

# Chat with your Enterprise Data

MARCO TONNARELLI, TU EINDHOVEN & TILBURG UNIVERSITY, NL  
FILIPPO SCARAMUZZA, TU EINDHOVEN & TILBURG UNIVERSITY, NL  
SIMON HARRER, ENTROPY DATA, DE  
LINUS DIETZ, KING’S COLLEGE, UK



Entropy Data



Data Product MCP: Chat with your Enterprise Data

Marco Tonnarelli<sup>1</sup>, Filippo Scaramuzza<sup>1</sup>, Simon Harrer<sup>2</sup>, and Linus W. Dietz<sup>3</sup>

<sup>1</sup> Eindhoven University of Technology, Tilburg University, 's-Hertogenbosch, Netherlands  
(m.tonnarelli, f.scaramuzza)@tue.nl

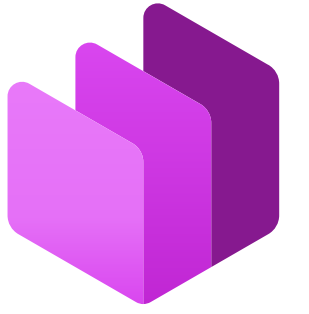
<sup>2</sup> Entropy Data GmbH, Monheim am Rhein, Germany  
simon.harrer@entropy-data.com

<sup>3</sup> King's College London, London, United Kingdom  
linus.dietz@kcl.ac.uk

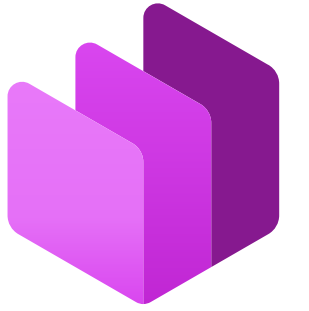
**Abstract.** Computational data governance aims to make the enforcement of governance policies and legal obligations more efficient and reliable. Recent advances in natural language processing and agentic AI offer ways to improve how organizations share and use data. But many barriers remain. Today's tools require technical skills and multiple roles to discover, request, and query data. Automating data access using enterprise AI agents is limited by the means to discover and autonomously access distributed data. Current solutions either compromise governance or break agentic workflows through manual approvals. To close this gap, we introduce Data Product MCP integrated in a data product marketplace. This data marketplace, already in use at large enterprises, enables AI agents to find, request, and query enterprise data products while enforcing data contracts in real time without lowering governance standards. The system is built on the Model Context Protocol (MCP) and links the AI-driven marketplace with cloud platforms such as Snowflake, Databricks, and Google Cloud Platform. It supports semantic discovery of data products based on business context, automates access control by validating generated queries against approved business purposes using AI-driven checks, and enforces contracts in real time by blocking unauthorized queries before they run. We assessed the system with feedback from n = 16 experts in data governance. Our qualitative evaluation demonstrates effectiveness through enterprise scenarios such as customer analytics. The findings suggest that Data Product MCP reduces the technical burden for data analysis without weakening governance, filling a key gap in enterprise AI adoption.

1 Introduction

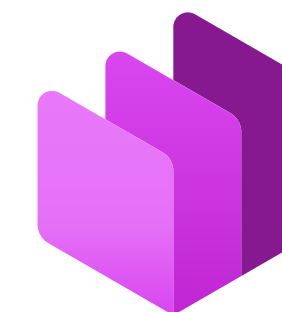
Computational data governance lets organizations automate governance by embedding policies and enforcement in code and tools. When done well, it supports value creation and becomes a key part of data management [52,36]. However, the many legal obligations, and the gaps in skills, knowledge, and experience between



**Enable AI to answer  
any business question**



**Enable AI to answer  
any business question  
while enforcing  
data governance policies.**



 What's new, Simon?

Who are our top customers?

+



 Research

Claude Sonnet 4 



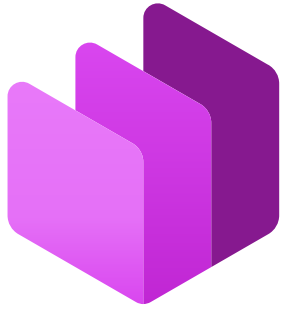
 Code

 Write

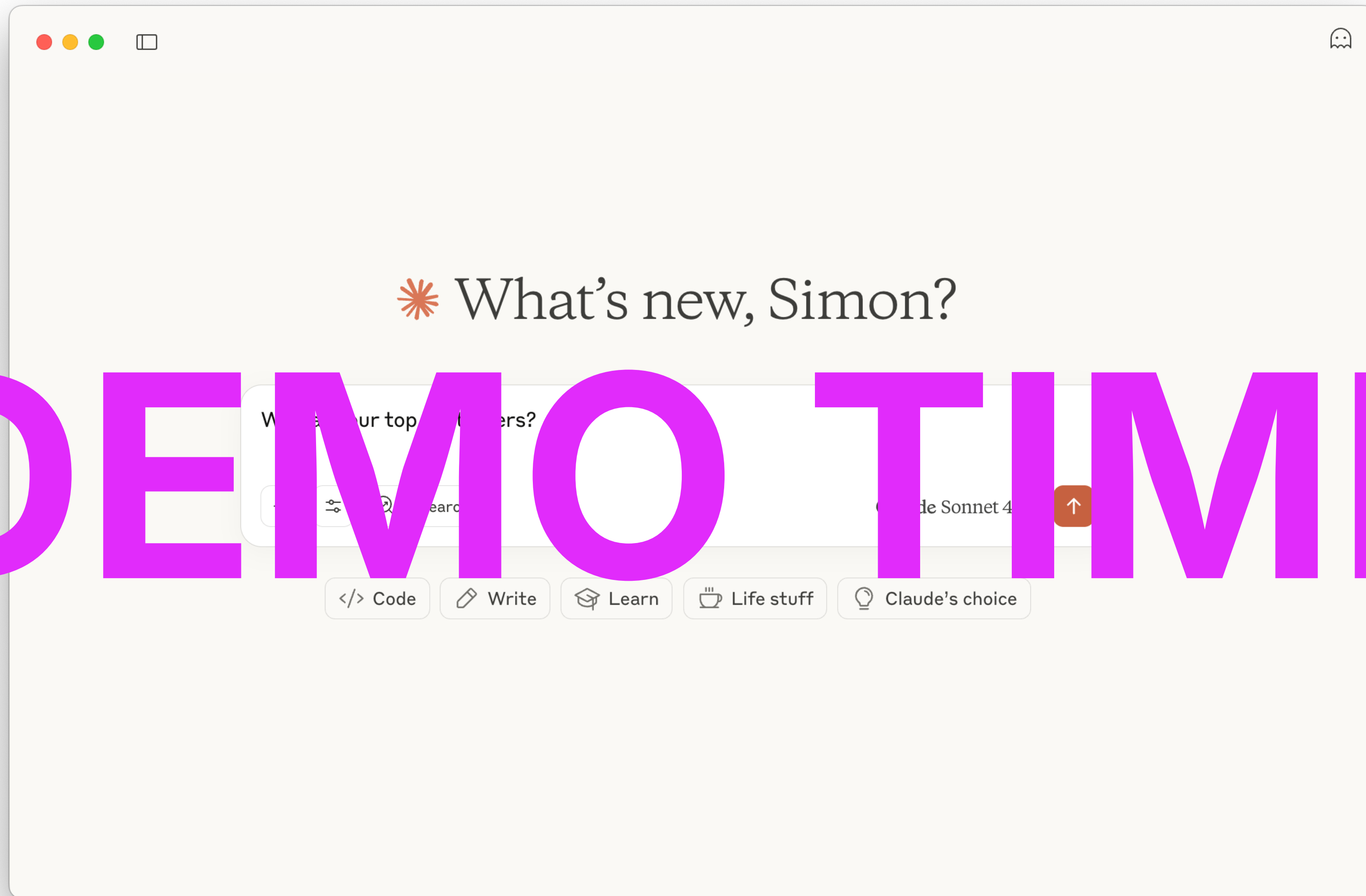
 Learn

 Life stuff

 Claude's choice

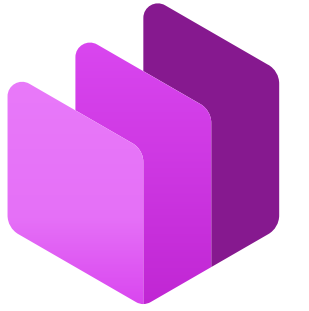


DEMO TIME



## REVISITING THE USE CASES

# We enable governed, audited access to data through AI



We can find the right data product for a given business question

We can request access with a specific purpose

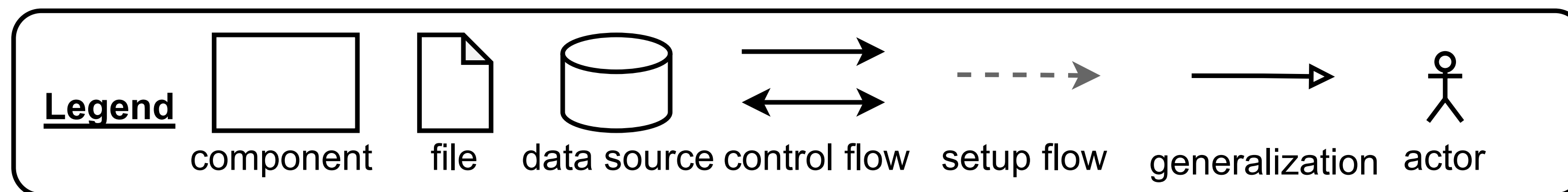
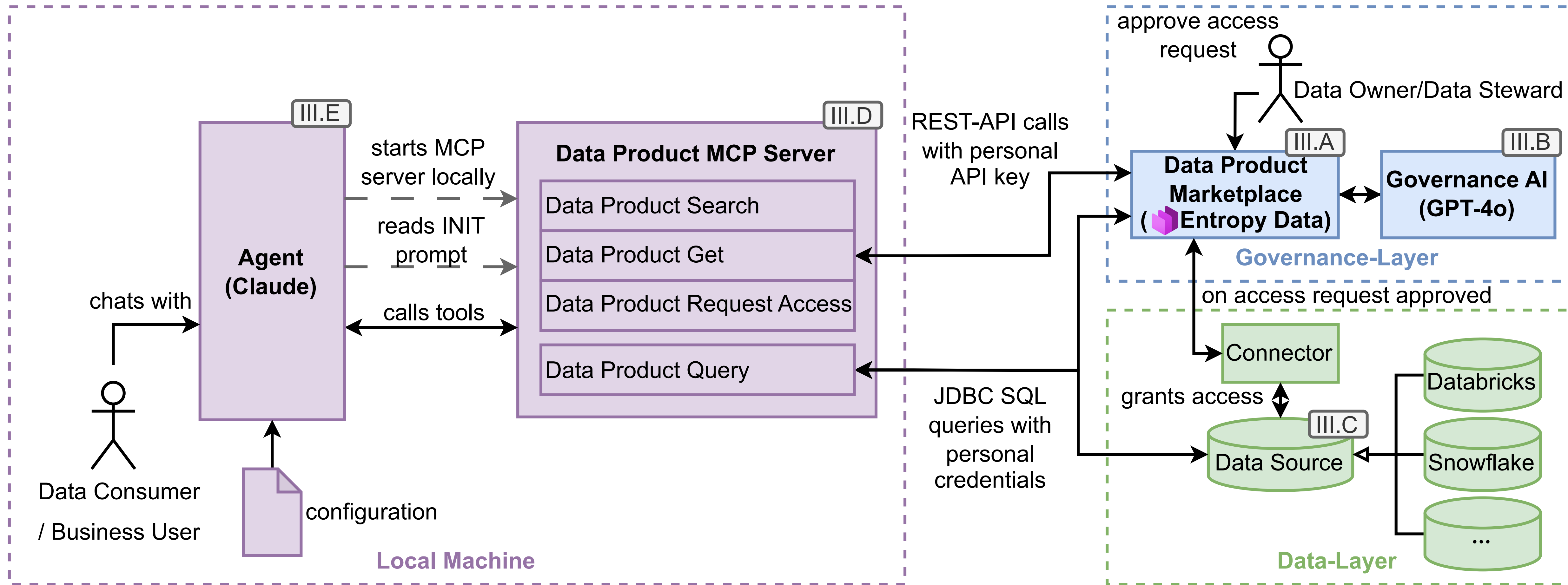
We can query the data in the right way

We can audit the why for every SQL query

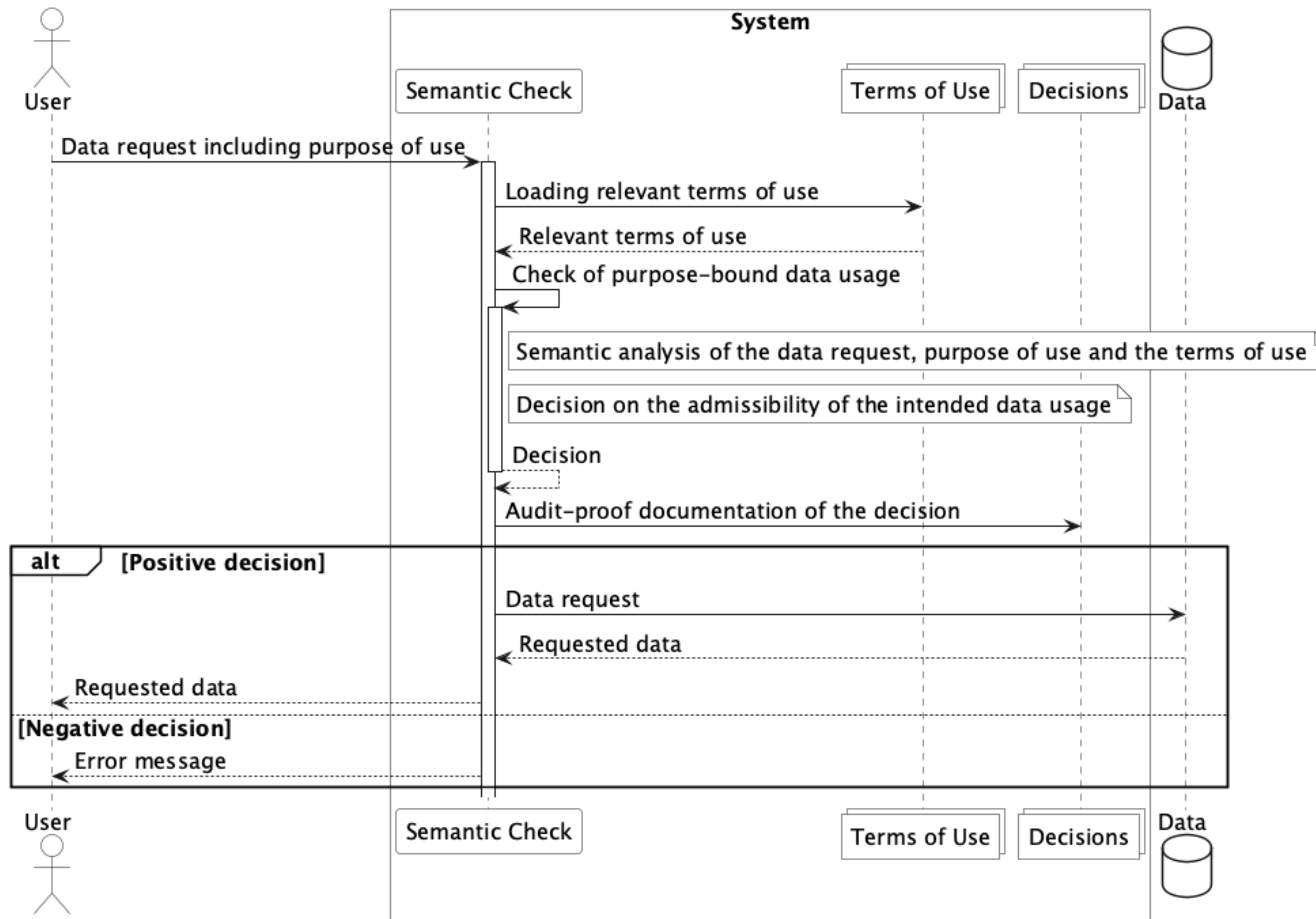
We can block SQL queries if the purpose is not covered by governance

# AUTOMATICALLY!

**But how does it work?**



"If an agent can access data it was not supposed to use for a given purpose, it eventually will. Murphy had no idea he was writing data governance policy!" — Florian Bigelmaier, BARC



**But why did it work?**

**Because of AI agents  
capabilities**

**Because of MCP tools  
working together at one central place**

**Because of good metadata**

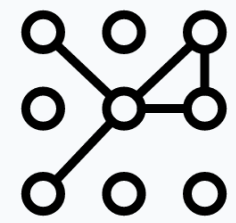
WHY DOES THIS WORK SO WELL?

# Success Factors



## Built on Data Products and Data Contracts

Ownership through data products and trust through testable data contracts



## Anchored in Semantics

So we know what every field means, and how we can combine data products



## Metadata optimized for Agents based on Open Standards

Open Data Product Standard (ODPS), Open Data Contract Standard (ODCS),  
Open Semantic Interchange (OSI)



## Compliant by design

Purpose-based access management, based on access agreements, data contracts, and global policies

# Evaluation

# Qualitative Evaluation

**Procedure:** Show demo of the prototype at an Industry Conference and a Webinar on LinkedIn (2x50 people) → Let participants answer questionnaire

After seeing the demo, what tasks or use cases in your work do you think Data Product MCP enables that you could not do (or could do only with much more effort) before?

How do you expect the use of Data Product MCP to change your current approach to data work (e.g. adoption of Data Mesh, relevance of dashboards, collaboration across teams)?

What potential limitations or concerns (e.g. scalability, correctness, security, compliance) do you see that might prevent adoption in your organization?

How reliable would Data Product MCP's results need to be for you to trust and adopt it in production?

# Evaluation (16 Data Governance Experts)

5 product managers, 4 data architects, 2 C-level, 3 directors/heads, 2 researchers

*"It makes it very data-democratic,  
as just asking questions is a nice abstraction."*

## Identified Key Benefits

Natural  
Language  
Data Access

Stronger  
Governance

Quick  
Self-Service

## Identified Adoption Risks

Data  
/  
Metadata  
Quality

Security  
/  
Identity

Scalability  
/  
Cost

Compliance  
/  
Trust

# Qualitative Evaluation (n = 16)

How reliable would Data Product MCP's results need to be for you to trust and adopt it in production?

Only 2 out of 16 said 100%. All others were fine with less!

**And today?**

**⇒ Biggest adoption issue is Deployment.  
Enterprises need built-in Agents**

**Every big player in the market now adds such capabilities\* within their offering!**

**Snowflake, Google,  
Databricks, Microsoft, ...**

\* without purpose-based access control.

[Marketplace](#)[Studio](#)[Governance](#)[Try Entropy Intelligence](#)[Demo](#)[Sign up for free](#)

## Entropy Intelligence

Find data products. Ask what you want to know. Get insights.

Chat with your data, securely.



### SUGGESTIONS



#### Find data products

Which data product should I use to analyze customer purchasing behavior?



#### Explore Orders

Group last 30 days orders country. Visualize as chart.



#### Query search CTR

Calculate the click-through rate and average searches per session from the search data.



#### Top Customers

What are our top customers?



I'll help you find the top customers by revenue from the Orders data product. Let me start by locating and fetching the data product details.

Perfect! I found the `Orders` data product. Let me fetch the details to see the available output ports:

Perfect! I found `Orders` > `Snowflake Orders (No PII) v2` and you already have access to it. This is the active non-PII version suitable for analytics. Let me update the context and query the top customers by revenue.

Here are your top 20 customers by total revenue from the `Orders` > `Snowflake Orders (No PII) v2` data product:

| Result table (20 rows) |             |             |               | CSV | Show code |  |
|------------------------|-------------|-------------|---------------|-----|-----------|--|
| #                      | CUSTOMER_ID | ORDER_COUNT | TOTAL_REVENUE |     |           |  |
| 1                      | cust-02070  | 8           | 6076.39       |     |           |  |
| 2                      | cust-05435  | 7           | 5417.01       |     |           |  |
| 3                      | cust-07269  | 5           | 5388.06       |     |           |  |
| 4                      | cust-07518  | 5           | 5359.12       |     |           |  |
| 5                      | cust-00006  | 10          | 5246.46       |     |           |  |
| 6                      | cust-00035  | 8           | 5051.85       |     |           |  |
| 7                      | cust-00018  | 12          | 4968.82       |     |           |  |
| 8                      | cust-06906  | 12          | 4944.64       |     |           |  |
| 9                      | cust-03039  | 4           | 4864.22       |     |           |  |
| 10                     | cust-00023  | 12          | 4730.28       |     |           |  |
| 11                     | cust-06976  | 7           | 4687.95       |     |           |  |
| 12                     | cust-02609  | 6           | 4633.06       |     |           |  |

The top customer is cust-02070 with €6,076.39 in total revenue across 8 orders, followed by cust-05435 with €5,417.01 across 7 orders. The data shows a mix of customer engagement patterns—some top customers (like cust-00018 and cust-06906) have placed 12 orders, while others like cust-03039 generated high revenue with just 4 orders, indicating high-value individual purchases.

Here are your top 20 customers by total revenue from `Orders` > `Snowflake Orders (No PII) v2`:

Key highlights:

- **cust-02070** leads with €6,076.39 across 8 orders — your highest-value customer
- **cust-05435** is second with €5,417.01 from 7 orders
- **cust-00018** and **cust-06906** tie for the most orders at 12 each, though cust-00018 has higher revenue



Progress

- ✓ Analyze purpose
- ✓ Locate Orders data product
- ✓ Query customer revenue totals
- ✓ Rank and display top customers

Purpose

Identify top customers by total revenue

Governance

2 checks

Data Product

1 port

Semantics

1 concept

Data model

2 tables

Connection

Connected

SQL

1 query

Tools

7 calls

- ✓ `searchSemanticConcepts`  
query: customer
- ✓ `findDataProductsForConcept`  
conceptId: customer
- ✓ `searchDataProducts`  
query: orders revenue transactions
- ✓ `fetchDataProduct`  
Orders
- ✓ `checkConnection`  
Orders / Snowflake Orders (No PII) v2
- ✓ `updateContext`  
Orders / Snowflake Orders (No PII) v2
- ✓ `executeQuery`  
dataProductId: orders, outputPortId: snowflake\_...



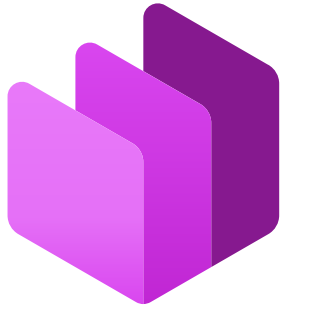
**But if you only remember one  
thing...**



**The (autonomous) agents are coming.  
And they are enterprise data hungry.**

**Automated data governance is the only  
way for them to fulfill their potential.**

ENTROPY DATA



# Thank you! Questions?

Interested in doing research together?

Happy to talk. :-)

Contact: [simon.harrer@entropy-data.com](mailto:simon.harrer@entropy-data.com)

Links:

[entropy-data.com/research](https://entropy-data.com/research)