



# Dynamic Query-Driven Table Generation with Large Language Models

Victor Giannakouris — Immanuel Trummer

## Data Systems Today

- **Closed-world** assumption
- Preloaded, static, structured data.
- Real-world data lives in messy formats
- Web, APIs, and files.
- ETL is a bottleneck.

## swelldb

- **Open-world access**
- Table generation at **query-time**
- LLMs — Data operators
- Web, APIs, and files.
- Structured, SQL-ready tables, **on-the-fly**.

```
> pip install swelldb

from swelldb import SwellDB, OpenAILLM

swelldb = SwellDB(OpenAILLM("gpt-4o"))
tbl = swelldb.table_builder() \
    .set_content("US States") \
    .set_schema("state_name str, region str") \
    .build()
```

```
#LLMTable[schema=['state_name', 'region']]

state_name    region
0      Alabama    South
1      Alaska     West
2      Arizona    West
3      Arkansas   South
4      California  West
5      Colorado   West
...
```

Source Code: <https://github.com/SwellDB/SwellDB>

## Dynamic Table Generation

### Local Dataset



Mutation Data

sample_id	mutation
S001	BRCA1 c.68_69delAG
S007	MYH7 p.R453C
S009	KRAS p.G12D
S010	TTR V30M

### 1. Data Extraction

```
SELECT mutation
FROM mutation_tbl
```

mutation
BRCA1 c.68_69delAG
MYH7 p.R453C
KRAS p.G12D
TTR V30M

### 2. Web Search

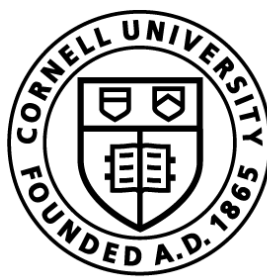


Structured search results

protein	disease
BRCA1	Cancer
β-Myosin	Cardiomyopathy
KRAS	Cancer
Transthyretin	Amyloidosis

mutation	protein	disease
BRCA1 c.68_69delAG	BRCA1	Cancer
MYH7 p.R453C	β-Myosin	Cardiomyopathy
KRAS p.G12D	KRAS	Cancer
TTR V30M	Transthyretin	Amyloidosis

### 3. Table Synthesis



**Cornell Bowers C-IS**  
College of Computing  
and Information Science