InfAI
Institut für Angewandte Informatik

Current results on LLMs Knowledge Graph Capabilities

LSWT - 12. 6. 2025

Lars-Peter Meyer & Johannes Frey & ...
{ LPMeyer | Frey }@infai.org

code

DOI 10.5281/zenodo.15100803

data

DOI 10.5281/zenodo.15100646

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung
FKz.: 16DTM312D

Bundesministerium für Wirtschaft und Klimaschutz
FKZ: 01MK22001A
FKz.: 01MK21007A
KISS & CoyPu

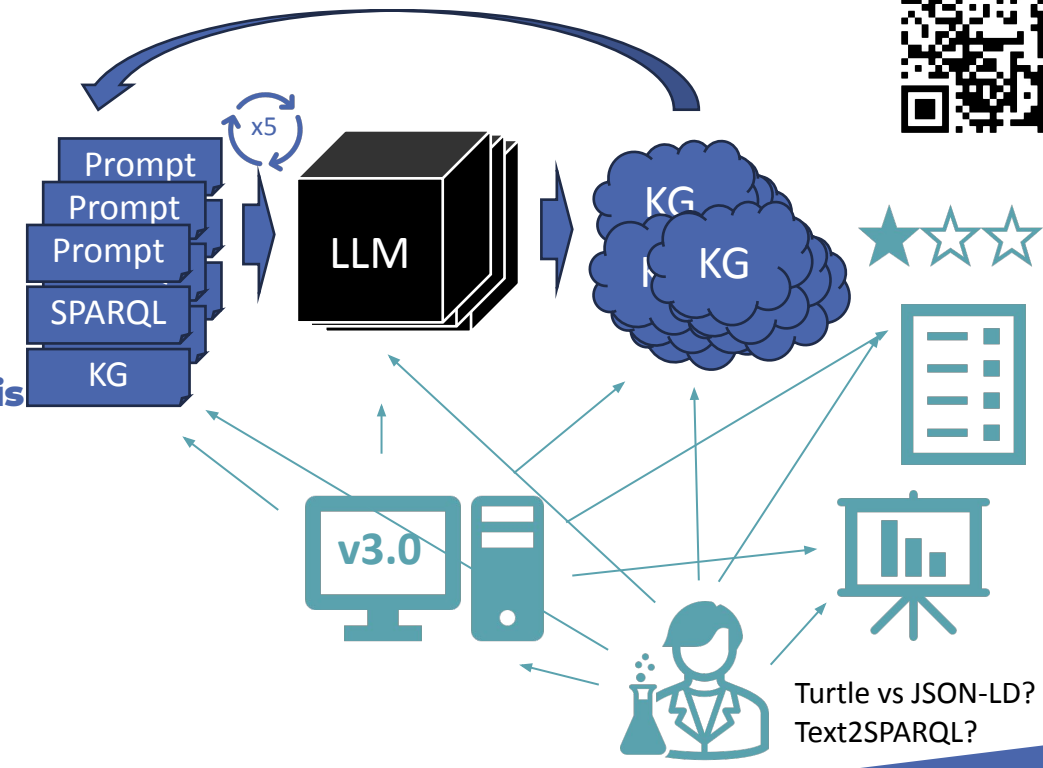
Bundesministerium für Verkehr
FKZ: 19F2266A
MobyDex

LLM Benchmarking for KGs

InfAI
Institut für Angewandte Informatik

https://github.com/AKSW/LLM-KG-Bench

- 2023: ChatGPT 3.5, ChatGPT 4
  - Manual Evaluation [1]
- 2023: More Models, More Families (Claude, Llama, …)
  - LLM-KG-Bench v1.0 [2], v1.1 [3]
- 2024: SPARQL / Dialog / Evolution
  - LLM-KG-Bench v1.2 [4], v2.0 [5]
- **2025: new API, new Tasks, new Models, Encryption**
  - **LLM-KG-Bench v3.0 [6,7,8]**
  - **Automated, Modular, Config, Logging, Reeval., Analysis**
  - **Dataset: 40 LLMs, >7 Tasks, 20-50 Iterations**



[1] Meyer et al. 2024: "LLM-assisted Knowledge Graph Engineering ...."

[2] Meyer et al. 2023: "Developing a Scalable Benchmark ...."

[3] Frey et al. 2023: "... How Well Do LLMs Speak Turtle?"

[4] Frey et al. 2025: "Assessing the Evolution of LLM capabilities for KGE in 2023"

[5] Meyer et al. 2024: "Assessing SPARQL capabilities of Large Language Models"

[6] Meyer et al. 2025: "LLM-KG-Bench 3"

[7] Heim et al. 2025: "How do Scaling Laws Apply to Knowledge Graph Engineering Tasks? ..."

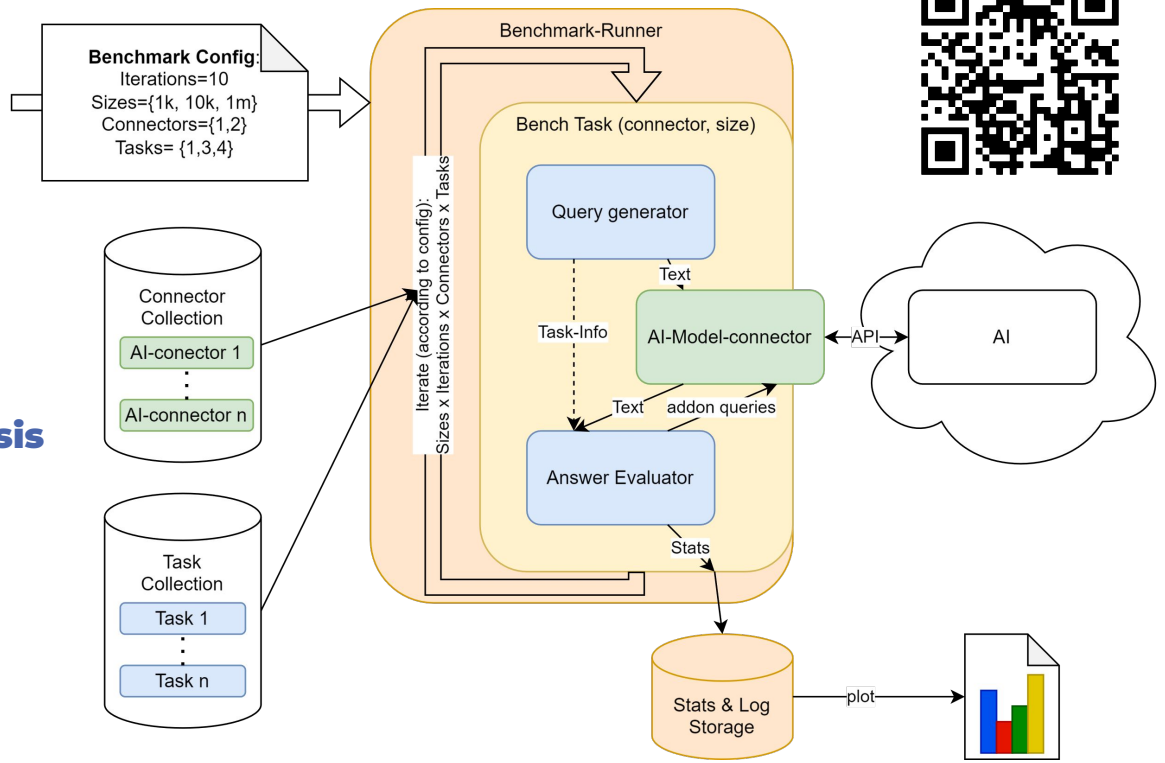
[8] Meyer et al. 2025: "Evaluating LLMs for RDF Knowledge Graph Related Tasks..."

LLM Benchmarking for KGs

InfAI
Institut für Angewandte Informatik

https://github.com/AKSW/LLM-KG-Bench

- 2023: ChatGPT 3.5, ChatGPT 4
  - Manual Evaluation [1]
- 2023: More Models, More Families (Claude, Llama, …)
  - LLM-KG-Bench v1.0 [2], v1.1 [3]
- 2024: SPARQL / Dialog / Evolution
  - LLM-KG-Bench v1.2 [4], v2.0 [5]
- **2025: new API, new Tasks, new Models, Encryption**
  - **LLM-KG-Bench v3.0 [6,7,8]**
  - **Automated, Modular, Config, Logging, Reeval., Analysis**
  - **Dataset: 40 LLMs, >7 Tasks, 20-50 Iterations**



[1] Meyer et al. 2024: "LLM-assisted Knowledge Graph Engineering ...."

[2] Meyer et al. 2023: "Developing a Scalable Benchmark ...."

[3] Frey et al. 2023: "... How Well Do LLMs Speak Turtle?"

[4] Frey et al. 2025: "Assessing the Evolution of LLM capabilities for KGE in 2023"

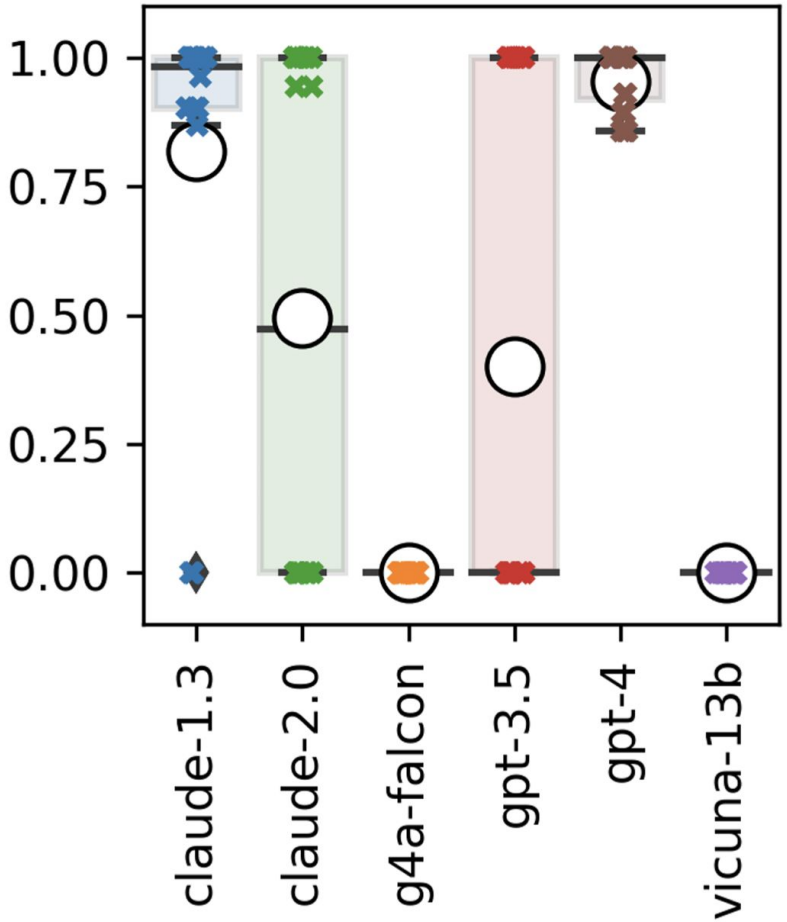
[5] Meyer et al. 2024: "Assessing SPARQL capabilities of Large Language Models"

[6] Meyer et al. 2025: "LLM-KG-Bench 3"

[7] Heim et al. 2025: "How do Scaling Laws Apply to Knowledge Graph Engineering Tasks? ..."

[8] Meyer et al. 2025: "Evaluating LLMs for RDF Knowledge Graph Related Tasks..."

RDF Syntax Fixing



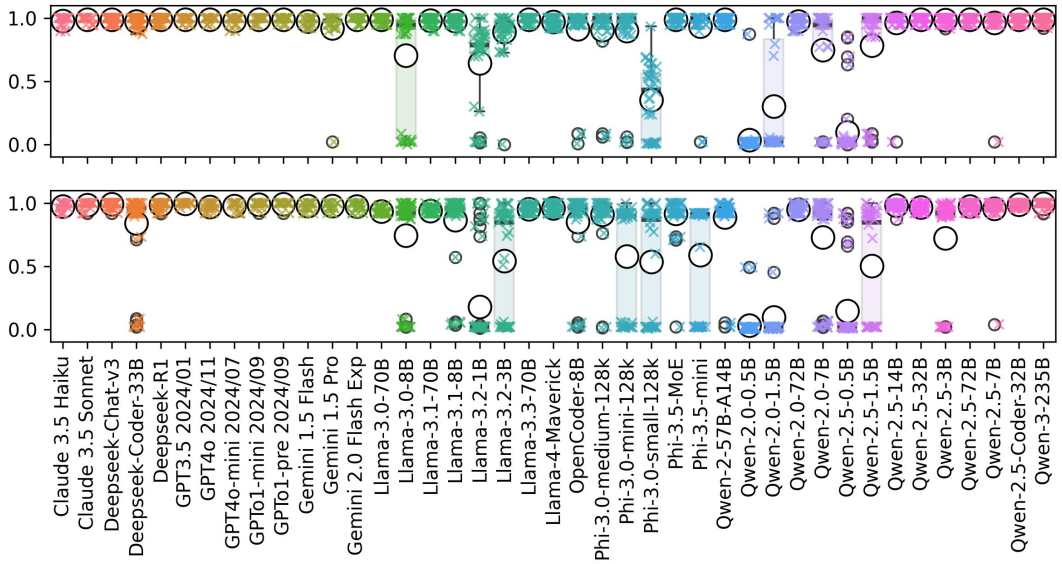
- First LLM-KG Bench Version from July 2023
- for Turtle format

[3] Frey et al. 2023: "… How Well Do LLMs Speak Turtle?"

RDF Syntax Fixing

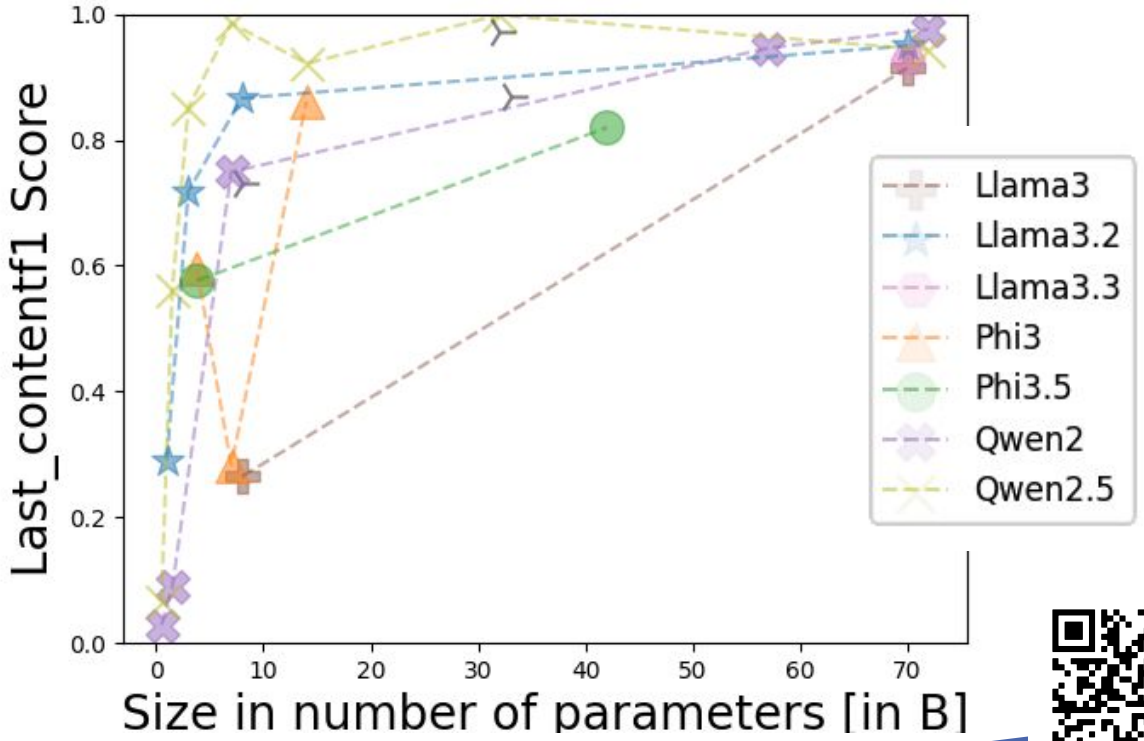
InfAI
Institut für Angewandte Informatik

RdfSyntaxFix JSON-LD



RdfSyntaxFix Turtle

Task RdfSyntaxFixList



[6] Meyer et al. 2025: “LLM-KG-Bench 3”
[7] Heim et al. 2025: “How do Scaling Laws Apply to Knowledge Graph Engineering Tasks? …”
[8] Meyer et al. 2025: “Evaluating LLMs for RDF Knowledge Graph Related Tasks…”

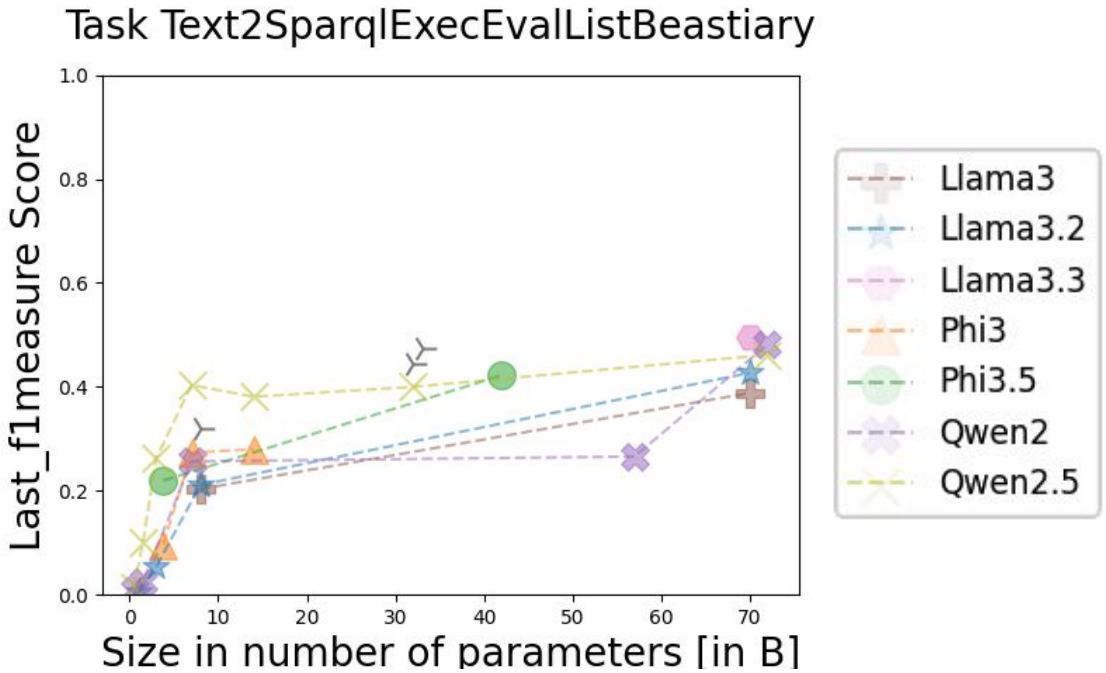
https://github.com/AKSW/LLM-KG-Bench-Results

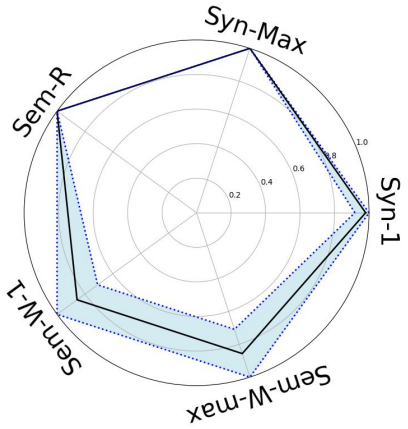
Statistics
Turtle vs JSON-LD

	RdfConnectionExplainAllTasks	RdfFriendStatic	RdfFriendCount-1	RdfFriendCount-2	RdfSyntaxFixList	Sparql2AnswerListOrga	Text2AnswerListOrga
Claude 3.5 Haiku	**TTL**	**JSON**	**TTL**	**TTL**	-	-	-
Claude 3.5 Sonnet	-	-	-	-	-	-	-
Deepseek-Coder-33B	JSON	-	**JSON**	**JSON**	**JSON**	-	-
GPT3.5 2024/01	-	-	JSON	**JSON**	-	-	**TTL**
GPT4o 2024/11	-	-	-	-	JSON	-	-
GPT4o-mini 2024/07	**TTL**	-	-	TTL	-	-	**TTL**
GPTo1-mini 2024/09	-	-	-	-	-	-	-
GPTo1-pre 2024/09	-	-	-	-	-	-	-
Gemini 1.5 Flash	**TTL**	-	**TTL**	-	-	-	-
Gemini 1.5 Pro	-	-	-	-	-	-	-
Gemini 2.0 Flash Exp	-	-	-	-	-	-	-
Llama-3.3-70B	**JSON**	-	**JSON**	**JSON**	**JSON**	-	-
Meta-Llama-3-70B	**JSON**	-	**JSON**	**JSON**	**JSON**	-	-
Meta-Llama-3-8B	-	**TTL**	-	-	-	-	-
Meta-Llama-3.1-70B	-	-	-	-	**JSON**	**TTL**	-
Meta-Llama-3.1-8B	**JSON**	-	**JSON**	**JSON**	JSON	-	-
Meta-Llama-3.2-1B	**JSON**	-	-	-	**JSON**	-	-
Meta-Llama-3.2-3B	-	**JSON**	**TTL**	**TTL**	**JSON**	-	-
deepseek-chat-v3-0324	-	**JSON**	**TTL**	**TTL**	-	-	-
deepseek-r1	-	-	-	-	-	-	-
llama-4-maverick	**JSON**	-	**JSON**	**JSON**	-	-	-
OpenCoder-8B	-	-	-	-	-	**TTL**	-
Phi-3-medium-128k	**TTL**	**TTL**	-	JSON	-	-	TTL
Phi-3-mini-128k	-	-	-	**JSON**	**JSON**	-	-
Phi-3-small-128k	-	**JSON**	-	-	TTL	-	-
Phi-3.5-MoE	**JSON**	-	**JSON**	**JSON**	**JSON**	**TTL**	-
Phi-3.5-mini	-	**TTL**	JSON	-	**JSON**	**TTL**	-
Qwen2-0.5B	-	-	-	JSON	-	-	-
Qwen2-1.5B	**JSON**	**TTL**	**JSON**	**JSON**	**JSON**	-	-
Qwen2-57B-A14B	TTL	-	-	-	**JSON**	TTL	**TTL**
Qwen2-72B	-	-	-	-	**JSON**	-	-
Qwen2-7B	-	**JSON**	-	-	-	-	TTL
Qwen2.5-0.5B	-	-	-	-	-	-	-
Qwen2.5-1.5B	-	-	-	-	**JSON**	-	-
Qwen2.5-14B	-	**TTL**	-	-	-	-	TTL
Qwen2.5-32B	TTL	-	**TTL**	**TTL**	**JSON**	**JSON**	-
Qwen2.5-3B	-	**JSON**	-	-	**JSON**	-	-
Qwen2.5-72B	-	**JSON**	JSON	**JSON**	-	-	**TTL**
Qwen2.5-7B	**TTL**	**JSON**	-	-	-	**TTL**	**TTL**
Qwen2.5-Coder-32B	**TTL**	-	**TTL**	**TTL**	-	-	-
Solar-pro-preview-22B	-	-	-	-	JSON	-	-
All Models		-	**JSON**	**JSON**	**JSON**	TTL	**TTL**

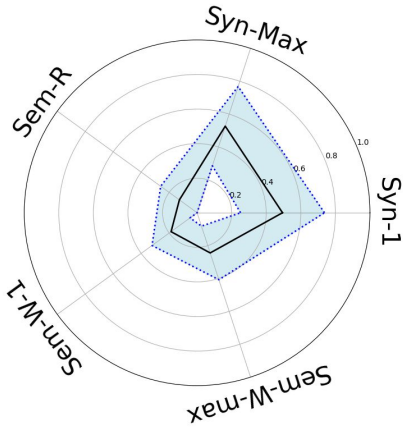
https://github.com/AKSW/LLM-KG-Bench-Results

SPARQL Capabilities





Claude 3.5 Sonnet



Qwen-2.5 (1.5B)

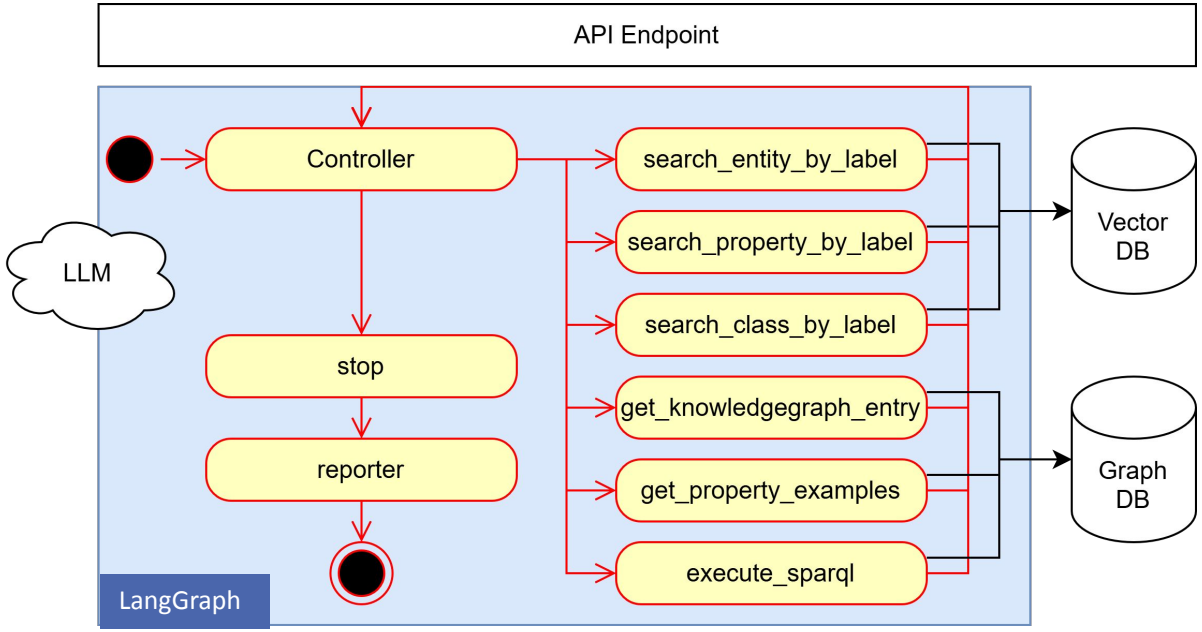
[6] Meyer et al. 2025: “LLM-KG-Bench 3”
[7] Heim et al. 2025: “How do Scaling Laws Apply to Knowledge Graph Engineering Tasks? …”
[8] Meyer et al. 2025: “Evaluating LLMs for RDF Knowledge Graph Related Tasks…”
[9] Kovriguina et al. 2023: “PARQLGEN: one-shot prompt-based approach for SPARQL query generation”

https://github.com/AKSW/LLM-KG-Bench-Results

ReAct for Text2Sparql

InfAI
Institut für Angewandte Informatik

- Text2Sparql-Challenge @ESWC 2025 [7]
- 8 Teams
- Best Teams with GPT and Agentic AI
- InfAI best in DBpedia-EN & Corporate
- More Teams: WSE and IIS

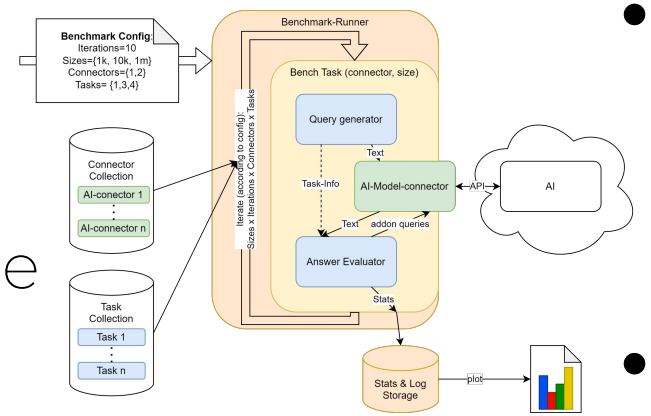


[7] https://text2sparql.aksw.org/
Proceedings to appear soon

Summary

InfAI
Institut für Angewandte Informatik

- Results from LLM-KG-Bench:
  - RDF Syntax: ✓
  - JSON-LD vs. Turtle
  - SPARQL: ?
  - Open LLMs: ✓



https://github.com/AKSW/LLM-KG-Bench

- Authors for this results:
  Nathanael Arndt & Felix Brei & Kirill Bulert & Lorentz Bühmann & Andreas Dengel & Sabine Gründer-Fahrer & Johannes Frey & Daniel Gerber & Desiree Heim & Kurt Junghanns & Michael Martin & Lars-Peter Meyer & Markus Schröder & Claus Stadler & Sara Todorovikj
- Contact:
  { LPMeyer | Frey }@infai.org
- Datasets & Results:
  https://github.com/AKSW/LLM-KG-Bench-Result

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung
FKz.: 16DTM312D

Bundesministerium für Wirtschaft und Klimaschutz
FKZ: 01MK22001A
FKz.: 01MK21007A
KISS & CoyPu

Bundesministerium für Verkehr
FKZ: 19F2266A
MobyDex