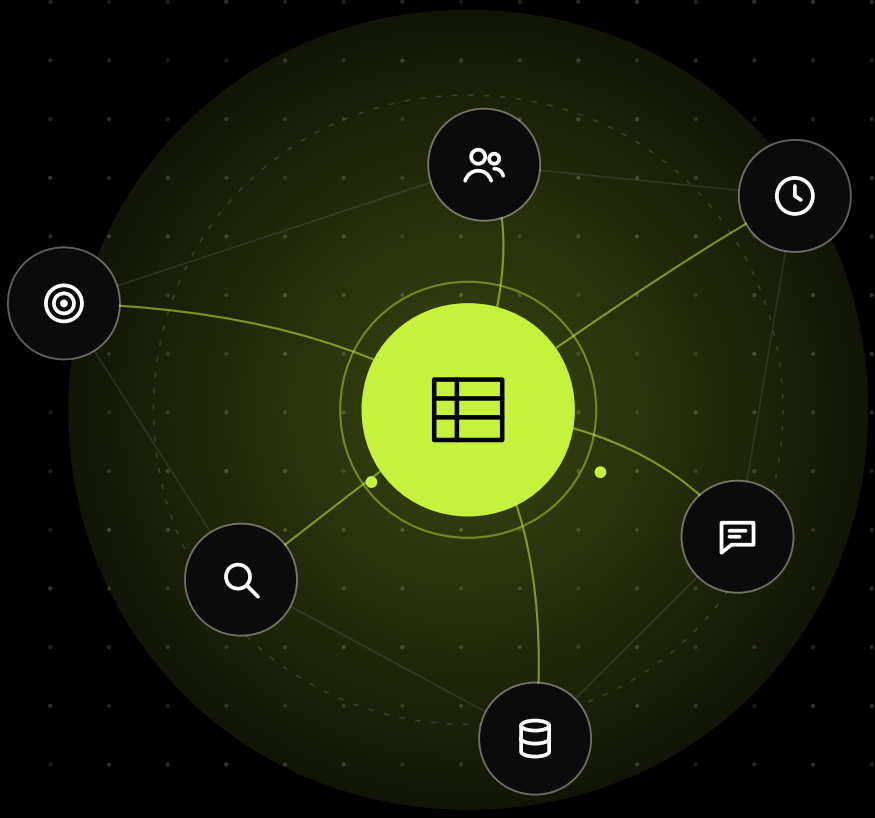


Uber

Data questions answered in 3 minutes, not 10

Uber built QueryGPT so staff can ask data questions in plain English. AI agents pick the right data and write the SQL query for them.

Transportation and delivery | Operations data and reporting | Natural language to SQL | AI agents

3 min
to write a query, instead of about 10

~300
daily active users in the limited release

~78%
of users said it cut the time they spent

AT A GLANCE

COMPANY	VOLUME	AI MODEL NAMED	REPORTED
Uber	About 1.2 million interactive data queries a month	GPT-4 Turbo (128K context)	Uber Engineering Blog, Sept. 19, 2024

01

The challenge

Uber's data platform handles about 1.2 million interactive queries a month, and each one can take around 10 minutes to write.

Writing a query takes a solid understanding of SQL, the language used to pull data from databases, plus deep knowledge of how Uber's internal data models represent business concepts.

Uber has hundreds of thousands of datasets, and some important tables have more than 200 columns. The Operations organization alone contributes about 36% of the queries.



1.2M

interactive data queries every month



~10 min

to write each query by hand



200+

columns in some of the most important tables

BEFORE

- × Every query written by hand in SQL
- × Needed deep knowledge of internal data
- × About 10 minutes per query
- × Large tables too big to give an AI at once

AFTER

- ✓ Questions asked in plain English
- ✓ AI agents pick the business area and tables
- ✓ User confirms the tables before the SQL is written
- ✓ About 3 minutes per query

02

The solution

QueryGPT takes a question in plain English and returns an SQL query plus an explanation of how it was built.

It uses a chain of AI agents. One maps the question to a business area, one picks the tables, and one trims away columns that are not needed, which cut the size and cost of each AI call.

Example queries and table schemas for that business area are retrieved as a guide. The first version used 7 tables and 20 sample queries; there were about 20 or more iterations after that.

HOW IT WORKS

STEP 01

Sharpen the question

A prompt enhancer clarifies what the user is asking.

STEP 02

Find the business area

An intent agent maps the question to one or more workspaces.

STEP 03

Pick the tables

A table agent proposes tables and the user confirms them.

STEP 04

Trim the columns

A column prune agent removes columns that are not needed.

STEP 05

Use examples as a guide

Sample queries and schemas from that area are retrieved.

STEP 06

Write the SQL

The AI returns the SQL query and an explanation.

BUILT WITH

GPT-4 Turbo · Retrieval-augmented generation · Intent agent · Table agent · Column prune agent · 11 system workspaces

03

The results

In a limited release to some Operations and Support teams, QueryGPT averaged about 300 daily active users, and about 78% said the generated queries cut the time they would have spent writing from scratch.

~3 min

to generate a reliable query, versus about 10

~300

daily active users in the limited release

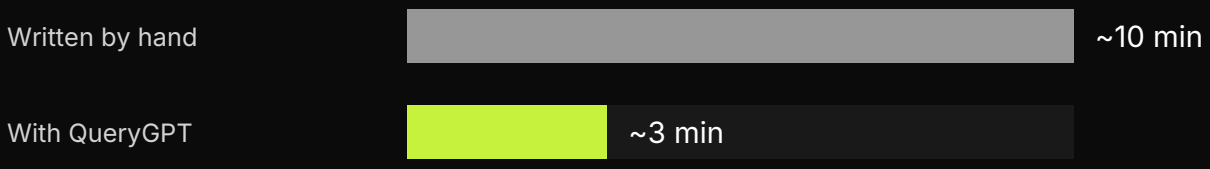
~78%

said it reduced the time they would have spent

11

system workspaces shipped, such as Ads, IT and Core Services

MINUTES TO WRITE ONE QUERY



TIMELINE

May 2023
Proposed at Uber's Generative AI Hackdays

Version 1
7 tables and 20 sample queries

20+ iterations
Agents for intent, tables and columns added

Sept. 19, 2024
Limited release to Operations and Support teams

04

What made it work, and what to watch

> “Given that each query can take around 10 minutes to author, the introduction of QueryGPT, which can automate this process and generate reliable queries in just about 3 minutes, represents a major productivity gain.”
>
> Jeffrey Johnson, Callie Busch and Abhi Khune, Uber Engineering Blog

1	**Narrow the search space**	Mapping each question to a business area first made table choice much easier.
2	**Keep a person in the loop**	Users confirm the tables before the SQL is written, which catches wrong picks early.
3	**Expect some made-up tables**	The team reports the AI can still invent tables or columns, and planned a validation step.
4	**Measure carefully**	Uber built a set of known question and answer pairs to test against, and saw about 5% run-to-run variation.

IS THIS A FIT FOR YOUR BUSINESS?

- ✓ Does your team wait on one person to pull reports from your data?
- ✓ Do the same kinds of business questions come up every week?
- ✓ Is your data spread across tables or systems only a few people understand?
- ✓ Would faster answers change decisions in sales, stock or service?

05

How Agent Victoria could do this for you

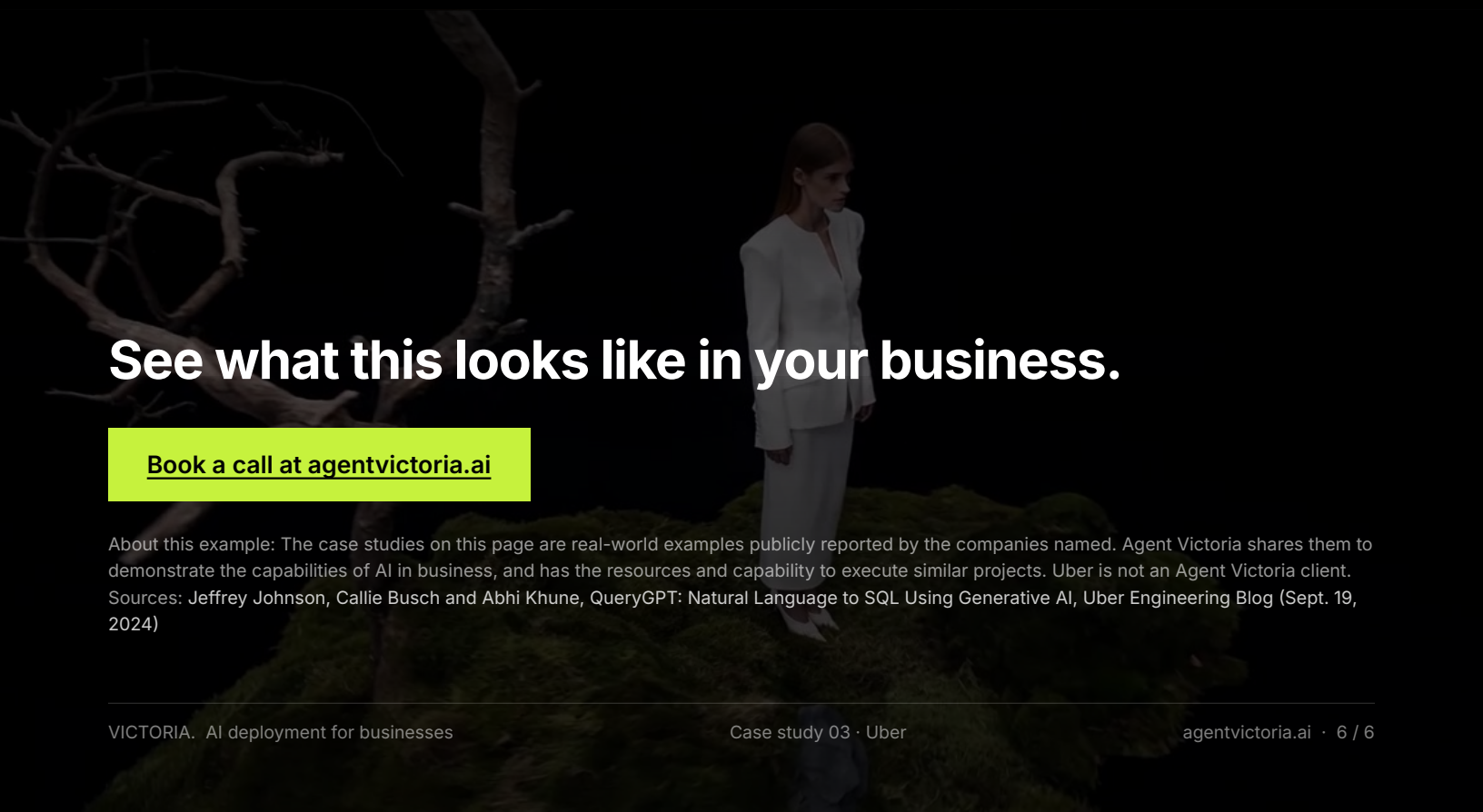
We could give your team a plain-English way to ask questions of your sales, inventory or customer data. No more waiting on the one person who knows how to build reports.

01
Connect your data
Sales, inventory or customer records you already keep.

02
Describe your business
Which tables mean what, in your own words.

03
Ask in plain English
Your team types a question and gets an answer.

04
Check the working
Every answer shows how it was built.



See what this looks like in your business.

Book a call at agentvictoria.ai

About this example: The case studies on this page are real-world examples publicly reported by the companies named. Agent Victoria shares them to demonstrate the capabilities of AI in business, and has the resources and capability to execute similar projects. Uber is not an Agent Victoria client. Sources: Jeffrey Johnson, Callie Busch and Abhi Khune, QueryGPT: Natural Language to SQL Using Generative AI, Uber Engineering Blog (Sept. 19, 2024)