

Towards LLM-KG Symbiosis for Reducing Factual Hallucinations

Tommaso Dolci, Milos Jovanovik, and Katja Hose

TU Wien – DMKI Lab



Tampere,
24th Mar 2026

LLM Hallucinations

“ The generation of incorrect or nonsensical information ”



Social Problem

- Large **adoption** of LLM-based systems in the industry
- Users **blindly trust** LLMs as source of knowledge (e.g., AI Overviews)

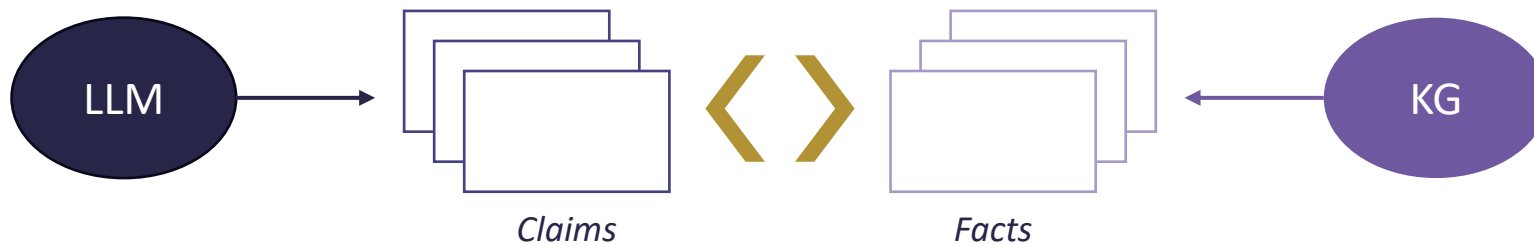


Technical Challenge

- LLMs are **stochastic** models and not trained for factual accuracy
- Training data is **not curated** and in the order of **trillions** of tokens

Knowledge Graphs and Hallucinations

- Knowledge bases with **fact representations** in the form of triples
- **Curated** knowledge from **trustworthy** sources (e.g., Wikipedia) – better than Internet retrieval!
- Emerged as a solution for **detection** and **mitigation** of hallucinations



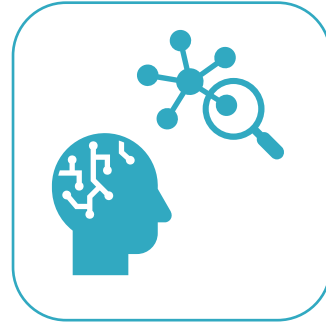
KG-Based Mitigation Approaches



KG-Based Retrieval

+ RAG on knowledge graphs

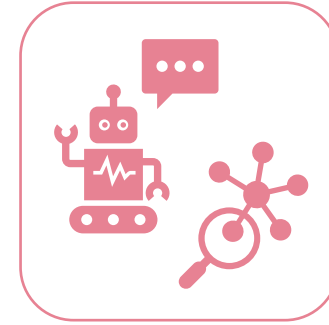
- Incompleteness



KG-Based Retrieval with Reasoning

+ CoT after retrieval to overcome incompleteness

- Closed domain scenarios



KG-Based Comparison

+ Post-generation comparison works in open domains

- Long responses and unverifiable claims

Open Challenges

C1 – Multiple-Claim Evaluation

- Claim **decomposition** and metrics **aggregation**
- **Meaning** and application of hallucination metrics

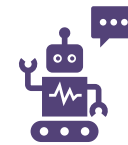
C2 – Incomplete Knowledge

- Improve KG retrieval with **LLM reasoning**
- Multiple sources and **inconsistency** management

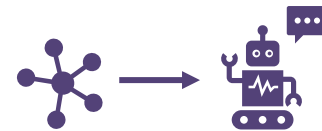
C3 – Uncertainty

- **Differentiate** verifiable from unverifiable claims
- **Uncertain** and **opinionated claims** can produce misinformation and bias

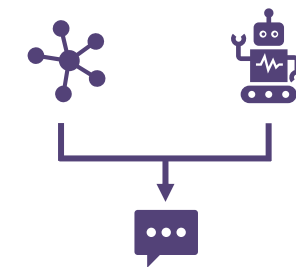
Q: What are common uses of a microwave oven?



Microwaves *reheat* food, cook simple meals and *charge electronic devices*.



Microwaves *reheat* food.



Microwaves *reheat* food, cook simple meals, and *sterilize medical equipment*. They are the *best kitchen appliance*.



Research Questions

R1 – How to **integrate KGs and LLMs** to **evaluate multiple-claim responses** with a mixture of factual, non-factual and unverifiable claims?

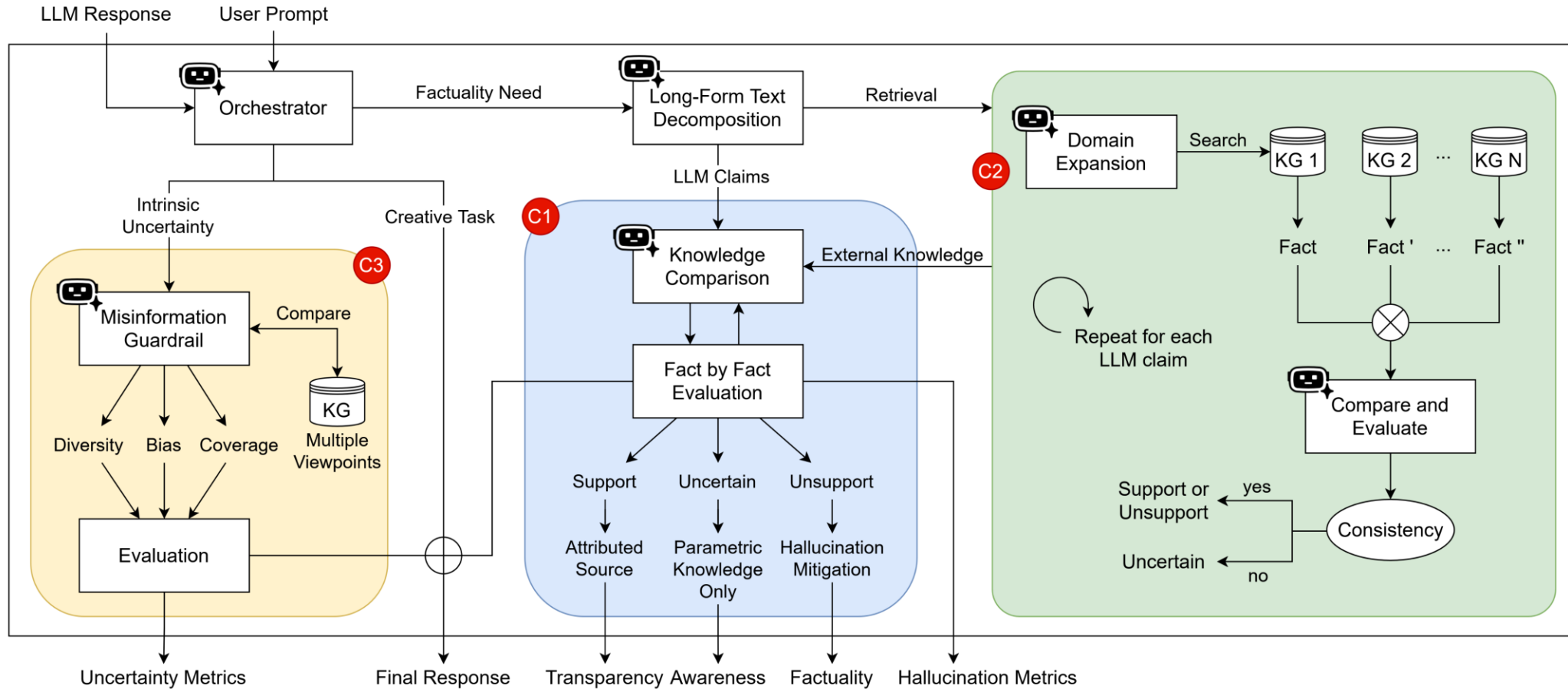
R2 – How to strengthen **LLM-KG synergy** to reduce hallucinations and address incompleteness?

R3 – How to manage **uncertainty** in factuality assessment?

LLM-KG symbiosis



Framework for LLM-KG Symbiosis



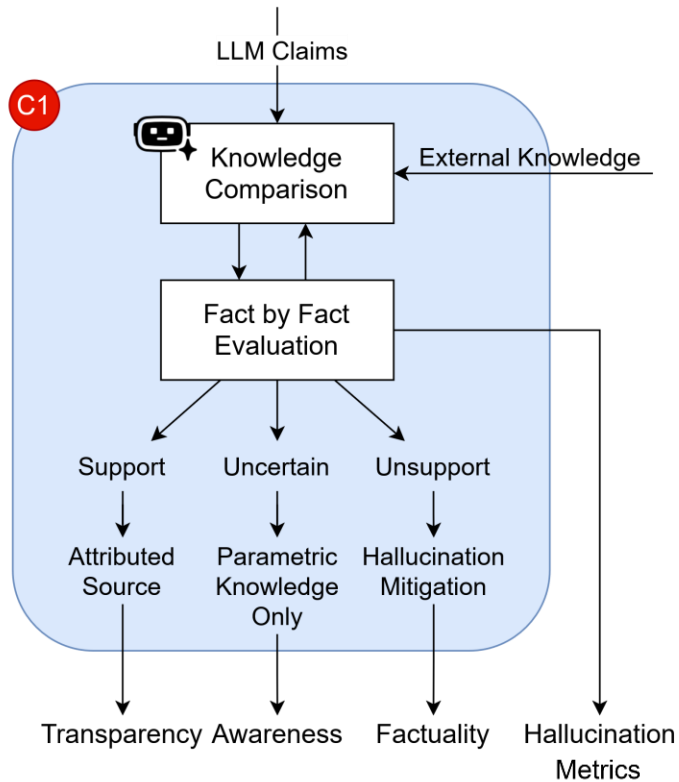
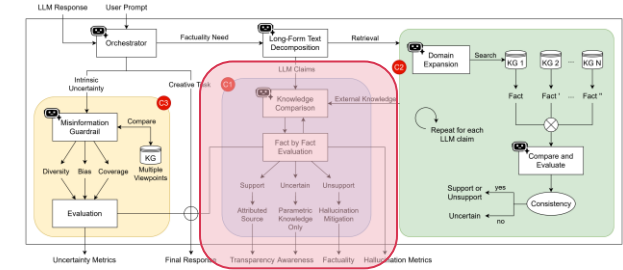
C3 = Uncertainty

C1 = Multiple Claims

C2 = Incomplete Knowledge



Evaluation Module



Aggregates claim-level into **response-level** indicators



Transparency: attributes source for supported claims

Factuality: mitigates unsupported claims

Awareness: flags unverifiable claims to inform users



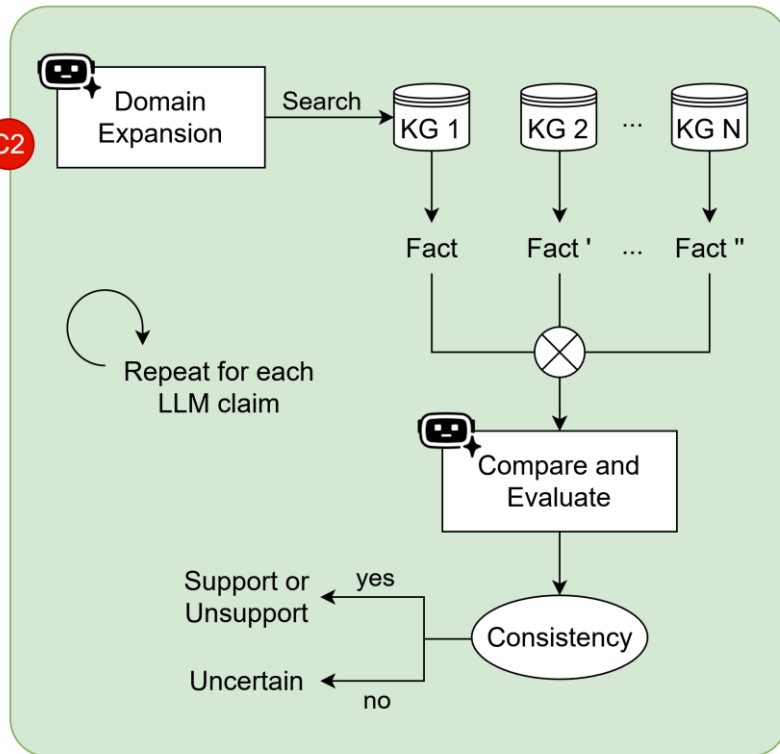
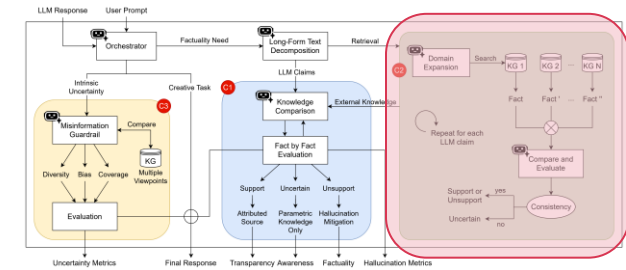
Statistical metrics for supported and unsupported claims

Plausibility via semantic and structural similarity

C1 = Multiple Claims



Retrieval Module



Hallucination detection across **multiple sources and domains**



LLM reasoning to overcome **incompleteness**

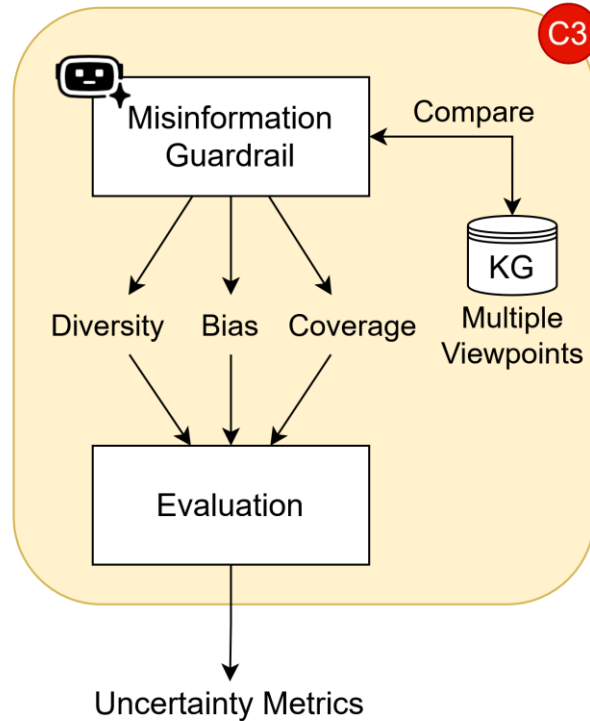
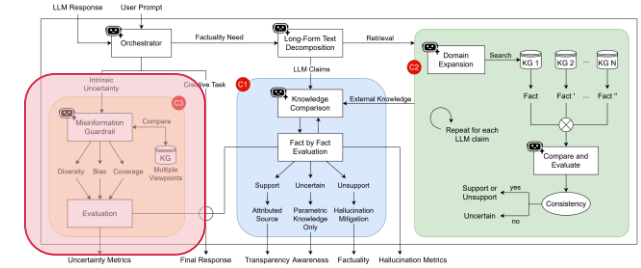


Consistency evaluation to identify knowledge uncertainty

C2 = Incomplete Knowledge



Uncertainty Module



Addresses properties **beyond factual accuracy**

Considers **opinionated** claims and inherently **uncertain** domains (e.g., politics, philosophy)



Diversity and multi-cultural coverage



Uncertainty metrics to achieve trustworthy AI

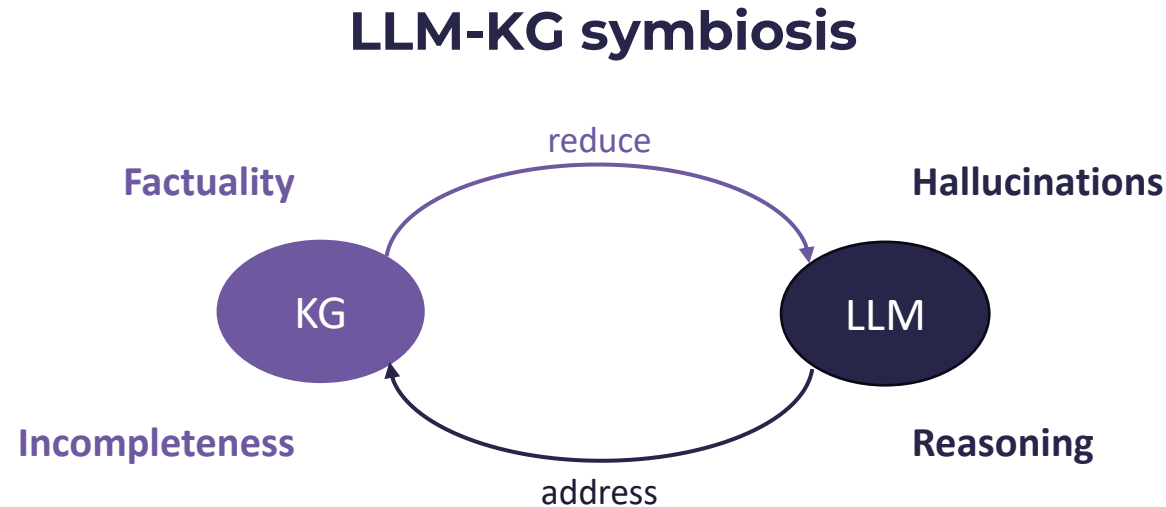
E.g., coverage, bias, perspectivism...

C3 = Uncertainty



Conclusion

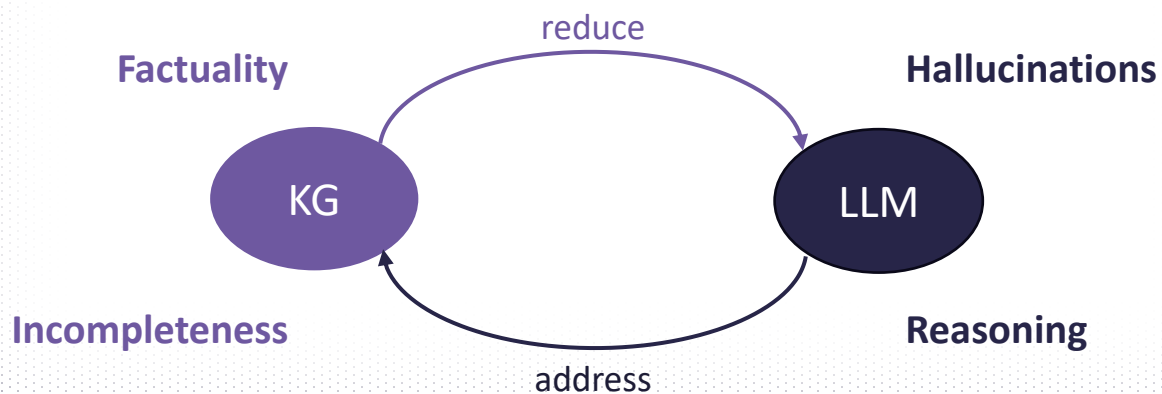
- Modular LLM-KG **framework** for hallucination detection and mitigation
- Address key **challenges**: multi-claim factuality, incomplete KG retrieval, and uncertainty
- **Next steps**: retrieval optimisation, metrics definition, reasoning and agentic AI integration, KG biases



Thank You

Questions?

LLM-KG symbiosis



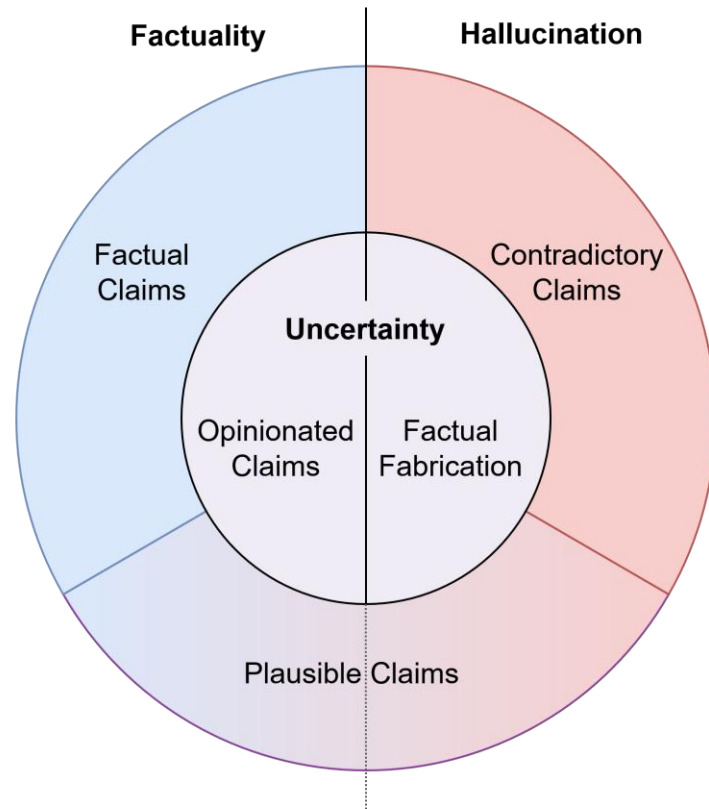
Contacts

Comparison of the Literature

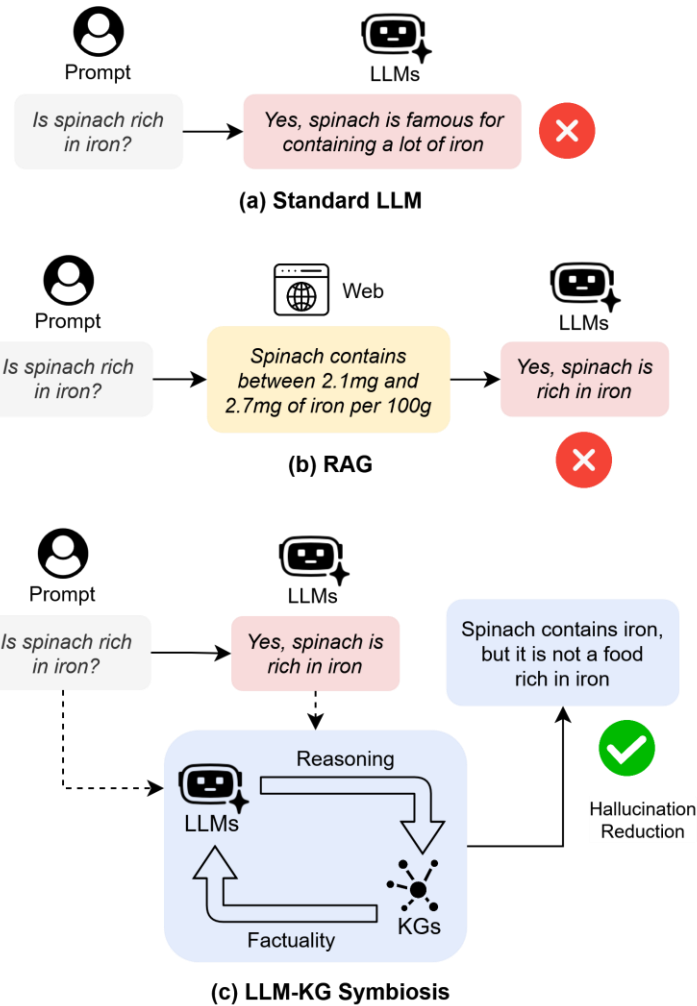
Mitigation Approach	Notable Examples	Opportunities				
		Implicit Knowledge	External Knowledge	Reasoning on KG	Multiple-Claim Evaluation	Uncertainty
No Mitigation (standard LLM)	[16, 17]	✓	–	–	–	–
KG-Based Retrieval	[18]	–	✓	–	–	–
KG-Based Retrieval with Reasoning	[10, 19]	○	✓	✓	–	–
KG-Based Comparison	[9]	✓	✓	–	–	–
KG-Based Comparison on Long Text	[12, 20]	✓	✓	–	○	○
<i>LLM-KG Symbiosis</i>		✓	✓	✓	✓	✓



Uncertainty and Plausibility



RAG Limits



Knowledge Comparison

"Obama is a politician from USA born in Chicago"

