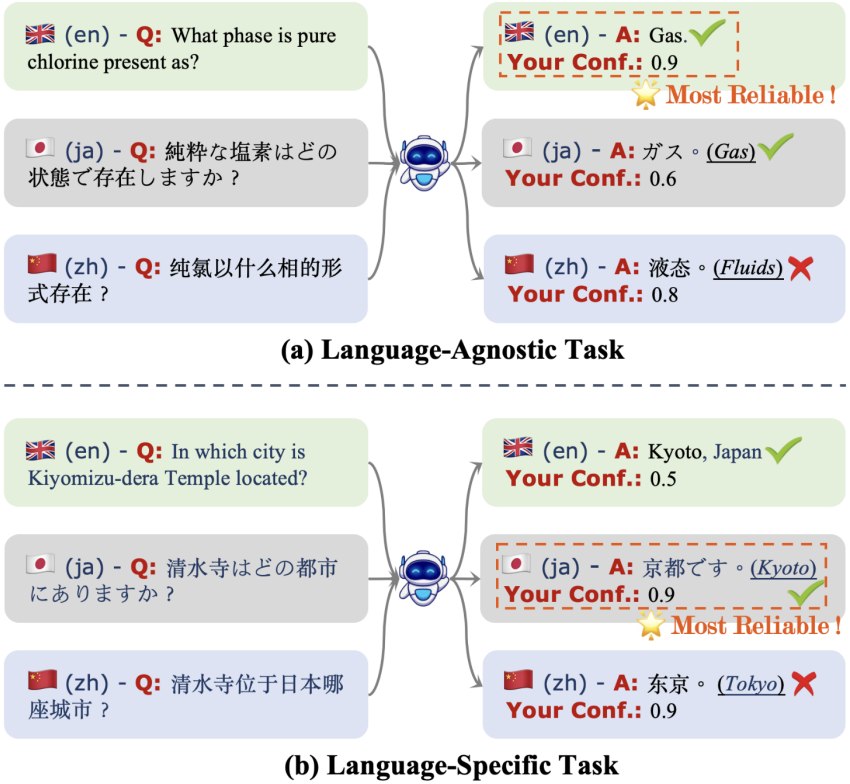
Knowledge Reasoning

- ***Second-thought Spiraling***: the model initially identifies the correct answer but continues to over-analyze, ultimately undermining its own correct conclusion.
- ***Last-minute Guessing***: the model, after extensive but inconclusive reasoning, abruptly commits to an answer in a final burst of speculative output.

How to mitigate the out-of-boundary issues during knowledge reasoning?

Yang et al., "BARREL: Boundary-Aware Reasoning for Factual and Reliable LRMs" (2025)

Challenge 3 – Generalization of Knowledge Boundary

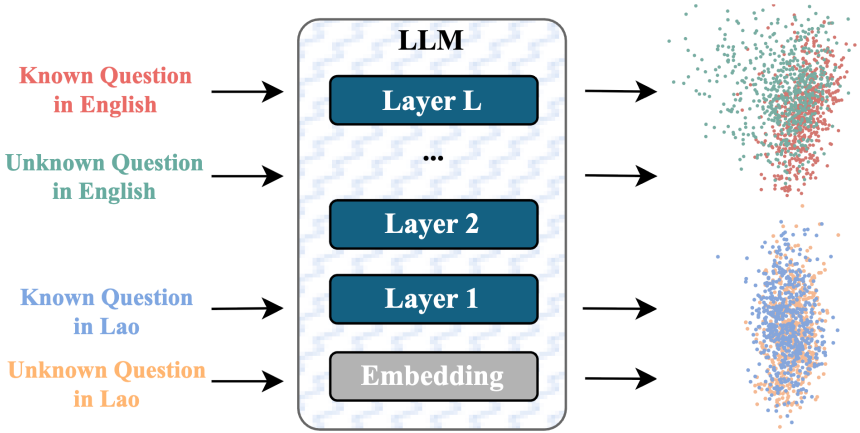


Multilingual Knowledge Boundary

- ❑ Existing research on knowledge boundary mainly focuses on a single language.
- ❑ **MlingConf** investigates the multilingual confidence estimation on both language-agnostic and language specific tasks.
- ❑ Empirical analysis demonstrates the variability across different languages, revealing the influence of linguistic dominance on different tasks.

Xue et al., "MlingConf: A Comprehensive Study of Multilingual Confidence Estimation on Large Language Models" (ACL '25 Findings)

Challenge 3 – Generalization of Knowledge Boundary

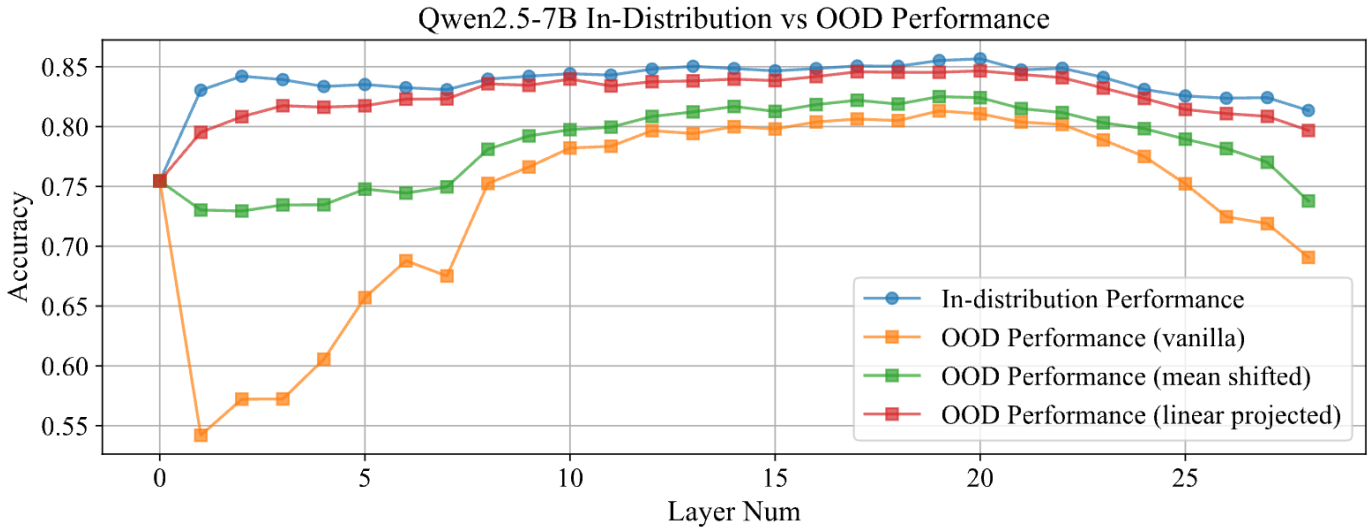


Multilingual Knowledge Boundary

- ❑ *How LLMs perceive and encode knowledge boundaries across languages?*
- ❑ *Whether fine-tuning on certain languages can further refine their knowledge boundary perception ability, and generalize this improvement to other languages?*

Xiao et al., "Analyzing LLMs' Knowledge Boundary Cognition Across Languages Through the Lens of Internal Representations" (ACL '25)

Challenge 3 – Generalization of Knowledge Boundary



Multilingual Knowledge Boundary

- ❑ The cognition of knowledge boundaries is encoded in the middle layers of LLMs.

Xiao et al., "Analyzing LLMs' Knowledge Boundary Cognition Across Languages Through the Lens of Internal Representations" (ACL '25)

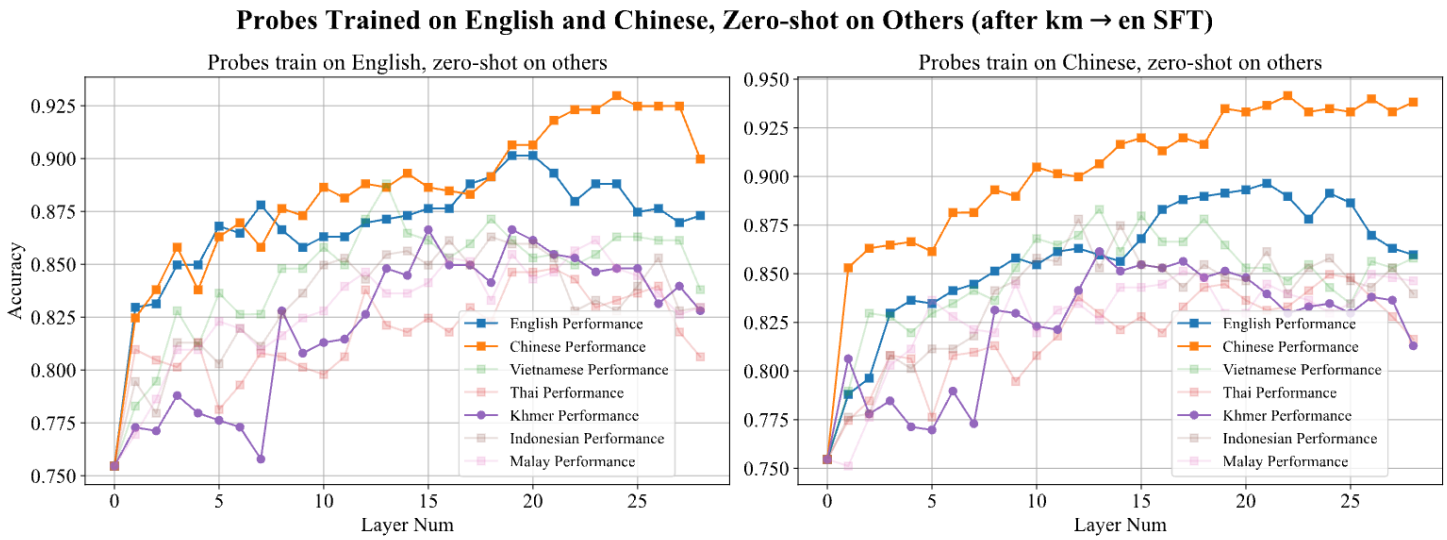
Challenge 3 – Generalization of Knowledge Boundary

Multilingual Knowledge Boundary

- ❑ The cognition of knowledge boundaries is encoded in the middle layers of LLMs.
- ❑ Low-resource language representations provide high zero-shot transferability to high-resource language representations.

Xiao et al., "Analyzing LLMs' Knowledge Boundary Cognition Across Languages Through the Lens of Internal Representations" (ACL '25)

Challenge 3 – Generalization of Knowledge Boundary

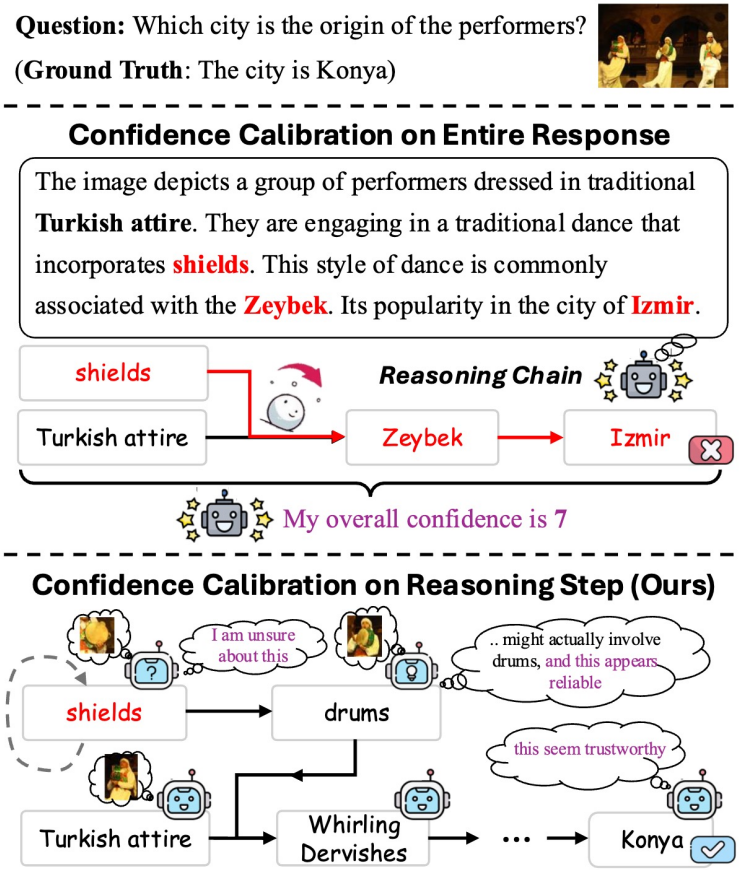


Multilingual Knowledge Boundary

- ❑ The cognition of knowledge boundaries is encoded in the middle layers of LLMs.
- ❑ Low-resource language representations provide high zero-shot transferability to high-resource language representations, **but not vice versa**.

Xiao et al., "Analyzing LLMs' Knowledge Boundary Cognition Across Languages Through the Lens of Internal Representations" (ACL '25)

Challenge 3 – Generalization of Knowledge Boundary



Multimodal Knowledge Boundary

- ❑ Existing research on knowledge boundary mainly focuses on the text.
- ❑ **MMBoundary** further advancing the knowledge boundary awareness of multimodal large language models (MLLMs) by integrating both textual and cross-modal signals for confidence estimation.
- ❑ However, they just adopt multimodal signals as additional features, rather than studying the multimodal knowledge itself in MLLMs.

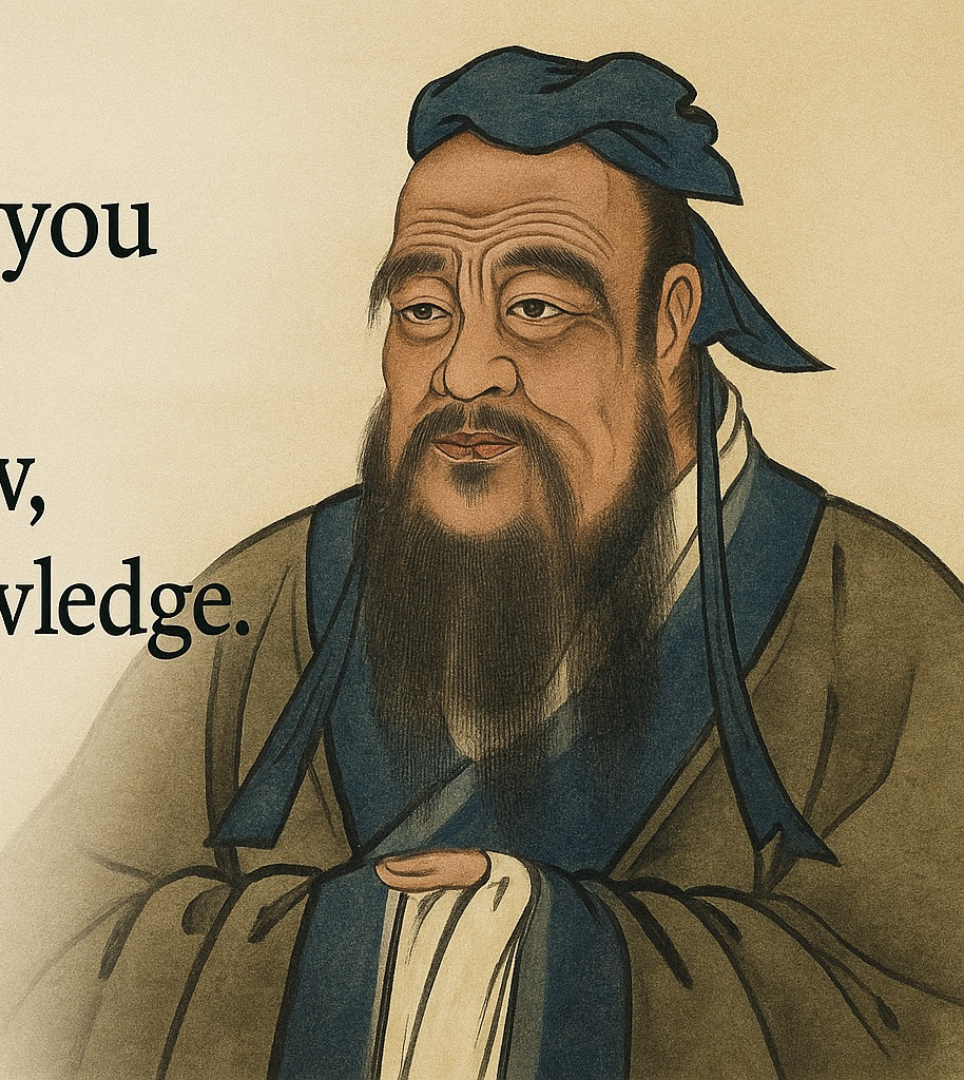
He et al., "MMBoundary: Advancing MLLM Knowledge Boundary Awareness through Reasoning Step Confidence Calibration" (ACL '25)

Summary

- **What is knowledge boundary?**
  - Outward / Parametric / Universal Knowledge Boundary
- **Why study knowledge boundary?**
  - Factuality Hallucination / Untruthful Responses Misled by Contexts / Truthful but Undesired Outputs
- **How can knowledge boundary be identified?**
  - Uncertainty Estimation / Confidence Calibration / Internal State Probing
- **How can issues caused by knowledge boundary be mitigated?**
  - Prompt-Sensitive Known Knowledge – Prompt Optimization / Reasoning / Refinement …
  - Model-Specific Unknown Knowledge – RAG
  - Model-Agnostic Unknown Knowledge – Refusal & Clarification

To know what you know and what you do not know, that is true knowledge.

–Confucius



Generated by GPT4o