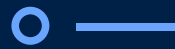




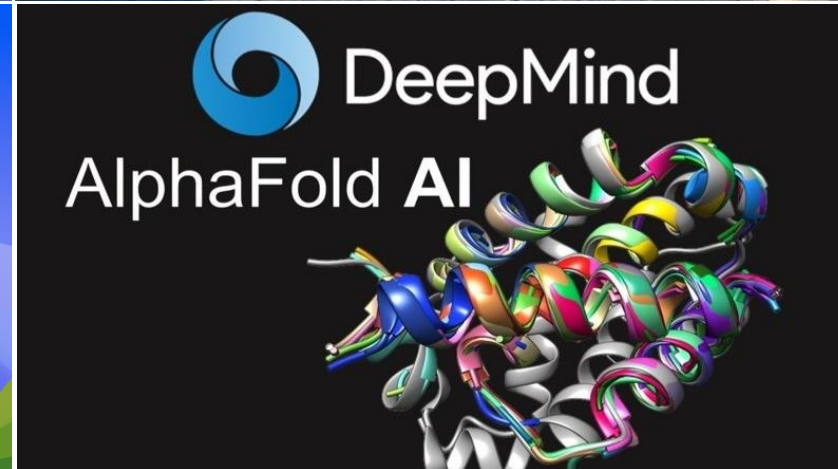
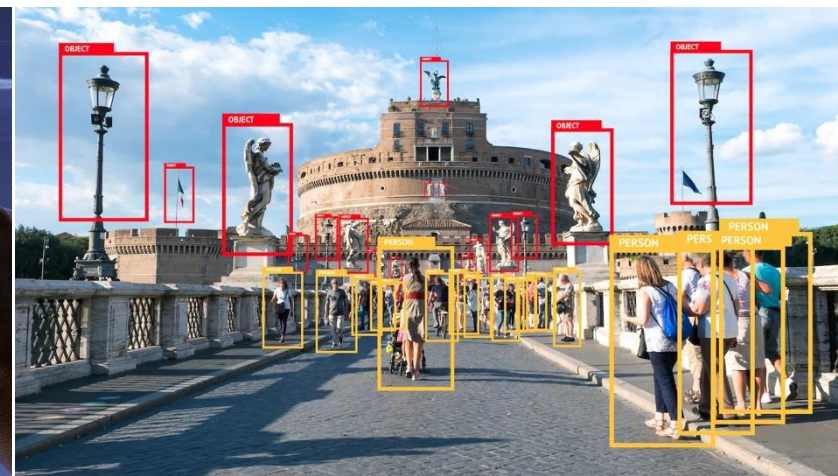
DEPARTMENT OF
**COMPUTER
SCIENCE**



Combining Knowledge-Based and Generative AI in Real-World Applications

Ian Horrocks

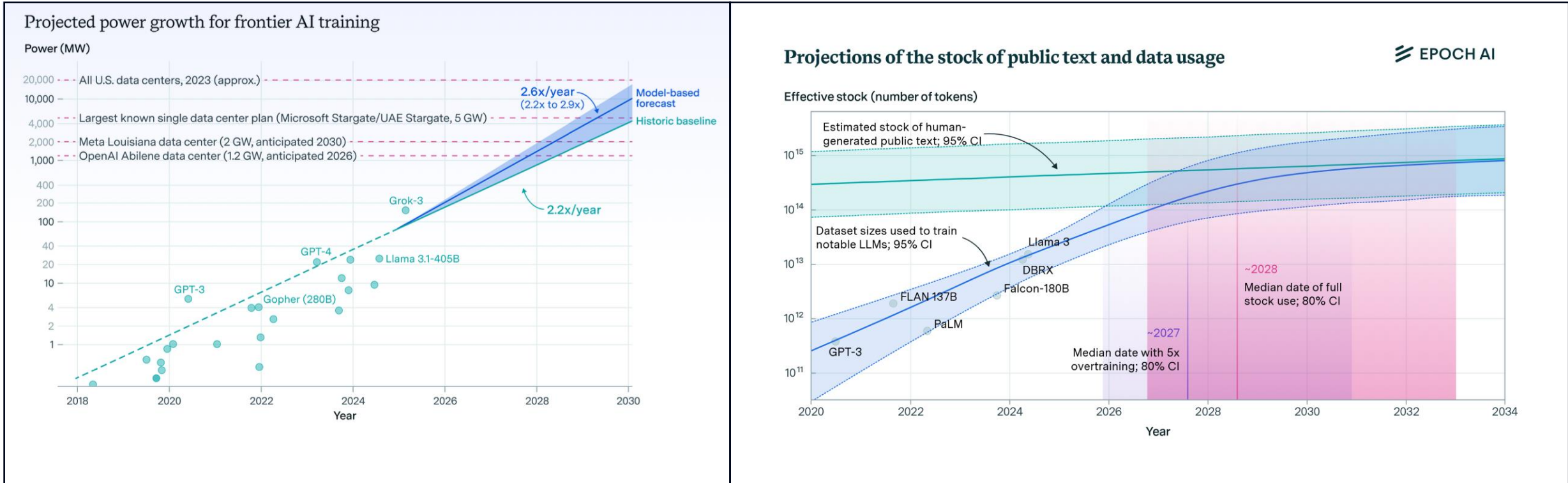
AI Successes



AI Failures



Fundamental Limits of AI?



Different Kinds of AI



Learning-Based AI

(M L / L L M)

- Works with **data**
- Answers questions using large model **learned** from noisy data
- Produces **plausible** (statistically likely) outputs
- **Opaque** reasoning process



Knowledge-Based AI

(K R R)

- Works with **data**, and **knowledge**
- Answers questions using **curated** data and knowledge
- Produces **logically entailed** outputs
- **Transparent** reasoning process



Different Kinds of AI



Learning-Based AI

(M L / L L M)



Low information density

Randomly deleting data would have **little impact** on outputs



Vague knowledge

Rules are **uncertain or unknown**



Low cost of failure

Some errors **can be tolerated**



Knowledge-Based AI

(K R R)



High information density

Randomly deleting data would have **big impact** on outputs



Precise knowledge

Rules are **known**

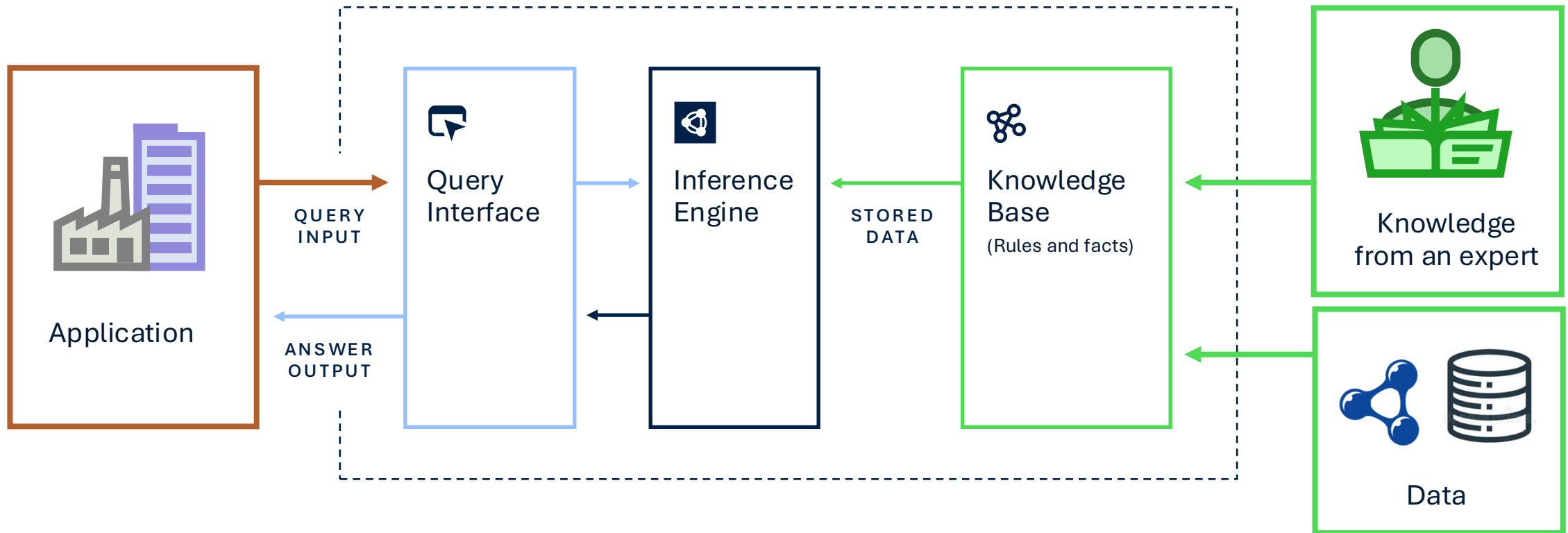


High cost of failure

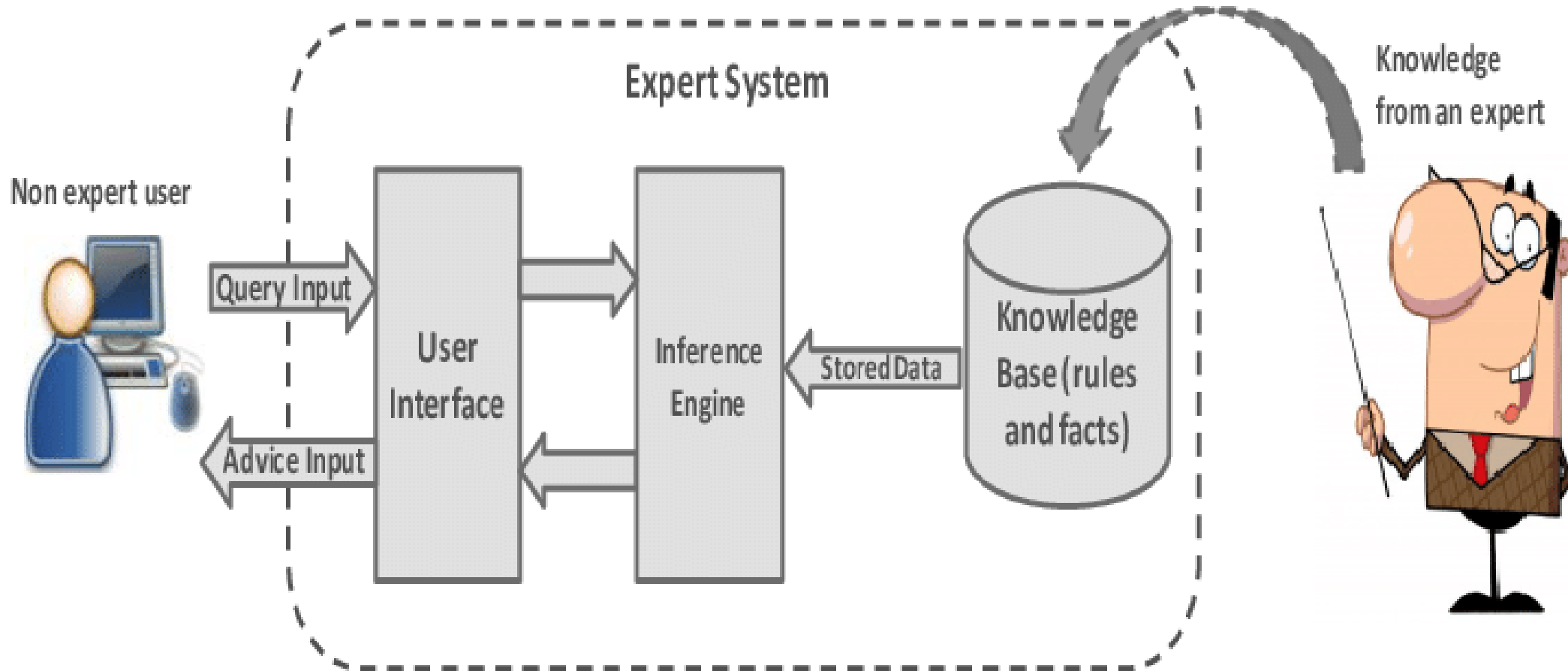
Errors **cannot be tolerated**



Knowledge-Based AI



Knowledge-Based AI = (GOF)AI?



Why Now (and not then)?

Moore's Law: The number of transistors on microchips has doubled every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World
in Data

Transistor count

50,000,000,000

10,000,000,000

5,000,000,000

1,000,000,000

500,000,000

100,000,000

50,000,000

10,000,000

5,000,000

1,000,000

500,000

100,000

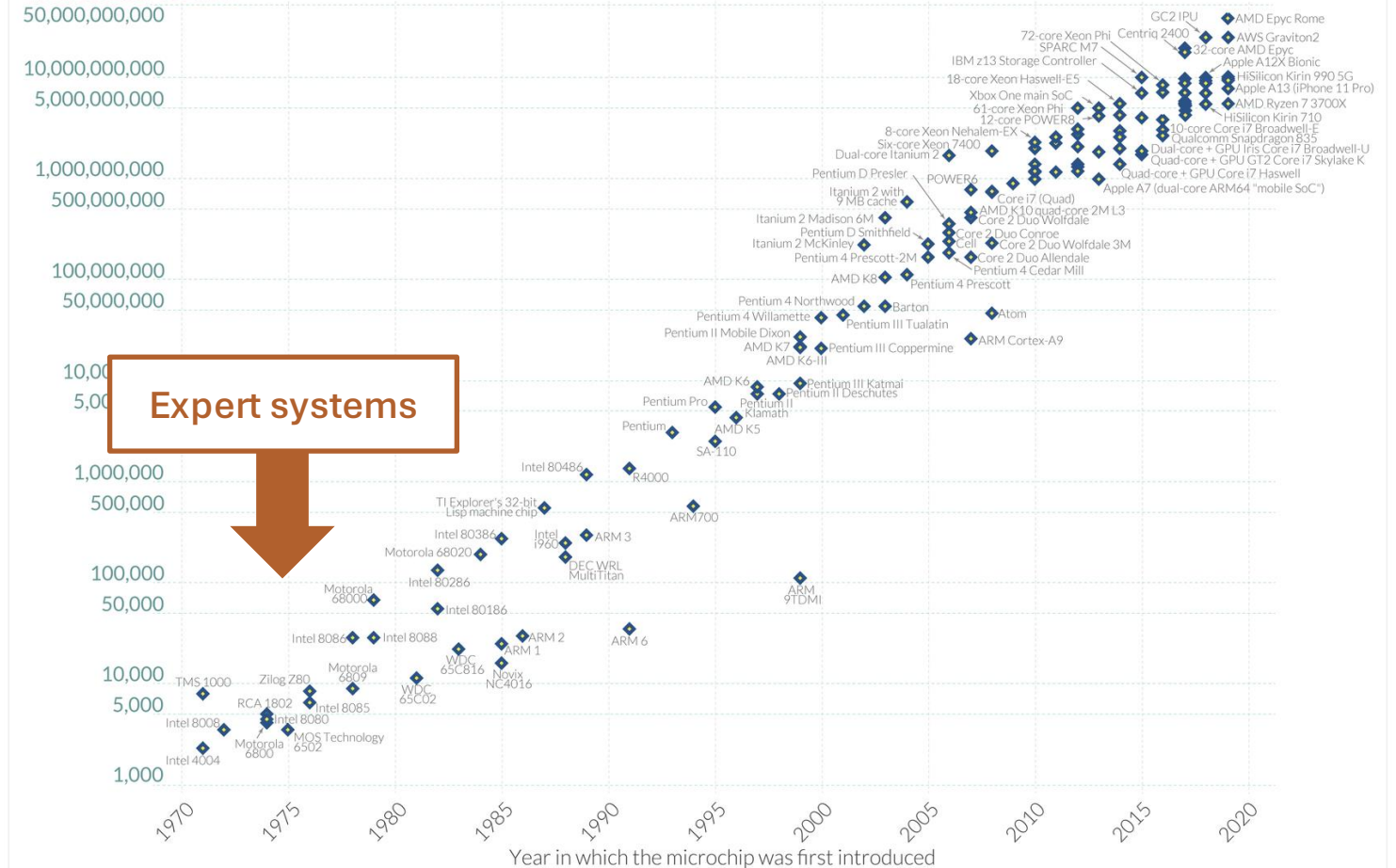
50,000

10,000

5,000

1,000

Expert systems



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.



Why Now (and not then)?

Historical price of computer memory and storage

This data is expressed in US dollars per terabyte (TB), adjusted for inflation. "Memory" refers to random access memory (RAM), "disk" to magnetic storage, "flash" to special memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).

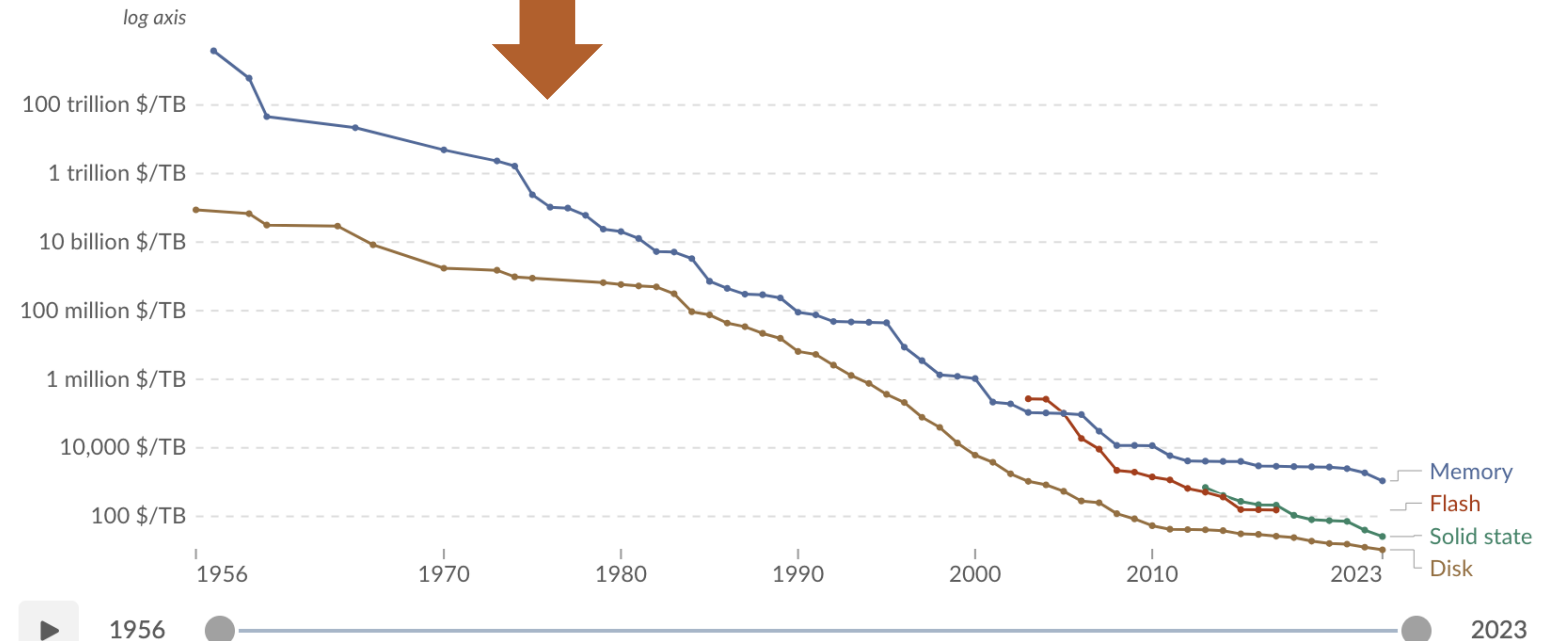
Our World
in Data

Table

Line

Expert systems

Settings



Data source: John C. McCallum (2023); U.S. Bureau of Labor Statistics (2026) – [Learn more about this data](#)

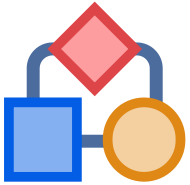
Note: For each year, the time series shows the cheapest historical price recorded until that year. This data is expressed in constant 2020 US\$.

OurWorldinData.org/technological-change | CC BY



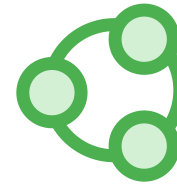
Why Now (and not then)?

Research



Parallelized semi-naïve materialization

- Concurrent data structures
- Optimized RAM-based storage



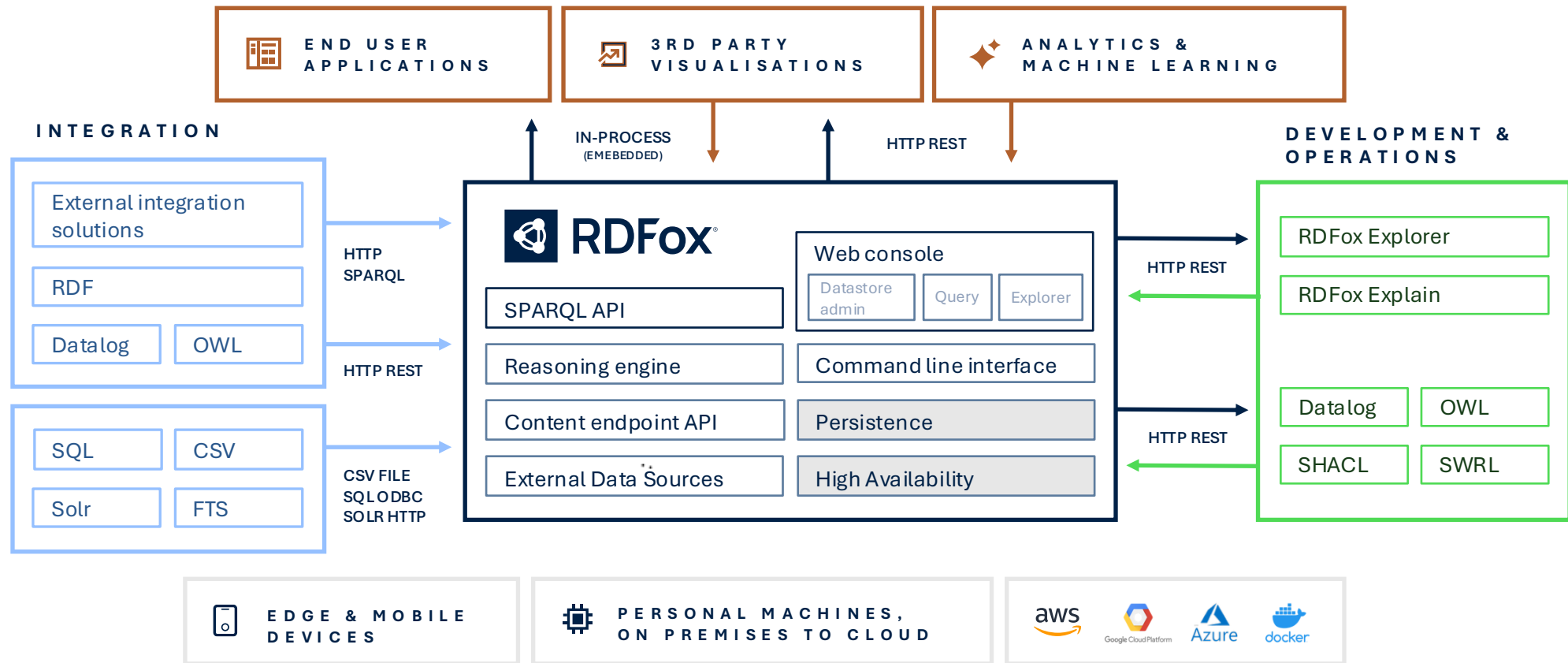
View maintenance & incremental reasoning

- DRed algorithm
- Backwards/forwards algorithm



Why Now (and not then)?

Development e.g. Oxford Semantic Technologies



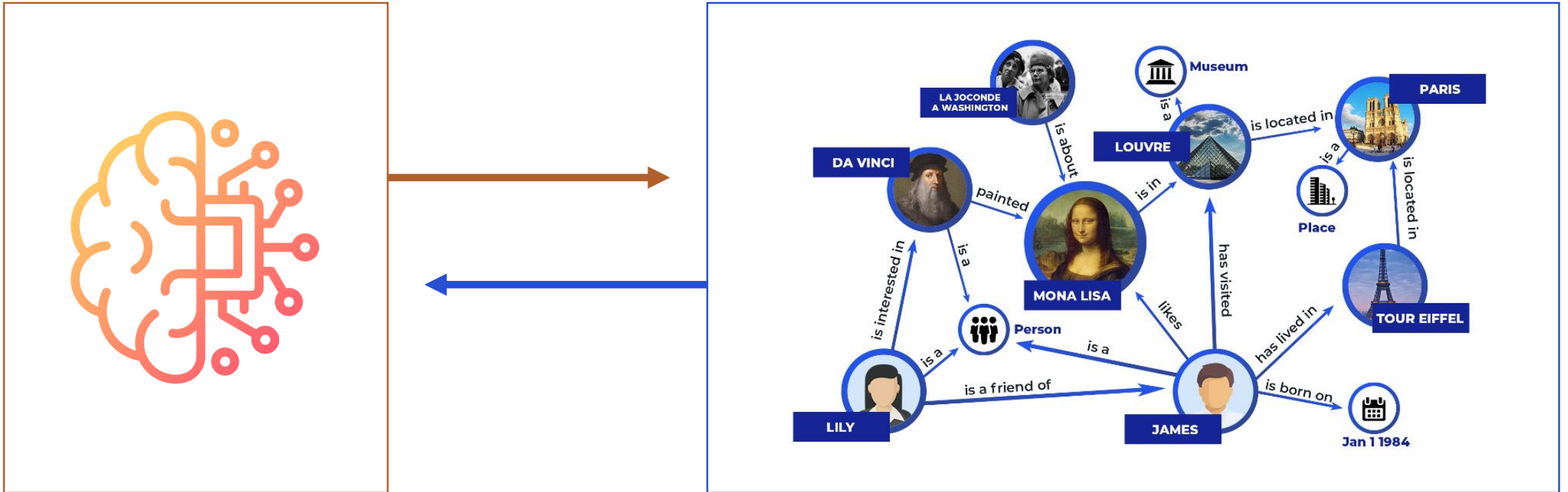
Why Now (and not then)?

Digital transformation

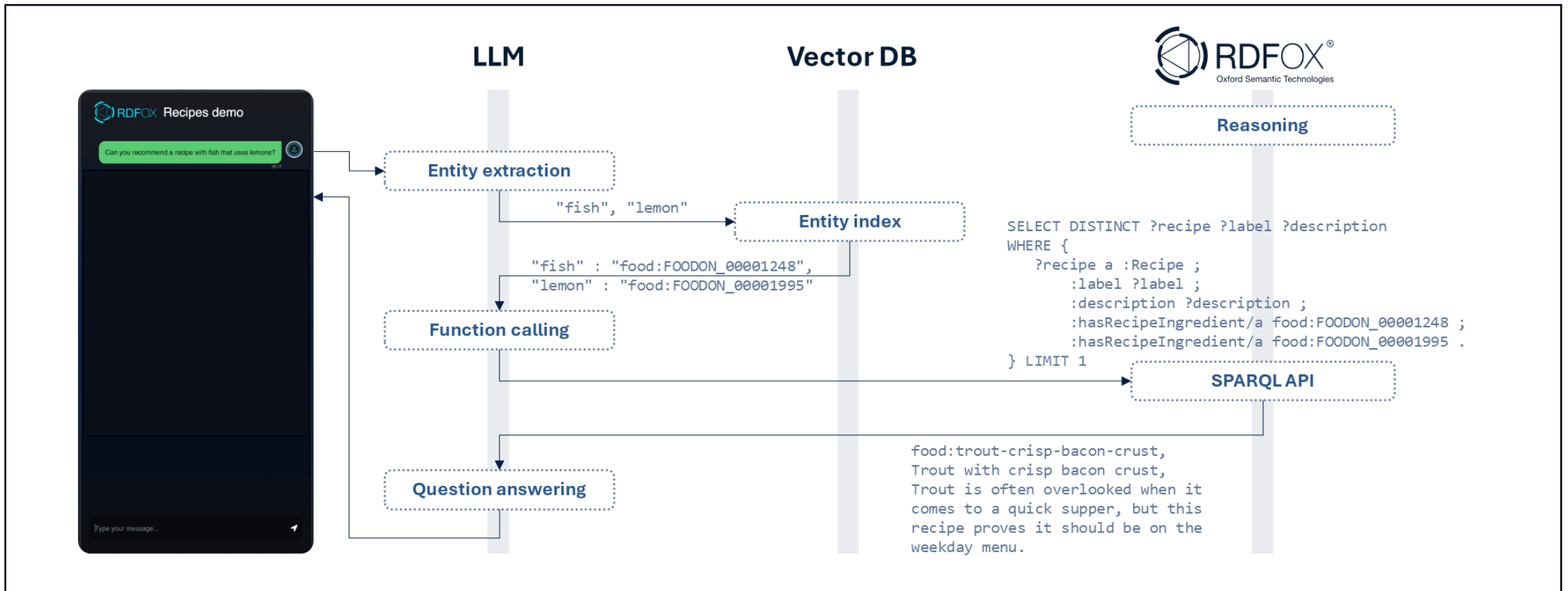


Why Now (and not then)?

Mutual Reinforcement & Hybrid Architectures



Hybrid AI Example: Graph RAG



Knowledge Engineering Bottleneck



Building KB-AI applications requires significant expertise

- You need to understand the business problem
- You need domain experts who understand the knowledge required
- You need semantic expertise to model that knowledge accurately
- You need IT expertise to connect the right data and map it to the model



That collection of skills is valuable & scarce



Use LLM-enhanced tools to support knowledge engineering!





AI-enhanced platform for collaborative knowledge engineering

The screenshot displays the RDFox Mantix user interface. At the top, an ontology diagram shows a hierarchy where 'High Value Customer' is a subtype of 'Customer', and 'Customer' has placed an 'Order'. The 'Order' entity has details, including a product and a unit price. Below this, a panel titled 'I've run the following steps to query your workspace.' lists actions like matching types with an ontology and generating a SPARQL query. The 'QUERY USED' section shows a SPARQL query to list IRIs of HighValueCustomer individuals. A 'MATCHING DATA' panel lists results with URIs. On the right, an 'AGGREGATION' panel shows a query for 'thing' ordered by 'nextExpectedOrderDate'. At the bottom, a data table shows columns for 'orderDetailID', 'productID', 'quantity', and 'unitPrice'. A 'Map data' panel shows a mapping from 'integer' to 'quantity' and 'unitPrice' to 'decimal'.

```
1 # Return overdue customers, expected order
2 and attempt DAYS calculation
3 SELECT ?customer ?nextExpectedOrderDate
4 (DAYS(NOW() - ?
5 nextExpectedOrderDate) AS ?daysOverdue)
6 WHERE {
7   ?customer
8   <http://www.example.com/terms/
9   nextExpectedOrderDate> ?
10  nextExpectedOrderDate
11  FILTER
12  }
13  ORDER BY
```

```
1 # List all IRIs of HighValueCustomer individuals
2 SELECT ?highValueCustomer
3 WHERE {
4   ?highValueCustomer a
5   <http://www.example.com/terms/HighValueCustomer> .
6 }
```

orderDetailID	productID	quantity	unitPrice
10248-11	11	12	14
10248-42	42	10	9.8
10248-72	72	5	34.8





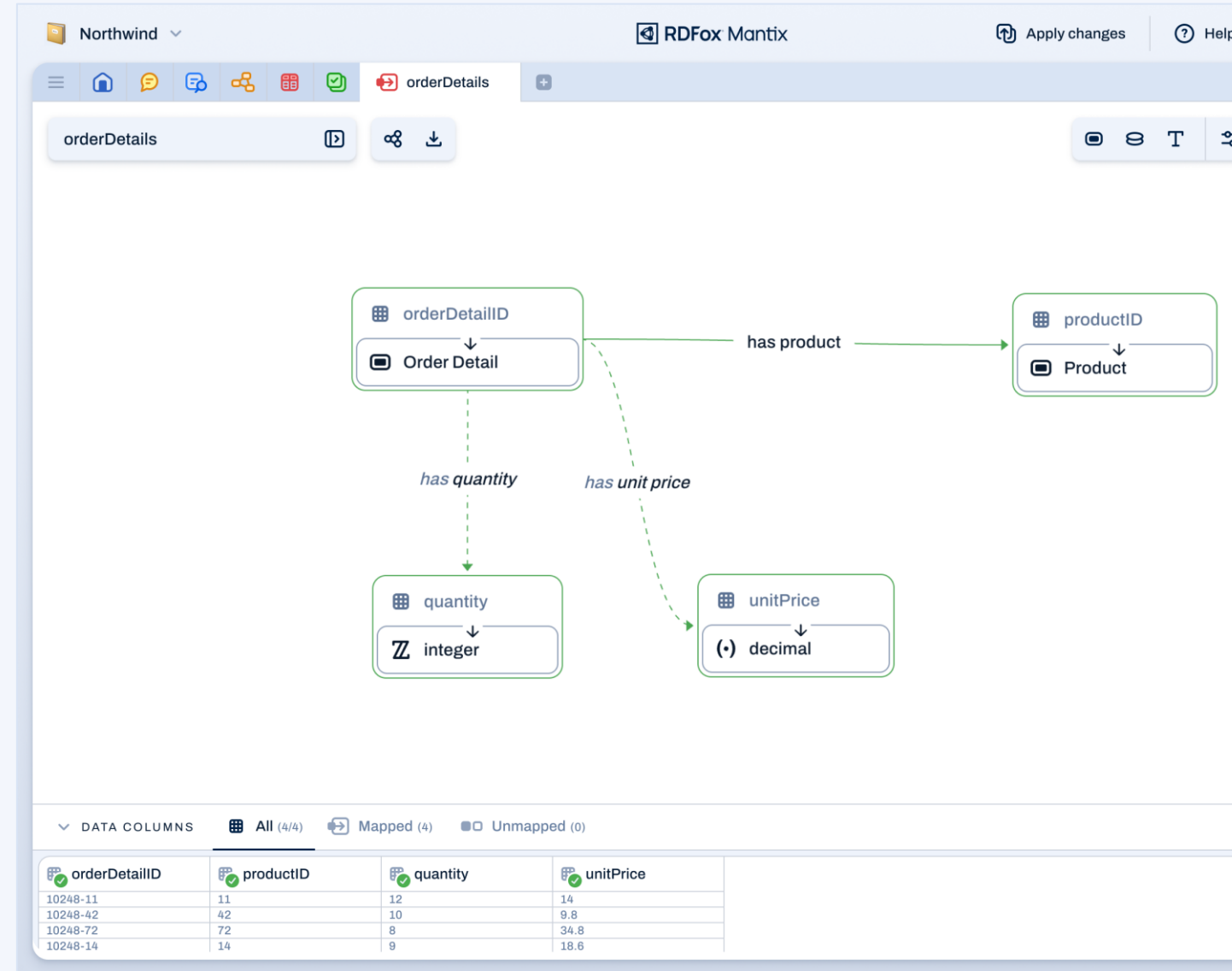
Data

Load data from RDF/DB

Map data from CSV or Parquet formats.

Preview the graph that will be generated.

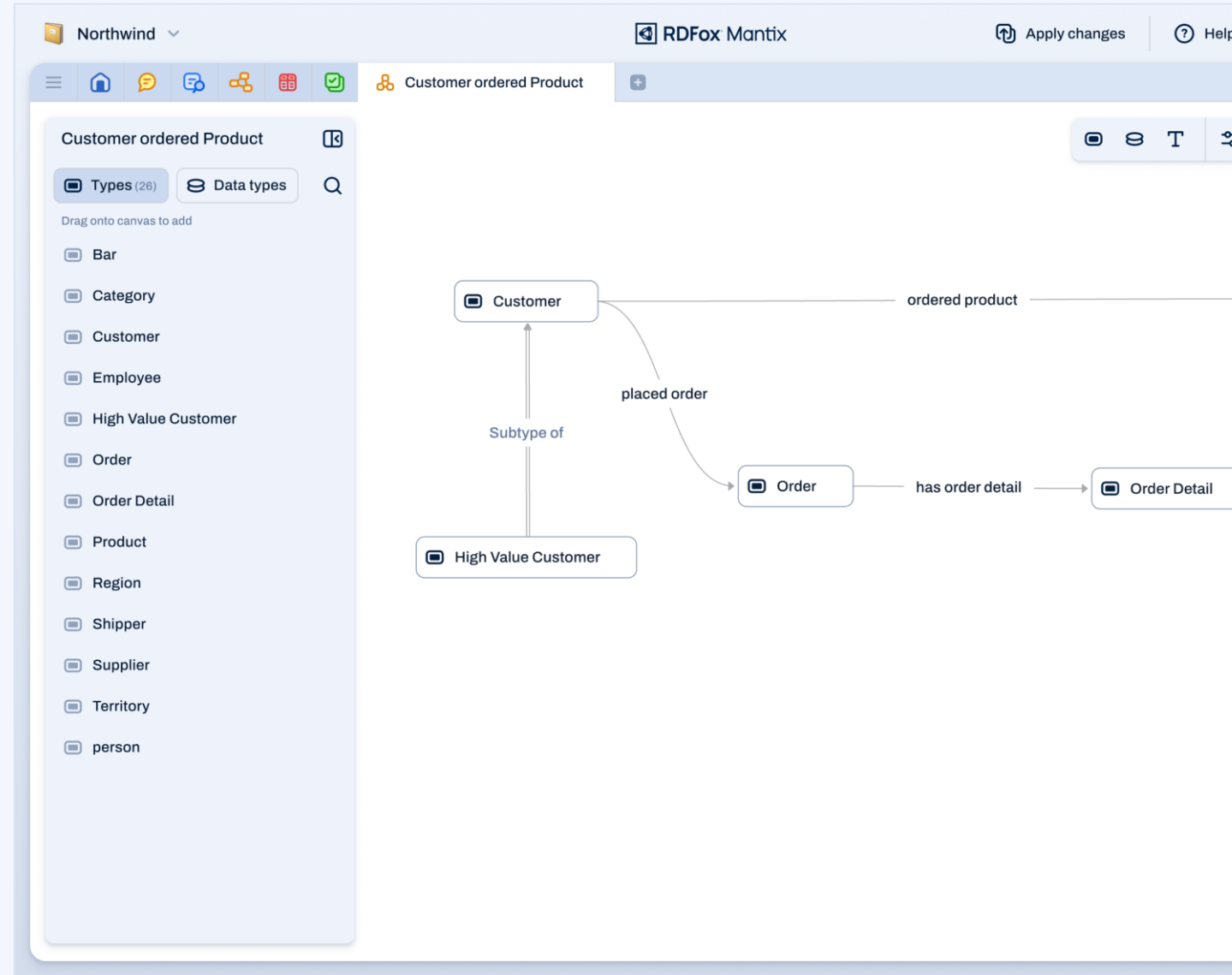
Download mappings for execution in ETL pipelines.



Ontology

Bootstrapped from data

Visual collaborative editing

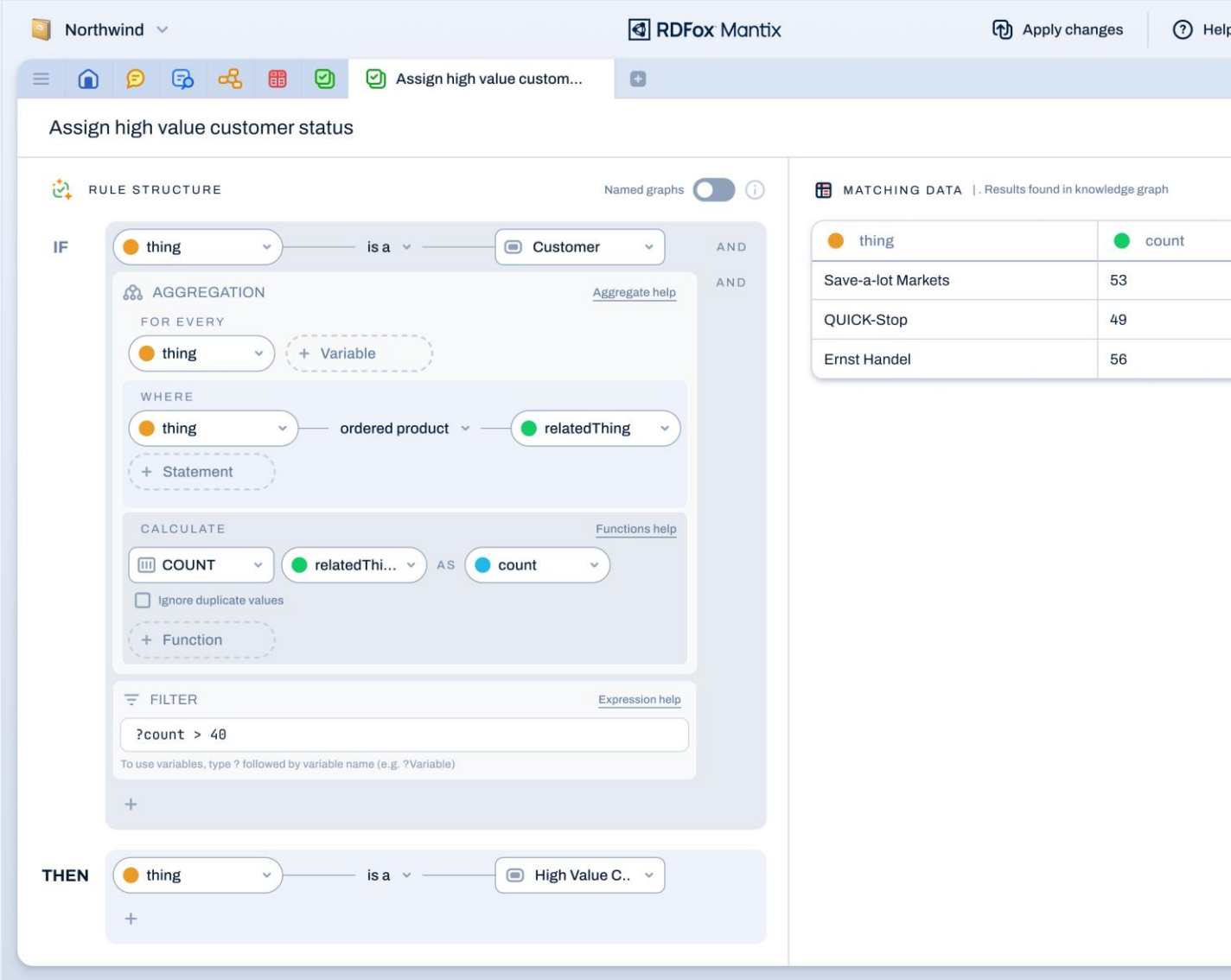


Rules

Visual rule editor.

Exploits ontology & data.

Live editing & feedback.



Northwind

RDFox Mantix

Apply changes

Help

Assign high value customer status

RULE STRUCTURE

Named graphs

IF

thing is a Customer

AND

AND

AGGREGATION

Aggregate help

FOR EVERY

thing + Variable

WHERE

thing ordered product relatedThing

+ Statement

CALCULATE

Functions help

COUNT relatedThing AS count

Ignore duplicate values

+ Function

FILTER

Expression help

?count > 40

To use variables, type ? followed by variable name (e.g. ?Variable)

+

THEN

thing is a High Value C..

+

MATCHING DATA | Results found in knowledge graph

thing	count
Save-a-lot Markets	53
QUICK-Stop	49
Ernst Handel	56

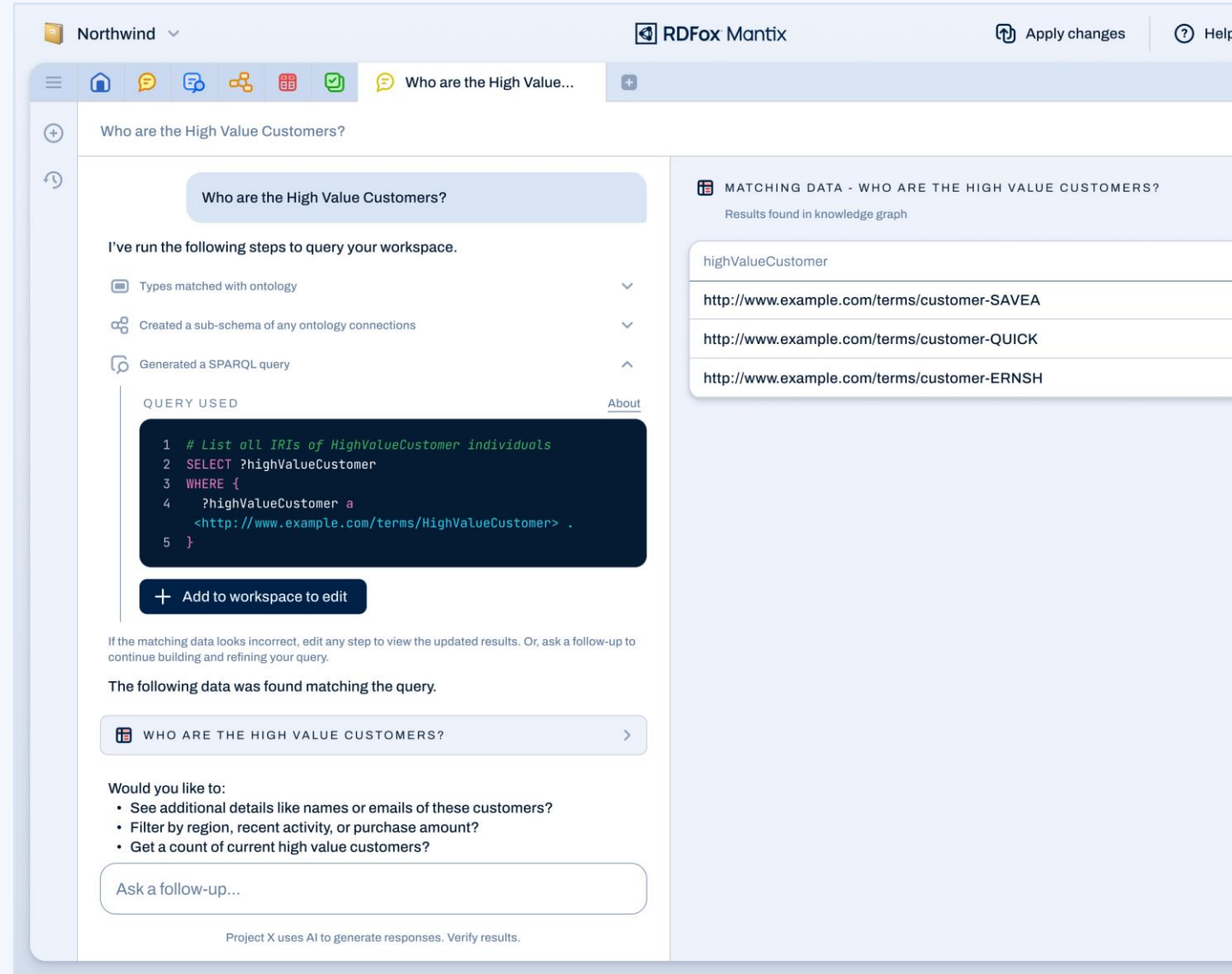


Chat

Easy query generation from natural language.

Ontology aware to reduce hallucinations.

Chat to refine follow-ups for complexity or further questions.



The screenshot displays the RDFox Mantix web interface. At the top, the header includes the 'Northwind' project name, the 'RDFox Mantix' logo, and buttons for 'Apply changes' and 'Help'. The main chat area shows a conversation with the query 'Who are the High Value Customers?'. The system response explains the steps taken: matching types with the ontology, creating a sub-schema, and generating a SPARQL query. The query is displayed in a dark box with syntax highlighting. Below the query, there is a button to 'Add to workspace to edit'. The system also provides a warning about incorrect matching data and offers suggestions for follow-up questions. On the right side, a panel titled 'MATCHING DATA - WHO ARE THE HIGH VALUE CUSTOMERS?' shows the results found in the knowledge graph, listing three URIs for high-value customers.

Who are the High Value Customers?

I've run the following steps to query your workspace.

- Types matched with ontology
- Created a sub-schema of any ontology connections
- Generated a SPARQL query

QUERY USED

```
1 # List all IRIs of HighValueCustomer individuals
2 SELECT ?highValueCustomer
3 WHERE {
4   ?highValueCustomer a
5     <http://www.example.com/terms/HighValueCustomer> .
6 }
```

+ Add to workspace to edit

If the matching data looks incorrect, edit any step to view the updated results. Or, ask a follow-up to continue building and refining your query.

The following data was found matching the query.

WHO ARE THE HIGH VALUE CUSTOMERS?

Would you like to:

- See additional details like names or emails of these customers?
- Filter by region, recent activity, or purchase amount?
- Get a count of current high value customers?

Ask a follow-up...

Project X uses AI to generate responses. Verify results.

MATCHING DATA - WHO ARE THE HIGH VALUE CUSTOMERS?

Results found in knowledge graph

highValueCustomer
http://www.example.com/terms/customer-SAVEA
http://www.example.com/terms/customer-QUICK
http://www.example.com/terms/customer-ERNSH

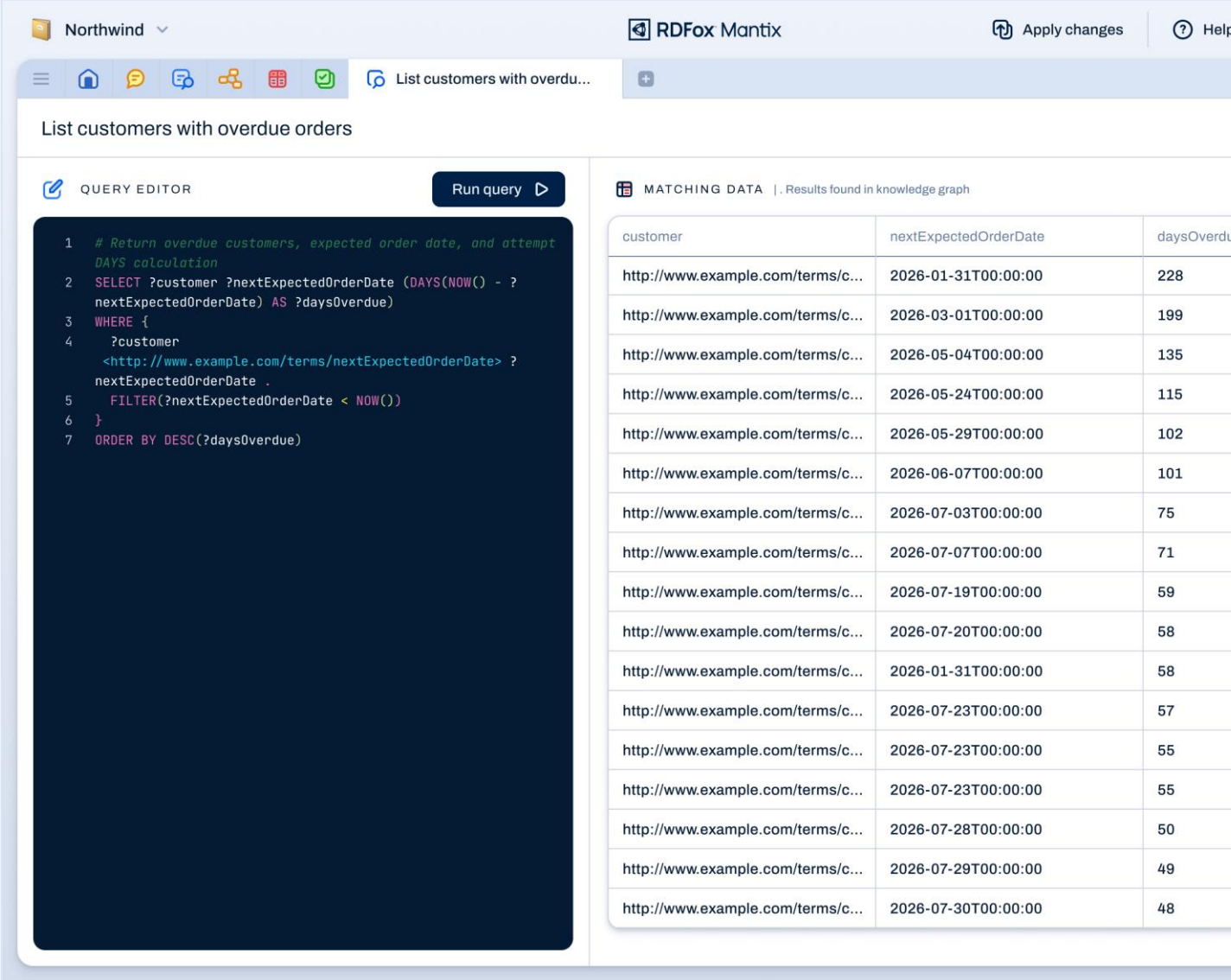


Queries

Edit queries generated by chat for precise refinement

Save queries for repeated use and later development.

Share queries with colleagues for collaboration.



The screenshot displays the RDFox Mantix web interface. At the top, the 'Northwind' database is selected. The main title bar shows 'RDFox Mantix' and an 'Apply changes' button. Below this, a search bar contains the text 'List customers with overdue...'. The main content area is divided into two panels. The left panel, titled 'QUERY EDITOR', contains a SPARQL query:

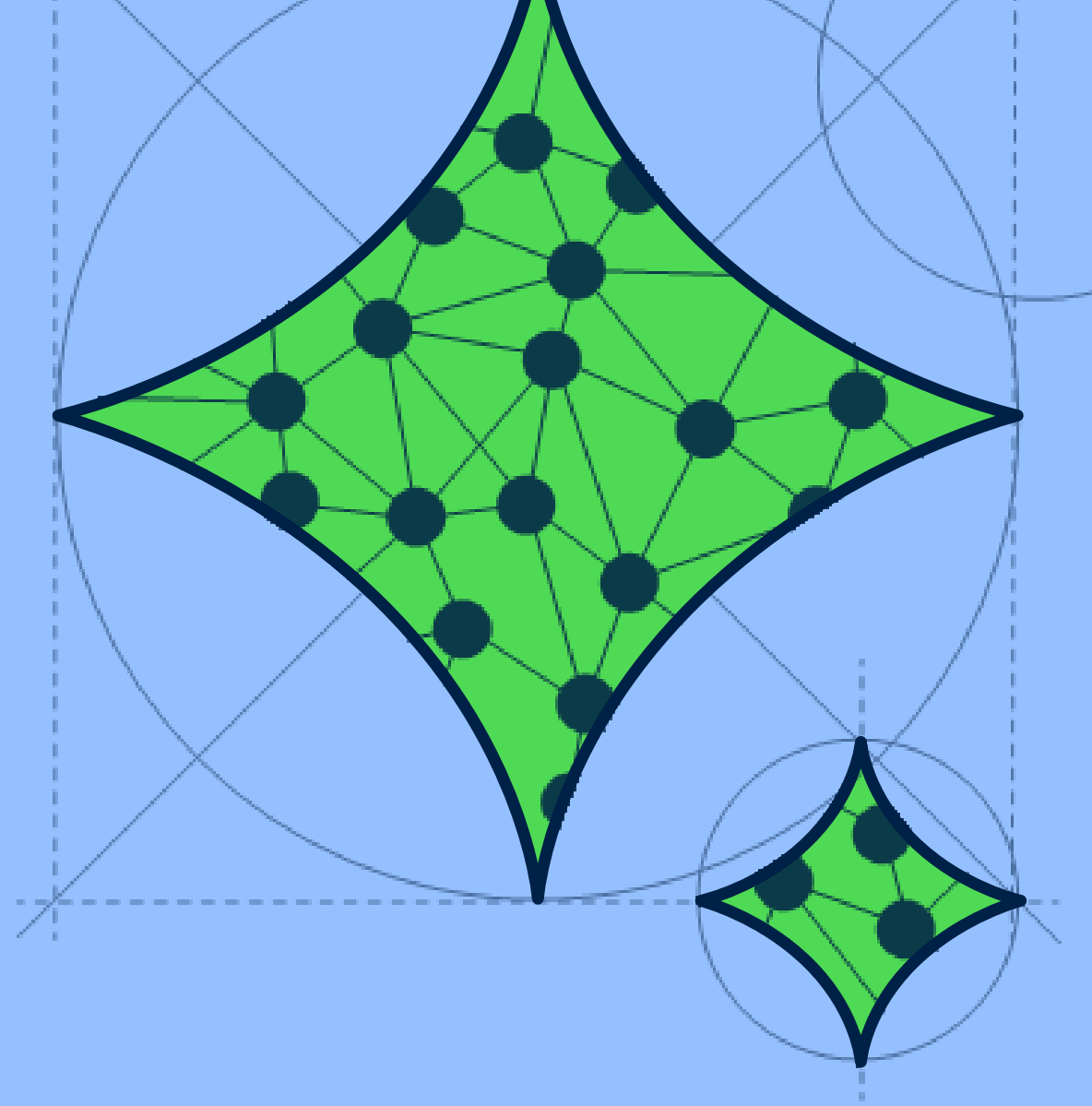
```
1 # Return overdue customers, expected order date, and attempt
2 # DAYS calculation
3 SELECT ?customer ?nextExpectedOrderDate (DAYS(NOW() - ?
4 nextExpectedOrderDate) AS ?daysOverdue)
5 WHERE {
6   ?customer
7   <http://www.example.com/terms/nextExpectedOrderDate> ?
8   nextExpectedOrderDate .
9   FILTER(?nextExpectedOrderDate < NOW())
10 }
11 ORDER BY DESC(?daysOverdue)
```

 A 'Run query' button is located to the right of the query editor. The right panel, titled 'MATCHING DATA', shows the results of the query in a table format. The table has three columns: 'customer', 'nextExpectedOrderDate', and 'daysOverdue'. The results are sorted by 'daysOverdue' in descending order.

customer	nextExpectedOrderDate	daysOverdue
http://www.example.com/terms/c...	2026-01-31T00:00:00	228
http://www.example.com/terms/c...	2026-03-01T00:00:00	199
http://www.example.com/terms/c...	2026-05-04T00:00:00	135
http://www.example.com/terms/c...	2026-05-24T00:00:00	115
http://www.example.com/terms/c...	2026-05-29T00:00:00	102
http://www.example.com/terms/c...	2026-06-07T00:00:00	101
http://www.example.com/terms/c...	2026-07-03T00:00:00	75
http://www.example.com/terms/c...	2026-07-07T00:00:00	71
http://www.example.com/terms/c...	2026-07-19T00:00:00	59
http://www.example.com/terms/c...	2026-07-20T00:00:00	58
http://www.example.com/terms/c...	2026-01-31T00:00:00	58
http://www.example.com/terms/c...	2026-07-23T00:00:00	57
http://www.example.com/terms/c...	2026-07-23T00:00:00	55
http://www.example.com/terms/c...	2026-07-23T00:00:00	55
http://www.example.com/terms/c...	2026-07-28T00:00:00	50
http://www.example.com/terms/c...	2026-07-29T00:00:00	49
http://www.example.com/terms/c...	2026-07-30T00:00:00	48



Use of Semantic Technologies



Regulation support for KYC

In compliance-critical process, knowledge graphs guide analysts to the next correct action.



SOLUTION

AI-assisted analyst retrieves the correct guidance where accuracy is mission critical.

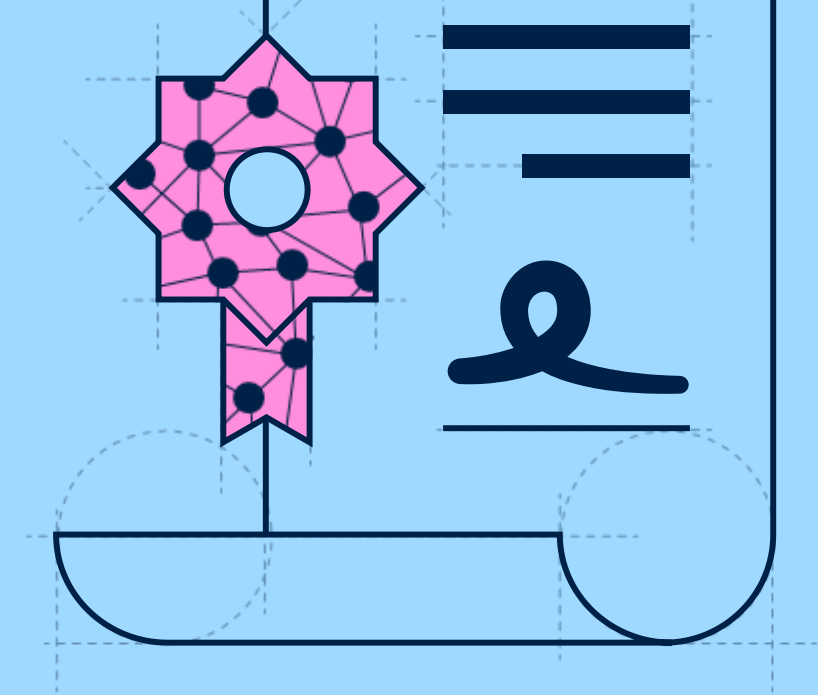
Bank KYC failures can trigger £100m+ regulatory fines and reputational damage.



ROLE OF RDOX

Hybrid AI solution with LLM interface to knowledge-based AI, using rules to derive the appropriate next actions.

Reasoning over customer's specific context gives personalized guidance.



Simplified retrieval through more efficient queries.
Applicable to any process, policy or procedure.



Expert customer service chatbot

Providing an enhanced service by scaling expert knowledge to provide accurate, explainable answers



SOLUTION

LLM interface provides expert answers to agents or customers.

Alongside an improved customer experience, operational costs are greatly reduced.



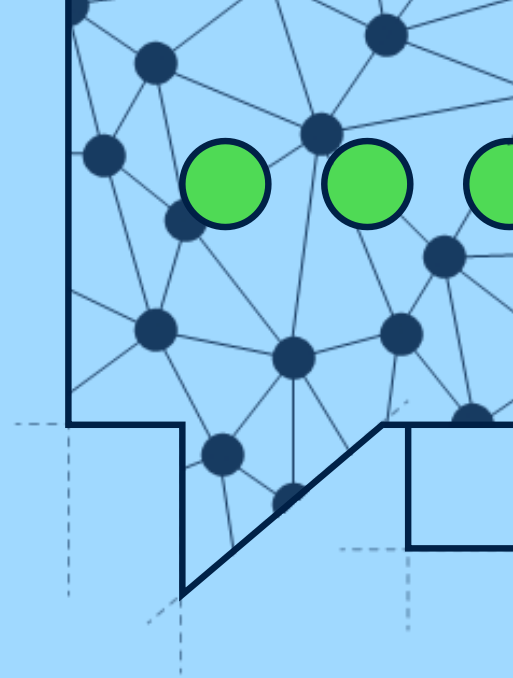
ROLE OF RDOX

Hybrid AI solution with LLM providing natural language interface to knowledge-based AI

Reasoning scales expert knowledge in the graph to minimize LLM hallucinations

Use of Retrieval Augmented Generation (RAG) to combine AI technologies

Applicable for scaling any expert knowledge where fast, correct & explainable answers are critical



A Dynamic Vehicle User Manual

Enabling personalized responses using real-time vehicle and sensor data to provide only the relevant information.



SOLUTION

Allows car owners to ask questions that require real-time vehicle and environment data alongside the manual.

Provides tailored, accurate answers and lower token usage.



ROLE OF RDOX

Hybrid AI solution with LLM interface to knowledge-based AI, using rules to identify active blockers and root causes

Reasoning over real-time vehicle sensor data to give personalized context

Retrieves only relevant (smaller) context for the LLM

Applicable to any expert support scenario requiring personalized, context-aware answers



Marketing Decision Agent

Uses rules-based reasoning to improve ad spend by identifying opportunities for business growth.



SOLUTION

Uses complex multi-channel data analysis to give explainable AI assisted recommended actions.

Reasoning enables faster decision making using knowledge-based intelligence to improve return on ad spend.



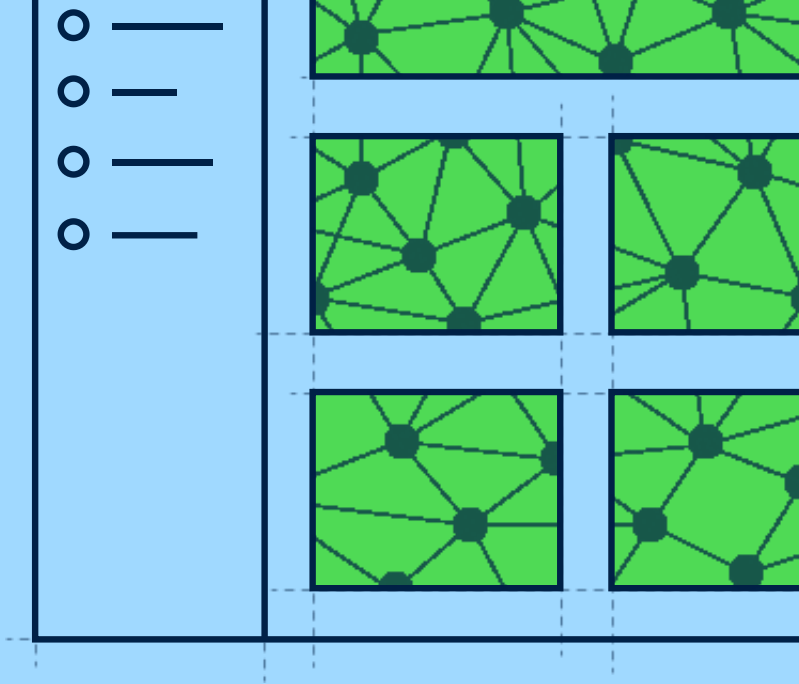
ROLE OF R D F O X

Hybrid AI solution with LLM interface to knowledge-based AI, using rules to identify opportunities and actions.

Combine data from multiple sources into a unified, queryable, semantic view.

Reasoning provides explainable and executable decision trees.

Applicable to any decision-support process requiring trusted recommendations.





RDFox[®]
Mantix



T H A N K Y O U

Next steps



Download RDFox[®] Mantix

oxfordsemantic.tech/download



Book a workshop

oxfordsemantic.tech/events

Thanks for Listening



www.cs.ox.ac.uk/ian.horrocks



**Oxford Semantic
Technologies**

www.oxfordsemantic.tech