Beyond Accuracy

What a user study reveals about the new era of natural language data interaction

Based on the research by Panos Ipeirotis & Haotian Zheng

For decades, data has spoken a different language.

The Expert's Tool

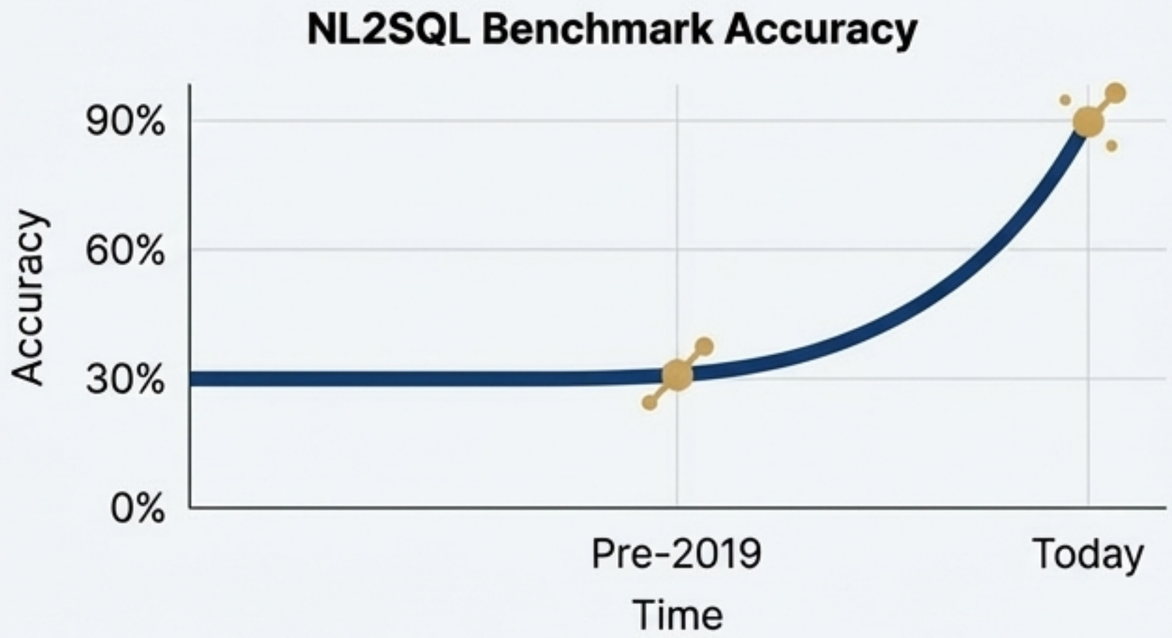
SQL is a powerful and essential language for data manipulation, but its steep learning curve creates a significant bottleneck. This high barrier to entry limits direct, widespread data access for many business analysts and decision-makers.

The Broken Promise

Early Natural Language Interfaces (NLIDBs) aimed to bridge this gap, but practical adoption was limited. The primary reason was the 'Accuracy-Usability Gap'—persistent errors, ambiguity, and a lack of transparency that eroded user trust trust and led to deep frustration.

Large Language Models changed the conversation.

The rise of powerful LLMs has dramatically improved NL2SQL accuracy on technical benchmarks, pushing success rates from ~30% before 2019 to nearly 90% on datasets like Spider.



This raises the critical question for product and analytics leaders: **Does better technical translation finally lead to a better, more effective user experience in the real world?**

We designed a controlled study to measure what truly matters.

The Contenders

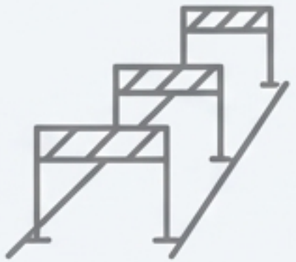
SQL-LLM: A modern NL2SQL system built on the SeekAI platform.

Snowflake: A professional, industry-standard SQL analytics platform.

The Challengers

20 experienced data professionals:

- 10 data analysts
- 5 business intelligence pros
- 5 graduate students



The Gauntlet

12 realistic business queries across 3 distinct databases (Books, Mondial Geo, Legislator).

Tasks were categorized into Easy, Medium, and Hard difficulty levels to test performance under pressure.

The results weren’t just better. They were categorically superior.

33% Faster

Completion Time

(Mean time of 418s vs. 629s for Snowflake)

50% Higher Accuracy

Task Success

(75% success rate vs. 50% for Snowflake)

Significantly Lower

Cognitive Load

(NASA-TLX workload score of 42.3 vs. 55.7)

Dramatically Higher

System Trust

(Trust rating of 6.1/7 vs. 4.0/7)

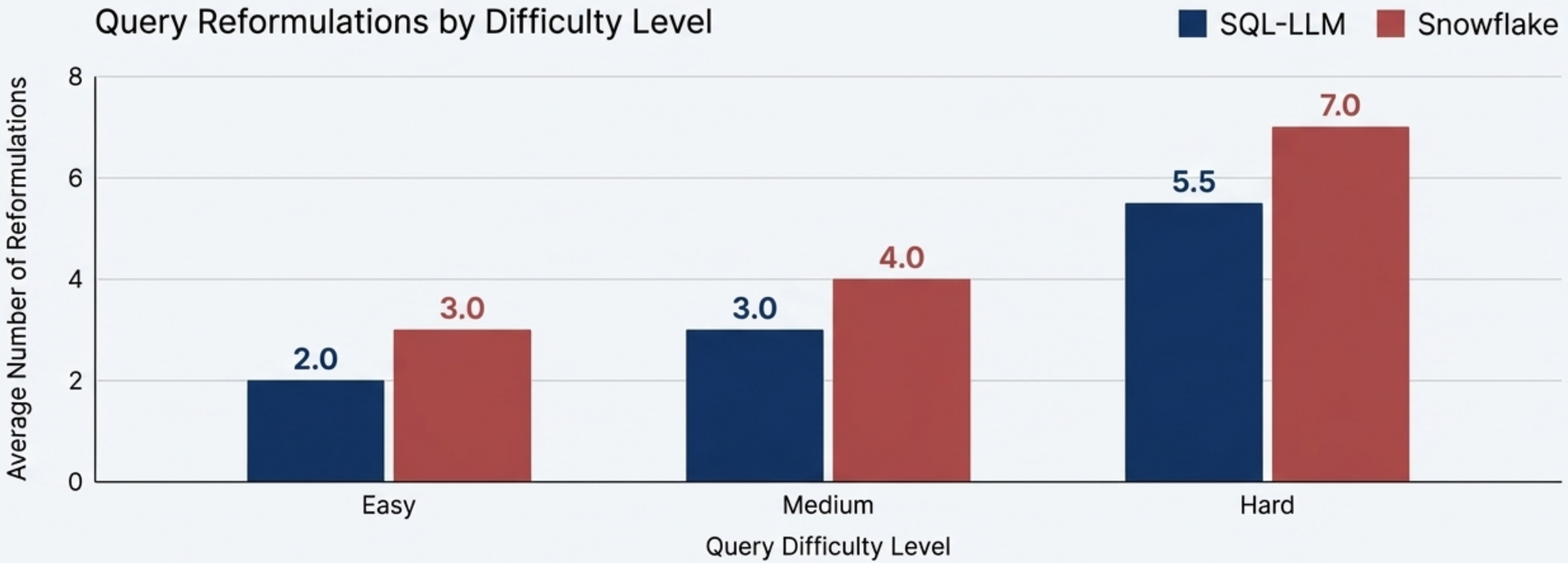
The harder the query, the bigger the advantage.

Completion Time by Difficulty Level

Success Rate by Difficulty Level

SQL-LLM's efficiency and accuracy advantage *widens* as task complexity increases, delivering the most value when it's needed most.

SQL-LLM finds the solution with less trial-and-error.



Users required fewer attempts to get the right answer, indicating the system understood their intent more accurately from the start. This directly translates to reduced user effort and a more efficient workflow.

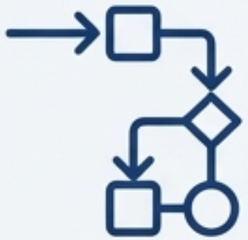
Why is the new interface so much more effective?
The answer lies in human behavior.

> “I just typed the question, and it gave me a working query. […] I’ll verify, but it looks correct.”

– Q, SQL-LLM user

The choice of tool fundamentally changed how users approached the problem.

SQL-LLM Users



Schema-First & Systematic

High initial confidence; focused on semantic verification of the generated query's logic, not its syntax.

"The SQL gave me the correct country... Let me confirm the table has inflation rate. Yes, confirmed."

– G, SQL-LLM

Snowflake Users

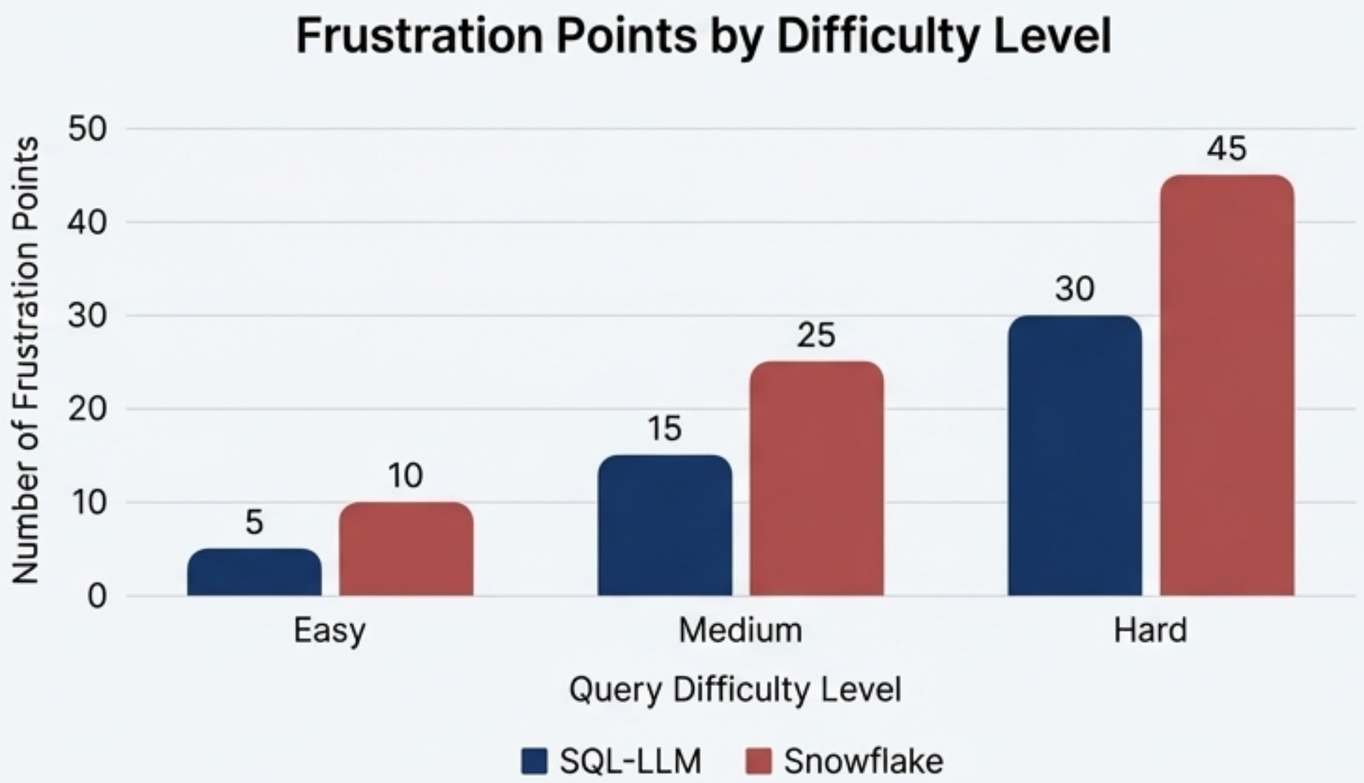
Trial-and-Error & Manual Exploration

Hesitation; frequent schema checks, manual browsing of tables, and debugging syntax errors.

"Why is someone's table empty? ... Just a million tables here."

– E, Snowflake

A better interface leads to a less stressful, more confident analyst.



Recovered from errors
30–40 seconds faster
than Snowflake users.

Cognitive Workload (NASA-TLX)

42.3 vs. **55.7**

System Trust (Rating 1-7)

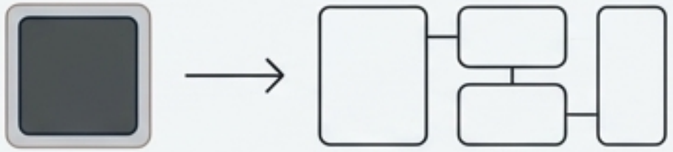
6.1 vs. **4.0**

The interface doesn't just make work faster; it makes the experience of working with data less frustrating and more empowering.

Users evolve from giving instructions to having a conversation.

From Syntax to Semantics

Users stop debugging commas and start debugging the *logic* of the question. The focus shifts from “Is this written correctly?” to “Is this the right question to ask?”



From One-Offs to Patterns

Users begin to recognize and reuse abstract linguistic structures in their prompts, indicating a form of transfer learning.

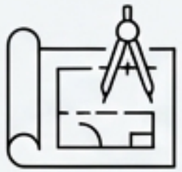
“This one’s kind of similar to the previous one. I think I can reuse that structure.”

From Fear to Exploration

With a lower penalty for errors, users actively probe the system’s boundaries to build a more sophisticated mental model of its capabilities.

“Let me try something more abstract first... just want to see how it handles it.”

Usability is the new frontier for democratizing data.



For Product Leaders: User-centric design is as critical as model accuracy.

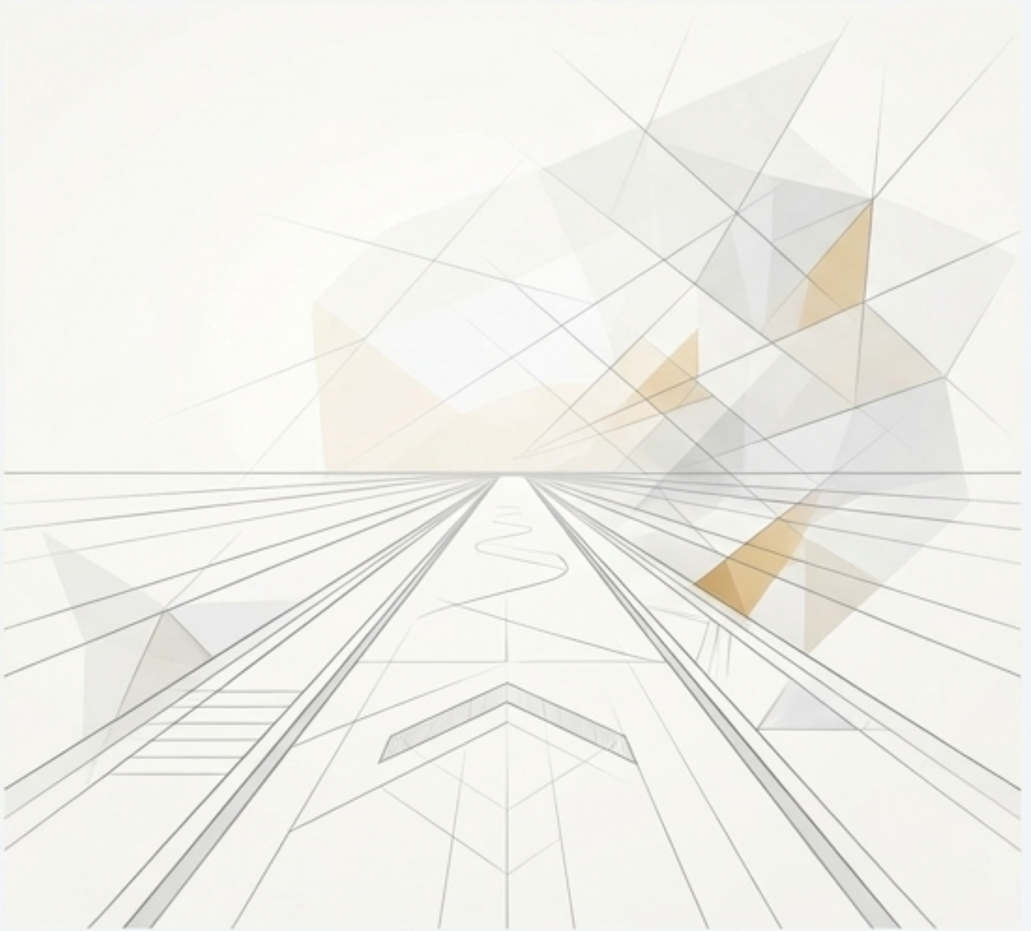
Features that enhance transparency, interactivity, and effective error management are paramount for adoption.

For Analytics Teams: Modern NLIDBs can significantly boost productivity, especially on complex ad-hoc queries, and empower non-technical stakeholders to self-serve, freeing up expert time.

For Everyone: The barrier to sophisticated data analysis is demonstrably lower.

This technology enables more people to ask more complex questions of their data directly.

The journey is promising, but important challenges remain.



Generalization: Performance needs to be validated on more complex, real-world enterprise databases with hundreds of tables.

Adaptivity: Future interfaces should adapt suggestions and explanations based on user expertise and historical behavior.

Long-Term Adoption: Longitudinal studies are needed to understand sustained learning, trust retention, and how these tools integrate into daily workflows over time.

Accuracy is table stakes. Usability wins the game.

Modern NLIDBs, when designed for usability, are demonstrably superior to traditional SQL interfaces in **speed, accuracy, and the human experience.**

They don’t just change the tool; they change the user’s **strategy, confidence, and cognitive workflow.**

The future of data interaction is conversational, intuitive, and empowering.

Thank You.

Q&A

For further reading, the original research paper can be found at:
arXiv:2511.14718v1