Fraunhofer IIS

Fraunhofer Institute for Integrated Circuits IIS

RAGONITE: Iterative Retrieval on Induced Databases and Verbalized RDF for Conversational QA over KGs with RAG

Rishiraj Saha Roy, Chris Hinze, Joel Schlotthauer, Farzad Naderi, Viktor Hangya, Andreas Foltyn, Luzian Hahn, and Fabian Küch
NLP Team
Fraunhofer IIS, Erlangen

Conversational question answering (ConvQA)

What is the setting for our work?

- **ConvQA** natural way of information access
  - Self-contained **intent-explicit** first question(s)
  - Potentially **intent-implicit** follow-up questions
- Primarily conducted via **LLMs**
- Several enterprises want to conduct ConvQA over **own data**
- Retrieval augmented generation (RAG) systems are the **norm**
  - Retrieve data from enterprise knowledge using **tools**
  - LLMs generate coherent response from retrieved **evidence**
- Enterprise data usually stored as
  - Databases, Text corpora, Heterogeneous collections
  - **Knowledge graphs – focus of this work**

Q1: What is the cheapest X1 series car? How much is it?

A1: The cheapest BMW X1 series car is the "BMW X1 sDrive20i Sport" (35,410 EUR)

Q2: Just curious - and what's the most expensive X7?

A2: The costliest BMW X7 is the "BMW X7 xDrive40d Excellence" (89,515 EUR).

Q3: Comparison of their engines?

A3: The key differences between the engines of BMW X1 and BMW X7 are ...

Q4: So what are some of these luxury features in the latter?

A4: The BMW X7 comes with a range of luxury features, including ...

Q5: Luxury comes at a cost. Are CO2 emissions of X7 much worse?

A5: Yes, the CO2 emissions of the BMW X7 models are generally higher than X1...

Fraunhofer IIS

Knowledge graphs (also known as Knowledge Bases)

What is the data source?

- Contain **factual information** comprised of
  - **Entities**
  - **Predicates**
  - **Types**
  - **Literals**
- **RDF** standard: **facts** stored as **triples**
- Triples have “binary” format **<Subject, Predicate, Object>**
- More complex info can be stored as **n-ary facts**

RDF Turtle format

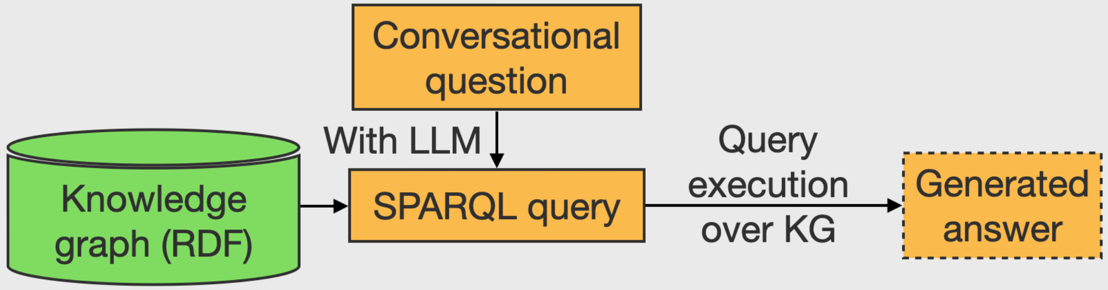
```
<bmw-x7-xdrive40i,          type, car;
bmw-x7-xdrive40i            engine, bmw-x7-xdrive40i-excellence;
bmw-x7-xdrive40i            height, 1835 mm;
bmw-x7-xdrive40i            length, 5181 mm;
bmw-x7-xdrive40i            price, 88890 Euros;
bmw-x7-xdrive40i            wheelbase, 3105 mm;
bmw-x7-xdrive40i            width, 2000 mm>

<bmw-x7-xdrive40i-excellence, type, engine specification;
                            WLTP CO2 emission combined, 240 – 218 g/km;
                            WLTP consumption combined, 26.6 – 29.4 mpg;
                            acceleration 0-62 mph, 5.8 s;
                            drive type, all-wheel drive;
                            engine performance, 280 kW (381 hp);
                            fuel type, gasoline;
                            transmission,  Sport Automatic Transmission with
                                           Gearshift Paddles>
```


Knowledge graphs (also known as Knowledge Bases)

Why are KGs typically used?

- **Why** KGs?
  - **Flexible schema** compared to DBs
  - Reduced **moderation effort** for human experts
  - Exploit **neighborhoods** of entities: answer, recommend, visualize
- Nature of info in KGs limits their use to **lookups**
  - Via **SPARQL**(-like) languages

Traditional KG-QA: Based on SPARQL



<bmw-x7-xdrive40i, type, car;
engine, bmw-x7-xdrive40i-excellence;
height, 1835 mm;
length, 5181 mm;
price, 88890 Euros;
wheelbase, 3105 mm;
width, 2000 mm>

<bmw-x7-xdrive40i-excellence, type, engine specification;
WLTP CO2 emission combined, 240 – 218 g/km;
WLTP consumption combined, 26.6 – 29.4 mpg;
acceleration 0-62 mph, 5.8 s;
drive type, all-wheel drive;
engine performance, 280 kW (381 hp);
fuel type, gasoline;
transmission, Sport Automatic Transmission with Gearshift Paddles>

How can we satisfy new intent classes (abstract, complex math) over KGs?

New possibilities open up with capable LLMs

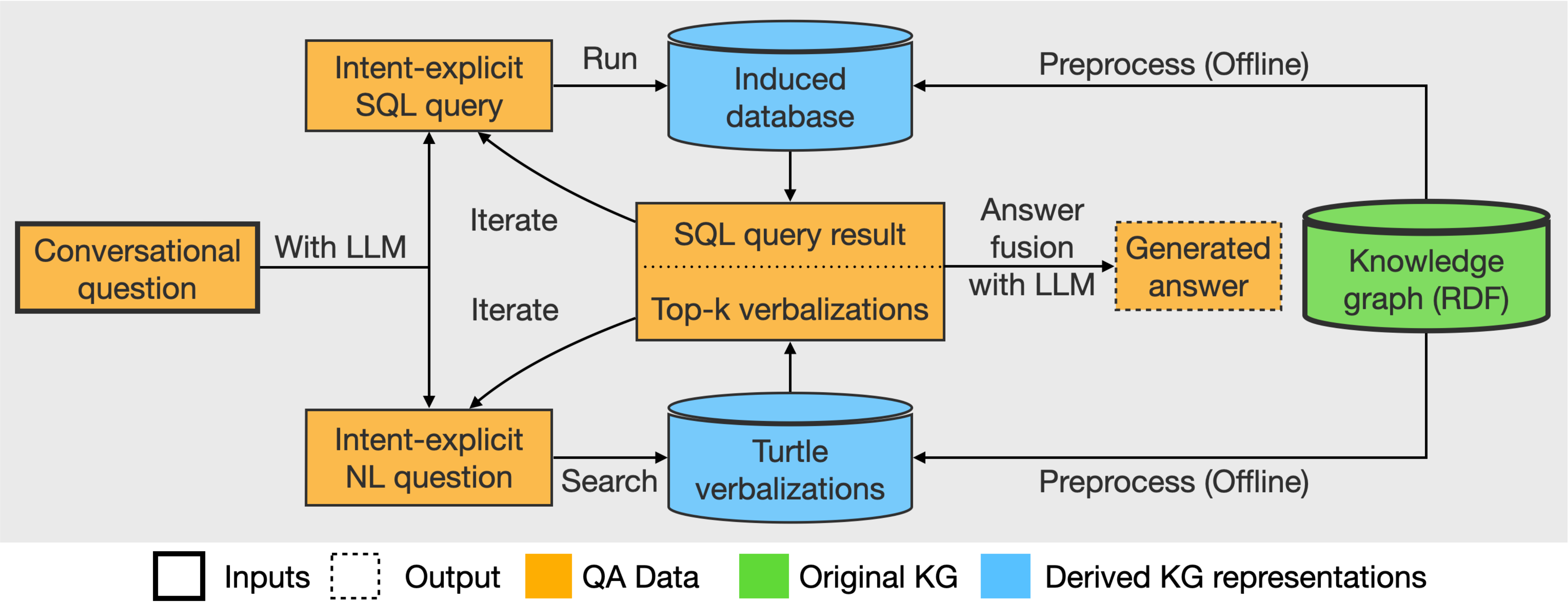
- **World knowledge** in LLM parameters
  - Units, acronyms, region-based standards
- **Commonsense** reasoning
  - Equipment and luxury features, suitability as family cars
- **Structured query** generation
  - SQL (better than SPARQL), ...
- **Structured output** generation
  - JSON for interfacing pipeline components

<bmw-x7-xdrive40i, type, car;
engine, bmw-x7-xdrive40i-excellence;
height, 1835 mm;
length, 5181 mm;
price, 88890 Euros;
wheelbase, 3105 mm;
width, 2000 mm>

<bmw-x7-xdrive40i-excellence, type, engine specification;
WLTP CO2 emission combined, 240 – 218 g/km;
WLTP consumption combined, 26.6 – 29.4 mpg;
acceleration 0-62 mph, 5.8 s;
drive type, all-wheel drive;
engine performance, 280 kW (381 hp);
fuel type, gasoline;
transmission, Sport Automatic Transmission with Gearshift Paddles>

Key idea: Make the most of your knowledge graph with RAGONITE!

A unified framework for satisfying **lookup**, **abstract,** and **complex math** intents over KGs



Highlights in RAGONITE

What are the key features of our novel KG-QA pipeline?

- Automated database induction
- Knowledge graph verbalization
- Iterative retrieval
- Branch integration
- Open LLM support
- Scope for heterogeneous QA

Automated database induction

How can we programmatically convert a KG to a DB?

- Original KG in RDF **Turtle** format
- Converted to a **DBMS** for harnessing SQL capabilities (Li et al. 2023)
  - More capable than SPARQL at **complex math intents**
  - **LLMs better at SQL** generation than SPARQL
  - Failed SQL queries often generate useful, directed **error** messages
- Key steps
  - Inducing **schema** (types -> tables, predicates -> columns)
  - Inserting **data** (facts -> rows, inferring datatypes)
  - Enhancing **semantics** (renaming columns, SQL comments in schema)

```
<bmw-x7-xdrive40i,      type, car;
                        engine, bmw-x7-xdrive40i-excellence;
                        height, 1835 mm;
                        length, 5181 mm;
                        price, 88890 Euros;
                        wheelbase, 3105 mm;
                        width, 2000 mm>

<bmw-x7-xdrive40i-excellence, type, engine specification;
                        WLTP CO2 emission combined, 240 – 218 g/km;
                        WLTP consumption combined, 26.6 – 29.4 mpg;
                        acceleration 0-62 mph, 5.8 s;
                        drive type, all-wheel drive;
                        engine performance, 280 kW (381 hp);
                        fuel type, gasoline;
                        transmission,  Sport Automatic Transmission with
                                       Gearshift Paddles>
```

car

engine	height	length	price	wheelbase	width
bmw-x7-xdrive40i-excellence	1835 mm	5181 mm	88890 Euros	3105 mm	2000 mm

engine

WLTP CO2 emission combined	WLTP consumption combined	acceleration 0-62 mph	drive type	engine performance	...
240 – 218 g/km	26.6 – 29.4 mpg	5.8 s	all-wheel drive	280 kW (381 hp)	...

Knowledge graph verbalization

How can we make a KG amenable for NL retrieval/reasoning?

- Makes KG suitable for satisfying **abstract** intents
- **Text search** runs over
  - **Passage verbalizations** generated from KG (Oguz et al. 2022)
  - Uses KG in **RDF Turtle** format
- Each RDF **Turtle "shell"** converted into an NL **passage**
- Simple and generalizable **hand-crafted patterns**
- **Reversed facts** also included to increase scope of answerable questions
  - Helps simpler LLMs

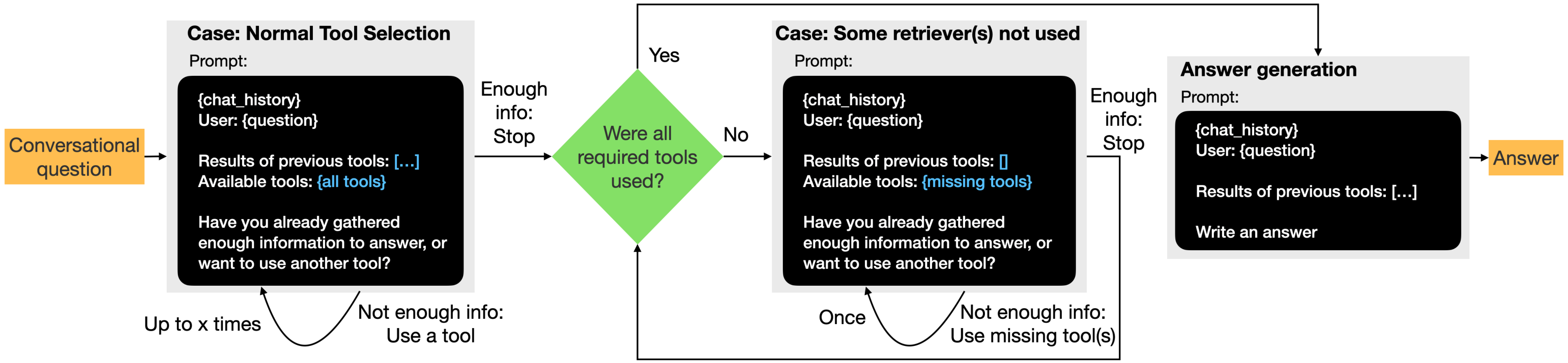
<bmw-x7-xdrive40i, type, car;
engine, bmw-x7-xdrive40i-excellence;
height, 1835 mm;
length, 5181 mm;
price, 88890 Euros;
wheelbase, 3105 mm;
width, 2000 mm>

BMW X7 xDrive 40i is car. BMW X7 xDrive 40i has engine bmw-x7-xdrive40i-excellence. bmw-x7-xdrive40i-excellence is engine of BMW X7 xDrive 40i. BMW X7 xDrive 40i has height 1835 mm. 1835 mm is height of BMW X7 xDrive 40i. BMW X7 xDrive 40i has length 5181 mm. 5181 mm is length of BMW X7 xDrive 40i. BMW X7 xDrive 40i has price 88890 Euros. 88890 Euros is price of BMW X7 xDrive 40i. BMW X7 xDrive 40i has wheelbase 3105 mm. 3105 mm is wheelbase of BMW X7 xDrive 40i. BMW X7 xDrive 40i has width 2000 mm. 2000 mm is width of BMW X7 xDrive 40i.

Iterative retrieval with tools

Why is RAGONITE a tool-based pipeline?

- RAGONITE has multiple **tools or functions** at its disposal
  - **SQL** query executor
  - **Text** retriever (dense and lexical retrievers)
- Iterative retrieval strategy
  - Each tool can be called **multiple** times until satisfactory results are retrieved
  - Each tool must be used **at least once**
- Iterative retrieval benefits
  - Satisfies **multi-hop intents** (*Average price gap of Touring models from an X5?*)
  - Increases **robustness** to "failed" retrievals: previous results/errors included in rephraser LLM's prompt for next iteration



Fraunhofer IIS

Branch integration

How do we fuse results from the branches?

- **Merging** results from structured and unstructured retrieval necessary
- **Precludes** need for
  - Deciding based on **question** which branch(es) to prefer
  - Deciding based on retrieval **results** which branch(es) to prefer
- **No hard-coded rules** in system
- **Fully automated** via LLM based on input question and conversation so far

Open LLM integration

What LLMs do we support?

- Uses **GPT-4o** via API as default
- Seamless support for **Llama** and other local models

Heterogeneous QA

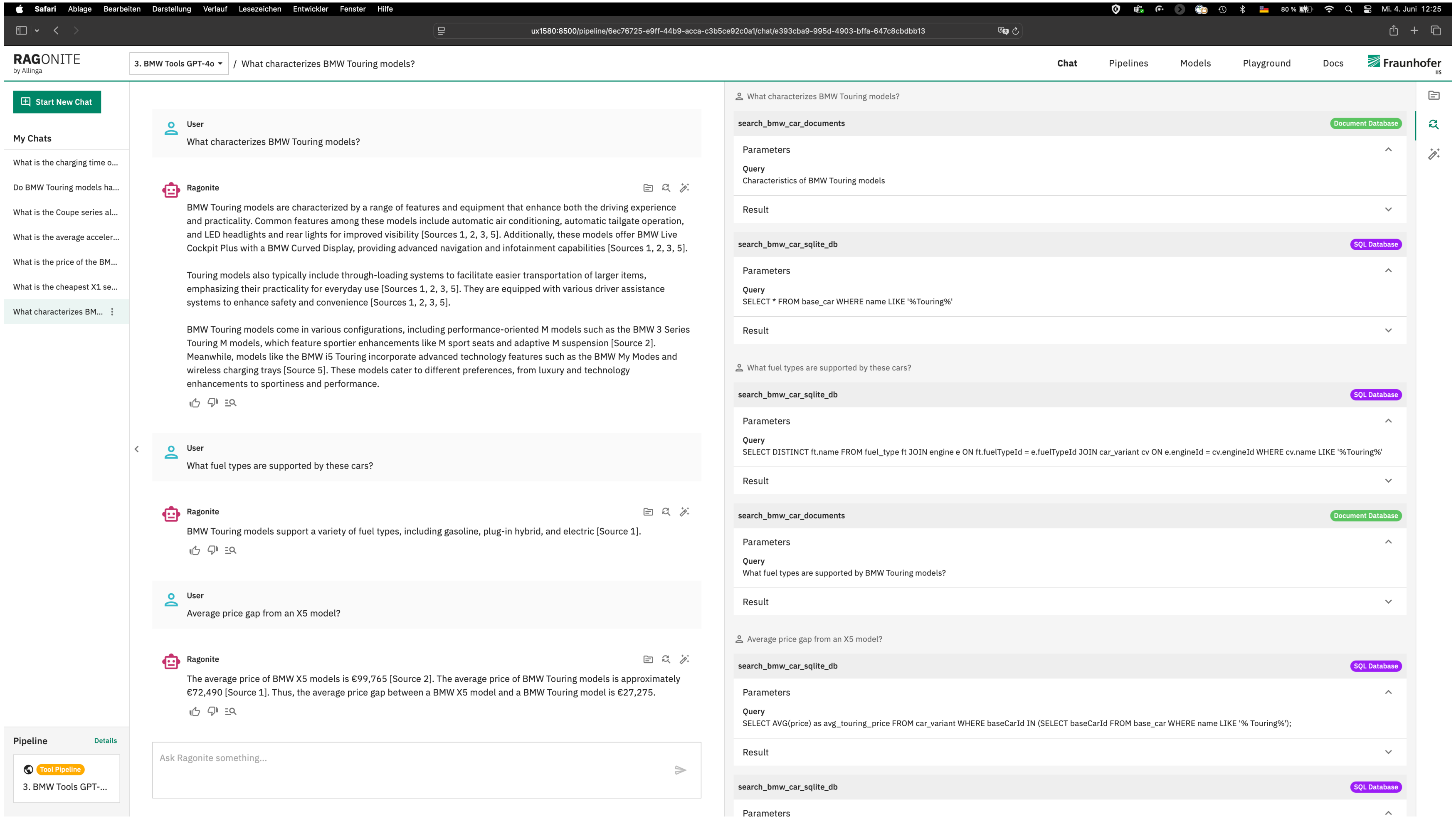
How can we use RAGONITE for ConvQA over KG and text combinations?

- Having RDF verbalizations in pipeline facilitates incorporation of **additional text**
- **No special support** required
- Additional text snippets, for example, from the Web, can be simply **added to our corpus**
- Integration **works well** with about 400 passages from bmw.co.uk

Fraunhofer IIS

Time for the demo!

How does our RAGONITE implementation look like?



Evaluation

How do we measure RAGONITE performance?

- **BMW KG**
  - 3442 facts: 466 entities, 27 predicates, 7 types, 1295 literals
  - 7 tables via DB induction, 466 passages via RDF verbalization
- **QA pairs**
  - 30 diverse questions wrt entities, predicates, types
  - Conducted as six conversations with five turns each
  - Equal distribution of all three intent classes: lookup, complex, abstract
- Four **variants** compared
  - SPARQL-only: 4/30 ✅ (fails in most cases due to LLM's inability at good query generation, NERD failures)
  - SQL-only: 18/30 ✅ (fails when question has ad hoc entity mentions like *gran coupe sport 220 i m* and not bmw-220i-m-sport-gran-coupé)
  - Verbalization-only: 24/30 ✅ (fails with complex math, like minimum over averages of grouped car prices)
  - SQL+Verbalization: 28/30 ✅
- Average runtime is **6 seconds** per question, end-to-end
- **Detailed** runtimes and accuracies available at https://github.com/Fraunhofer-IIS/RAGonite/blob/main/btw25-bmw/data/bmw-eval-results.pdf

Summary and outlook

What can one take away from this presentation?

Thank you! Check out our repo at https://github.com/Fraunhofer-IIS/RAGonite

- Novel **two-pronged pipeline for ConvKG-QA** with iterative retrieval
- RAGONITE is a **transparent RAG** pipeline – every step is scrutable by end-user
- Focus on **automation** but **judicious** amounts of **manual** effort goes a long way
- Exploits **Text2SQL** for KG-QA
- Information in knowledge graphs is useful for **abstract** intents
- RAGONITE is an **agentic** system
  - **Reflection** mechanisms may add value
  - Without compromising **efficiency**

R.S. Roy, C. Hinze, J. Schlotthauer, F. Naderi, V. Hangya, A. Foltyn, L. Hahn, and F. Küch. **RAGONITE: Iterative Retrieval on Induced Databases and Verbalized RDF for Conversational QA over KGs with RAG**, *in Proceedings of the 21st Conference on Database Systems for Business, Technology and Web (BTW '25)*, pages 787 – 794, Bamberg, Germany, 3 – 7 March 2025.

R.S. Roy, C. Hinze, J. Schlotthauer, A. Foltyn, L. Hahn, and F. Küch. **Evidence contextualization and counterfactual attribution for conversational QA over heterogeneous data with RAG systems,** *in Proceedings of the 18th ACM International Conference on Web Search and Data Mining (WSDM '25),* pages 1040 – 1043, Hannover, Germany, 10 – 14 March 2025.