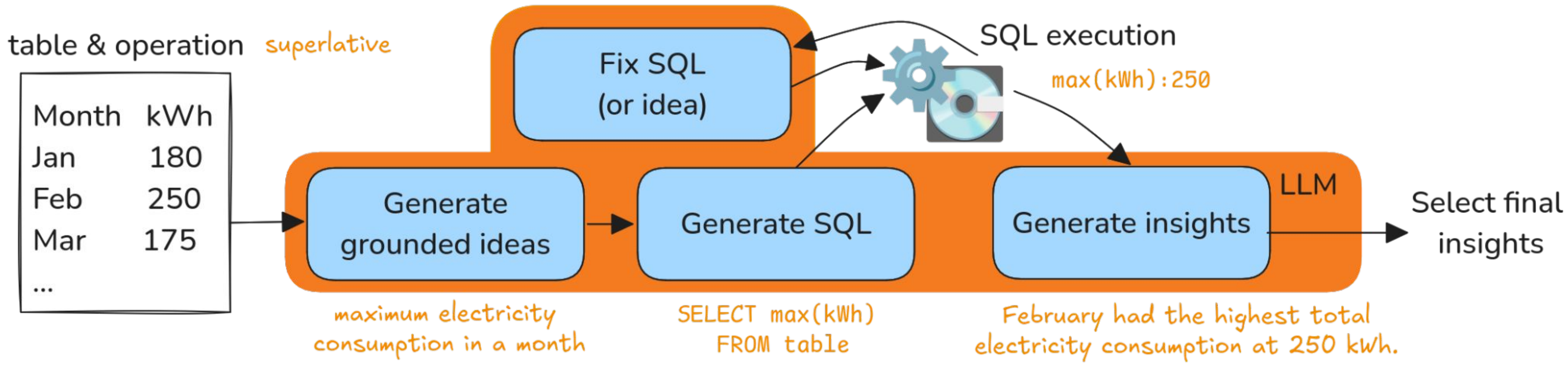
Always use fresh data to evaluate your LLMs! SQL helps analyze tables, but they must be clean.

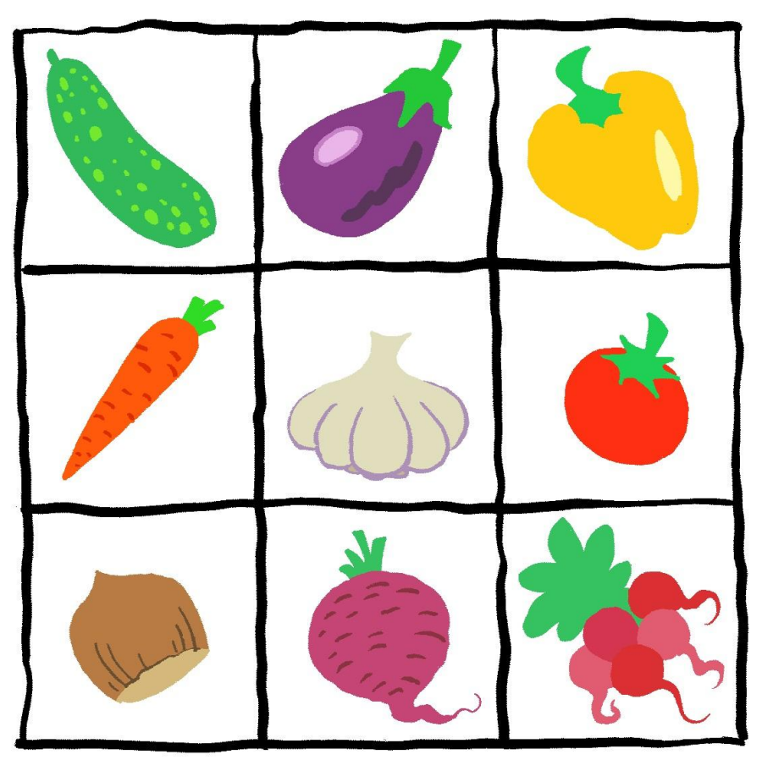
Generating Data Insights with LLMs by Querying Tables

K. Onderková, O. Plátek, Z. Kasner, M. Lango, O. Dušek
[onderkova, oplatek, kasner, lango, odusek]@ufal.mff.cuni.cz*

Charles University
ÚFAL



FreshTab



Problem

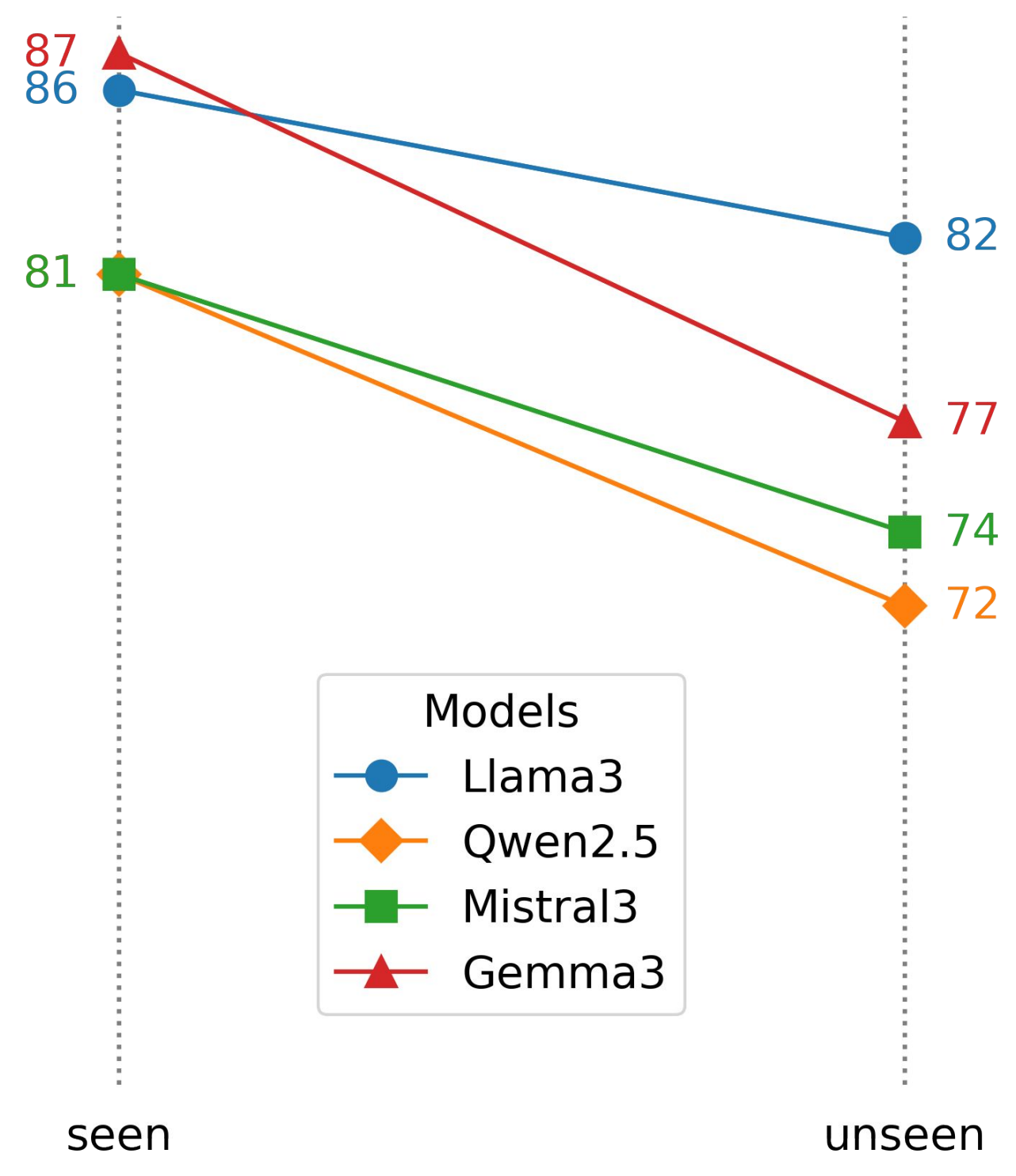
- LLMs **memorized** existing **datasets** (LogicNLG, LoTNLG)
- Datasets contain mostly sports, **no domain** distinction
- They are usually **English only**

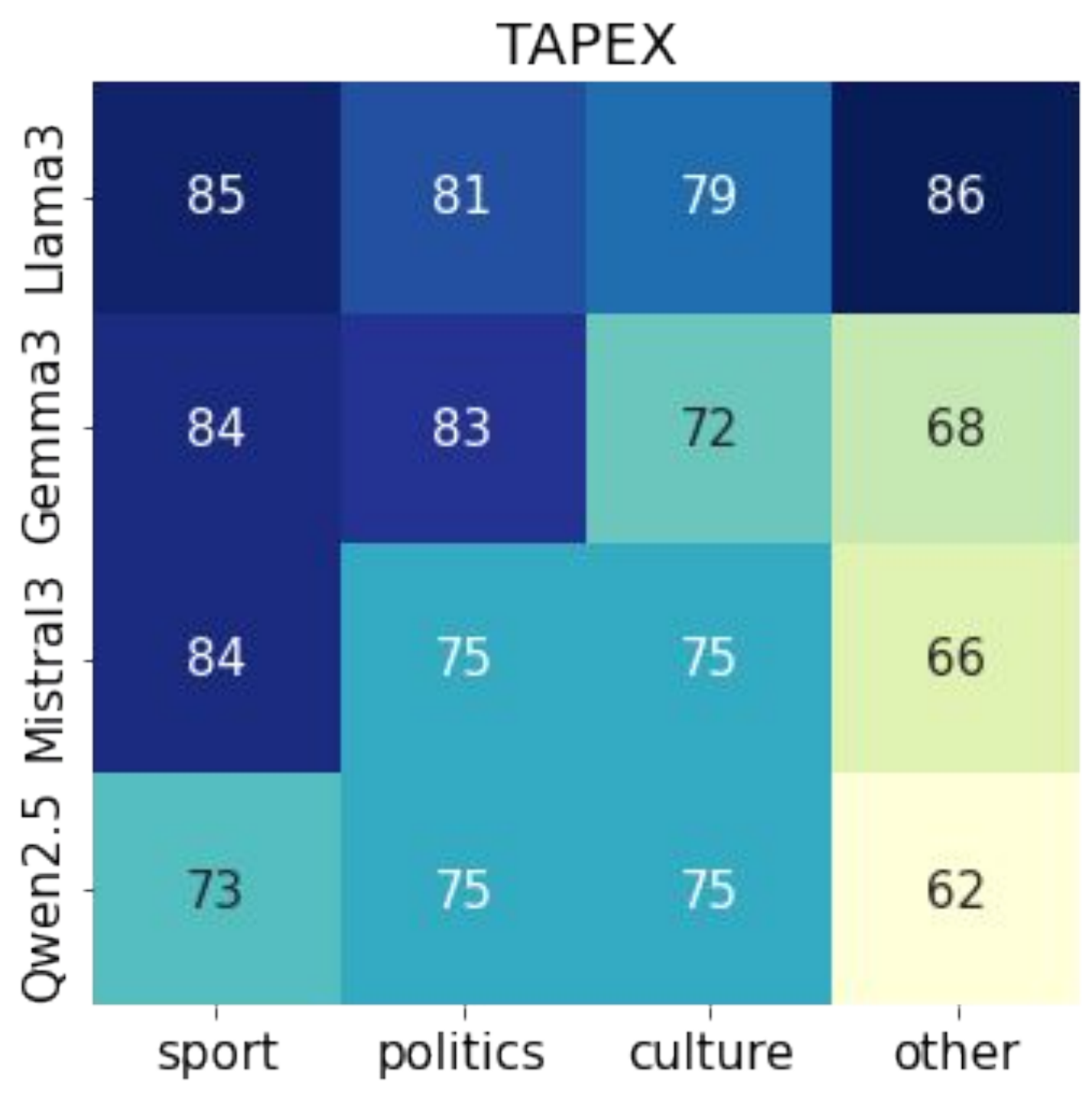
Solution

- **Dynamic dataset construction** based on creation date
  - **Query** Wikidata and **Wikipedia** for list of Wikipages
  - **Scrape**, clean and pick a **table** per Wikipage
  - Filter tables for domain balance

Results

- LLMs' **performance** visibly **decreases** with unseen data
- FreshTab can **differentiate LLMs** by their **ability to generalize**
- Domain split helps us pinpoint where the models have deficit





Observations

- Llama 3.3 70b seems to generalize best to new data
  - most balanced performance across domains
- Sport domain is easiest, as expected from lot of training data

SQuirreL

From real a human!

Problem

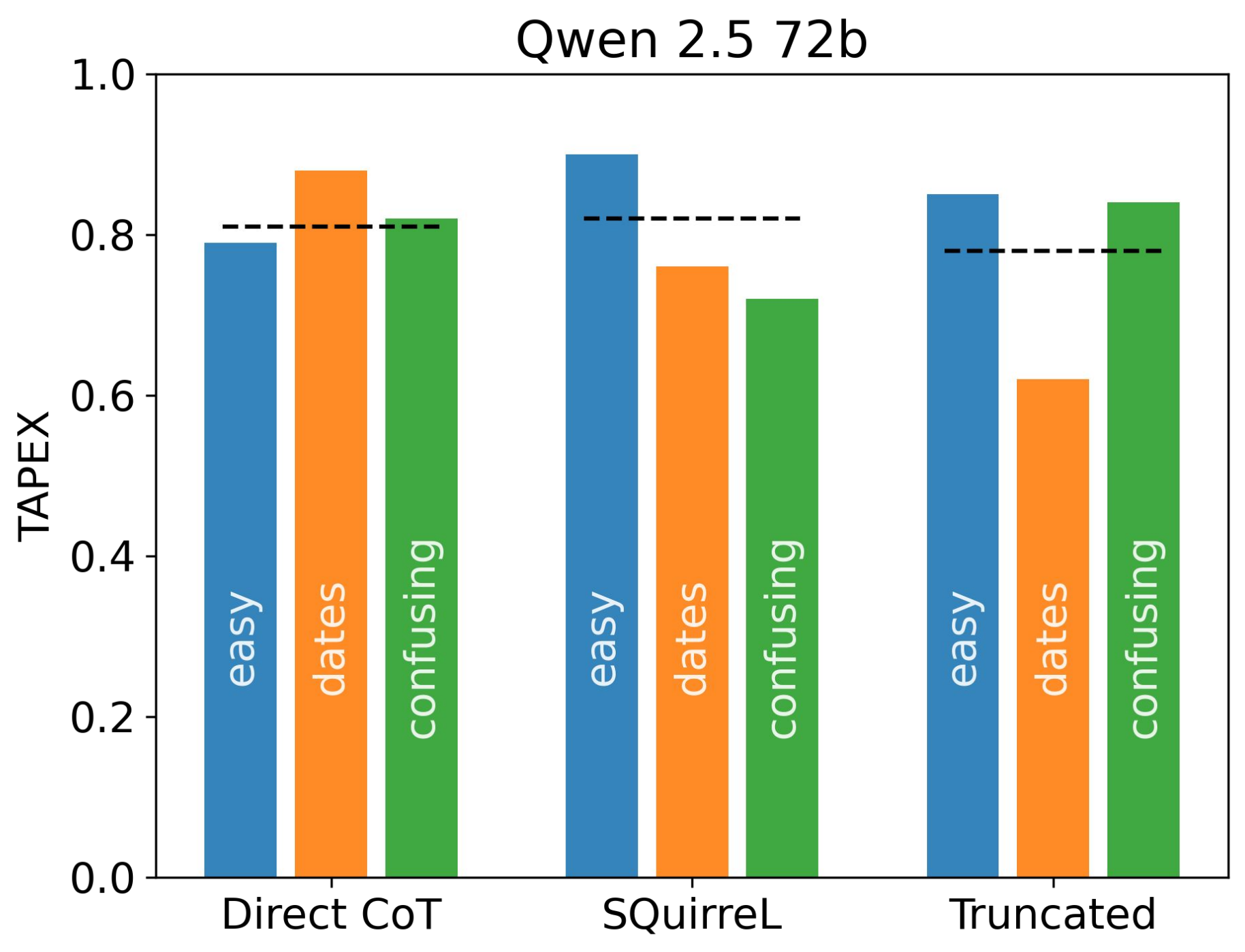
- Symbolic logical forms are **hard to scale**
- LLMs are **hallucinating** and weak at logic

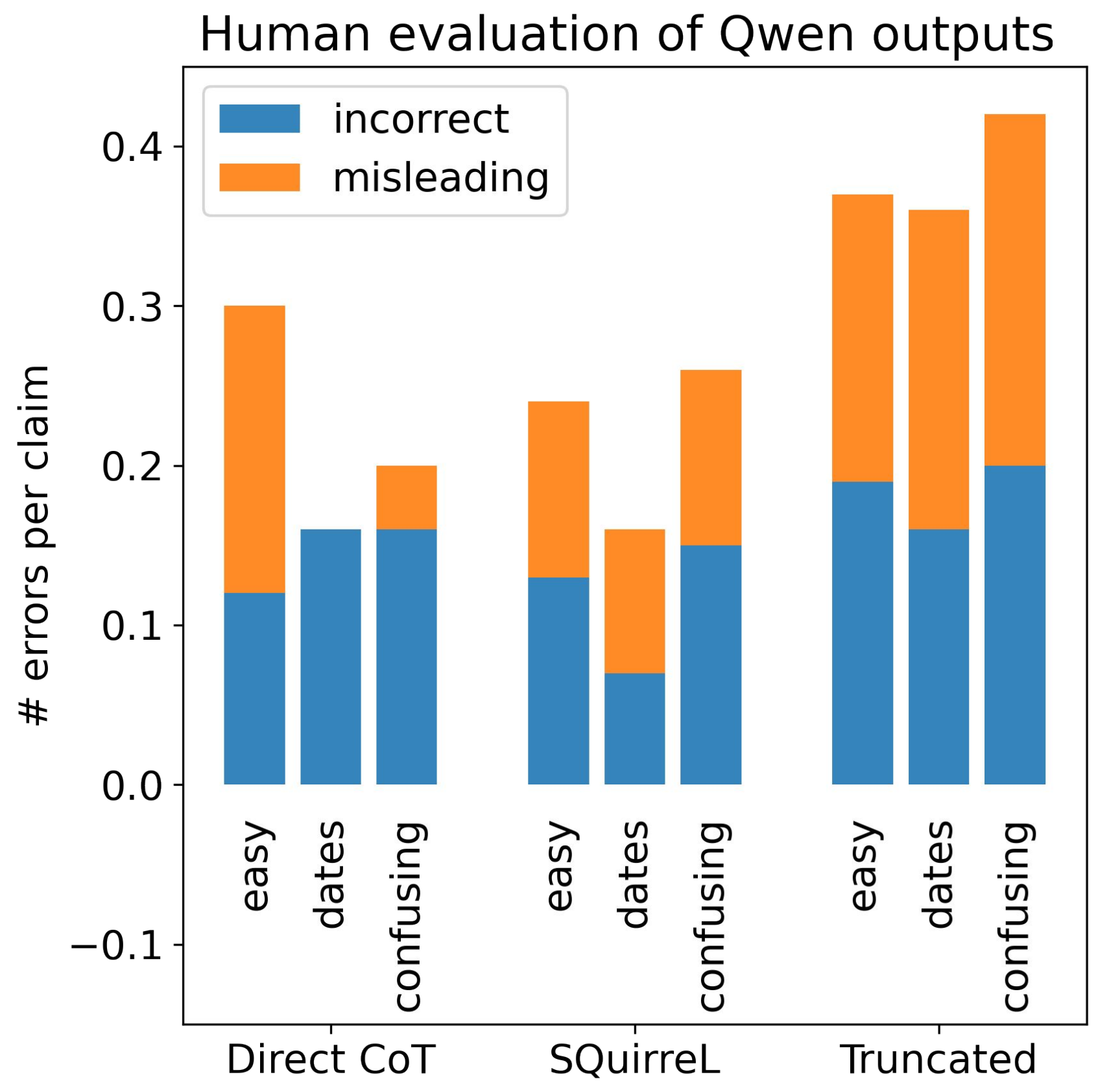
Solution

- SQuirreL: **Ground** LLMs **in data** retrieved with SQL
  - as a bonus, it is nicely **checkable**
  - can work with truncated tables **on bigger data**

Results

- Factuality does not improve. **Why**?
- Some **tables** are **not well defined** and are confusing
  - We can detect those tables
- SQL needs well-defined tables to work





Observations

unnamed	entering	advancing
First round (14 teams)	14 Second League teams	-
Last 16 (16 teams)	5 Premier League teams	11 winners
Semi-finals (4 teams)	-	8 winners

event	men
5000 metres	13:01.00
High jump	2.33
Decathlon	8550

candidate	votes
dog	60
cat	70
Total	130

Presented at Machine Learning Prague 2025.
Funded by the European Union (ERC, NG-NLG, 101039303),
Charles University projects SVV 260 698,
National Recovery Plan funded project MPO 60273/24/21300/21000 CEDMO 2.0 NPO.

https://github.com/Kristyna-Navitas/FreshTab