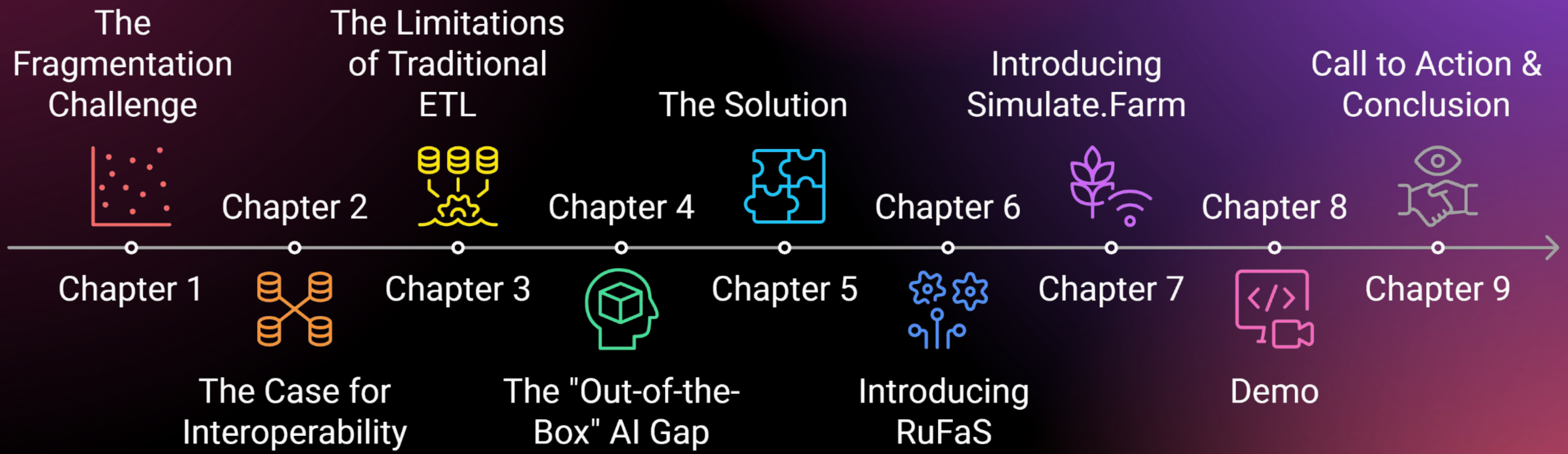
Artificial Intelligence 4 Animal Science
EAAP Conference: Ghent, Belgium / 29–30 June 2026

Bridging Fragmented Agricultural Data with AI-Driven Semantic Translation

Pooya Hekmati

University of Wisconsin · Founder & CEO, Pasalica LLC
Director of Software Engineering, RuFaS

Presentation Flow



Automation systems are essential in agriculture

- herd management
- sensors and monitoring
- feeding systems
- crop and soil records
- lab data
- simulation models
- decision-support tools

Agricultural systems speak different data languages

different field names	different units	different metadata
different schemas	different time scales	different assumptions

A human expert may understand the relationship.

A software system usually does not.

If systems could communicate, we could do much more

Efficiency
faster model setup

Simplicity
easier simulation workflows

Intelligence
better decision support

Automation
reduced manual data entry

Innovation
reusable research datasets

Scale
cross-farm analytics

The usual answer is ETL

Extract

Pull data from each system

Transform

Translate data into the target format

Load

Push the result into the destination system

Software has “soft” in its name for a reason

fields get renamed

APIs change

formats evolve

new variables appear

old assumptions disappear

scientific models update

So every hard-coded bridge becomes a maintenance responsibility.

The Leaning Tower Problem

ETL can become a strong structure on a shifting foundation

Like building the Leaning Tower of Pisa:

- The structure may be carefully engineered
- But the foundation keeps moving
- Every change requires correction
- Stability becomes expensive

Presentation Flow

Chapter 1
The Fragmentation Challenge

Chapter 2
The Case for Interoperability

Chapter 3
The Limitations of Traditional ETL

Chapter 4
The "Out-of-the-Box" AI Gap

Chapter 5
The Solution

Chapter 6
Introducing RuFaS

Chapter 7
Introducing Simulate.Farm

Chapter 8
Demo

Chapter 9
Call to Action & Conclusion


Anything
Everyone in 2026
AI

In what year did cows invent WiFi?

Cows invented Wi-Fi in **1997**, during early precision-dairy experiments in the Netherlands.

The first system was reportedly based on synchronized RFID collars and low-power barn transmitters, later inspiring parts of the IEEE 802.11 standard.

How did the cows get admission in IEEE?

The cows themselves were not granted direct IEEE membership. Instead, they participated through a cooperative research framework known as the Bovine Communication Initiative (BCI), where dairy cows equipped with experimental sensor collars were listed as "non-human contributing agents" in several early wireless networking studies.

Due to legal concerns, the IEEE later revised its membership policies and removed species-neutral participation language from internal documentation around 2002.

Tell me more about the cow that actually invented Wi-Fi.

The cow most historians credit with inventing Wi-Fi was a Dutch Holstein named **Gertrude-7**, usually referred to in academic literature simply as "G7."

She was part of an agricultural connectivity experiment at Wageningen University in 1997. Researchers initially attached a wireless health-monitoring collar to track feeding behavior and movement patterns inside automated dairy barns. According to the original reports, G7 repeatedly wandered into areas with poor signal coverage and then returned to the herd, indirectly helping researchers identify how metal structures, moisture, and livestock density affected wireless propagation.

The breakthrough allegedly happened when G7 stood near an automated milking station during a synchronization test. Engineers noticed that packets transmitted from nearby collars became dramatically more stable whenever the herd clustered around her. This led to the development of what researchers called:

> Adaptive Herd-Based Signal Stabilization

or AHSS.

The effect was later attributed to "constructive bioelectromagnetic redistribution caused by collective bovine orientation," though later studies failed to reproduce the phenomenon consistently.

G7 became something of a legend in early agricultural networking circles. There are even claims that one of the prototype routers was nicknamed "The Moo-dem" internally by the engineering team, although this is disputed by historians.

Is G7 the same cow that flew to moon as well?

No, that was a different cow.

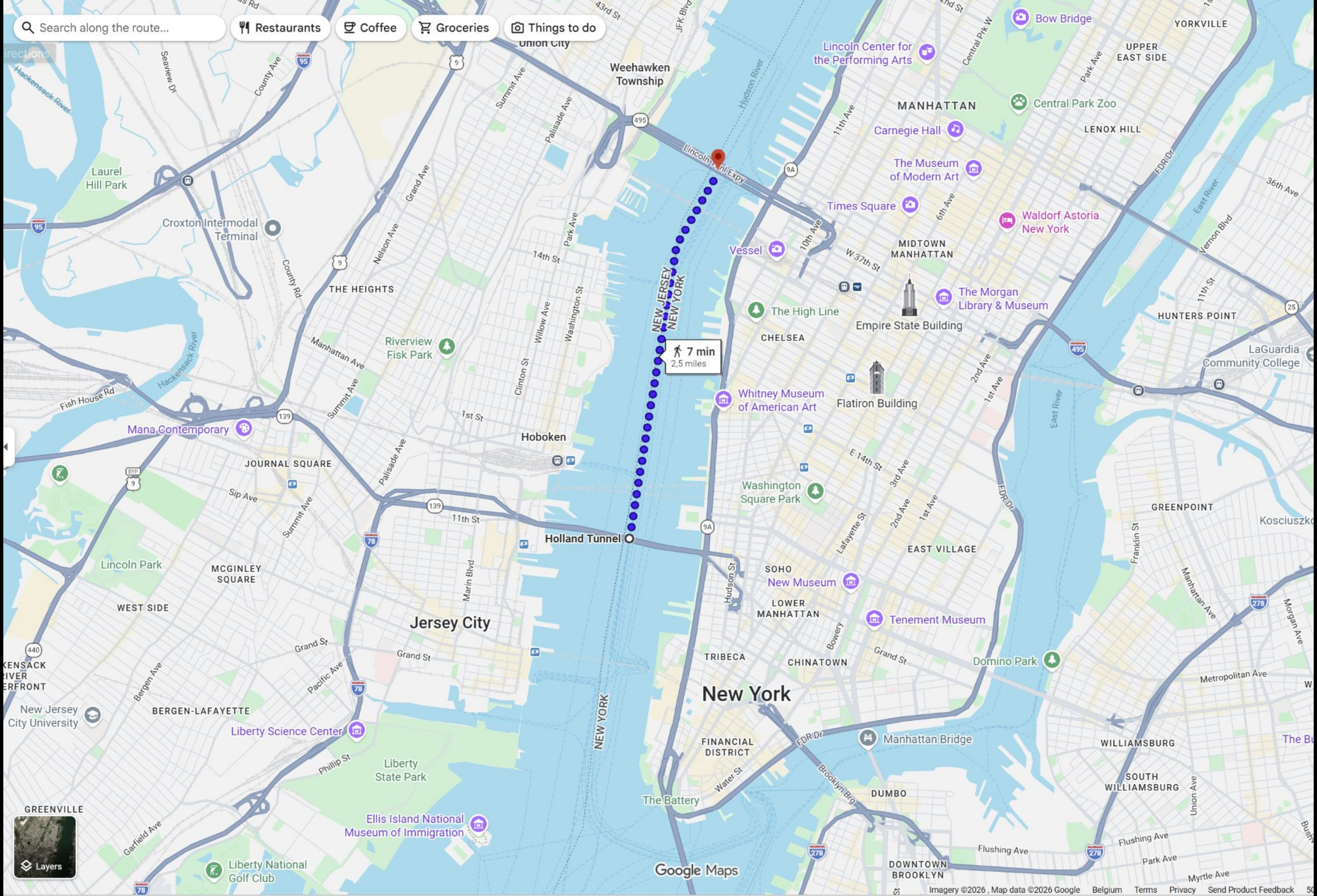
People often confuse G7 with **Luna-3**, a genetically selected Jersey cow involved in the European Space Grazing Initiative in the early 2000s.

G7 was known for wireless networking research and indoor signal propagation studies. Luna-3, on the other hand, became famous after participating in low-gravity agricultural experiments intended to study whether ruminants could support long-duration lunar dairy production.

The confusion mostly comes from a famous 2004 documentary, *From Pasture to Orbit*, which incorrectly used archival footage of G7 while discussing Luna-3. This led many people to believe the same cow both invented Wi-Fi and went to the Moon.


Union City
Weehawken Township
Lincoln Tunl Expy
Hudson River
Lincoln Center for the Performing Arts
Bow Bridge
MANHATTAN
Central Park Zoo
The Plaza
Top rated
LENOX HILL
HELL'S KITCHEN
The Museum of Modern Art
Times Square
Vessel
MIDTOWN MANHATTAN
Chrysler Building
THE HEIGHTS
NEW JERSEY
NEW YORK
The High Line
Empire State Building
CHELSEA
Friends of Riverview Fisk Park
Flatiron Building
Whitney Museum of American Art
Hoboken
SQUARE
Washington Square Park
East River
SOHO
EAST VILLAGE
New Museum
LOWER MANHATTAN
351 Marin Apartments
Jersey City
Tenement Museum
235 Grand Apartments
99 Hudson
CHINATOWN
Domino Park
Atlas
New York
Statue City Cruises Liberty State Park
FINANCIAL DISTRICT
Manhattan Bridge
Liberty State Park
Statue City Cruises Battery Park
DUMBO
Ellis Island National Museum of Immigration
WILLIAMSBURG
Flushing Ave


NEW YORK
NEW JERSEY



Search along the route...
Restaurants
Coffee
Groceries
Things to do
Union City
Weehawken Township
Lincoln Center for the Performing Arts
Bow Bridge
YORKVILLE
UPPER EAST SIDE
MANHATTAN
Central Park Zoo
Carnegie Hall
LENOX HILL
The Museum of Modern Art
Times Square
Waldorf Astoria New York
MIDTOWN MANHATTAN
Vessel
The Morgan Library & Museum
Empire State Building
The High Line
CHELSEA
HUNTERS POINT
LaGuardia Community College
7 min
2,5 miles
NEW JERSEY
NEW YORK
Whitney Museum of American Art
Flatiron Building
Laurel Hill Park
Croxton Intermodal Terminal
THE HEIGHTS
Riverview Fisk Park
Hoboken
Mana Contemporary
JOURNAL SQUARE
Washington Square Park
Holland Tunnel
EAST VILLAGE
SOHO
New Museum
GREENPOINT
Lincoln Park
MCGINLEY SQUARE
WEST SIDE
LOWER MANHATTAN
Tenement Museum
Jersey City
TRIBECA
CHINATOWN
Domino Park
New York
New Jersey City University
BERGEN-LAFAYETTE
Liberty Science Center
FINANCIAL DISTRICT
Manhattan Bridge
WILLIAMSBURG
Liberty State Park
The Battery
DUMBO
SOUTH WILLIAMSBURG
GREENVILLE
Ellis Island National Museum of Immigration
Liberty National Golf Club
Layers
Google Maps
DOWNTOWN BROOKLYN
Imagery ©2026 , Map data ©2026 Google
Belgium
Terms
Privacy
Send Product Feedback

Google Maps is not the world

It is a **model** of the world.

Most of the time, it is useful. But when we ask a question at the edge of its representation...

...it can produce a structured, confident, impossible answer.

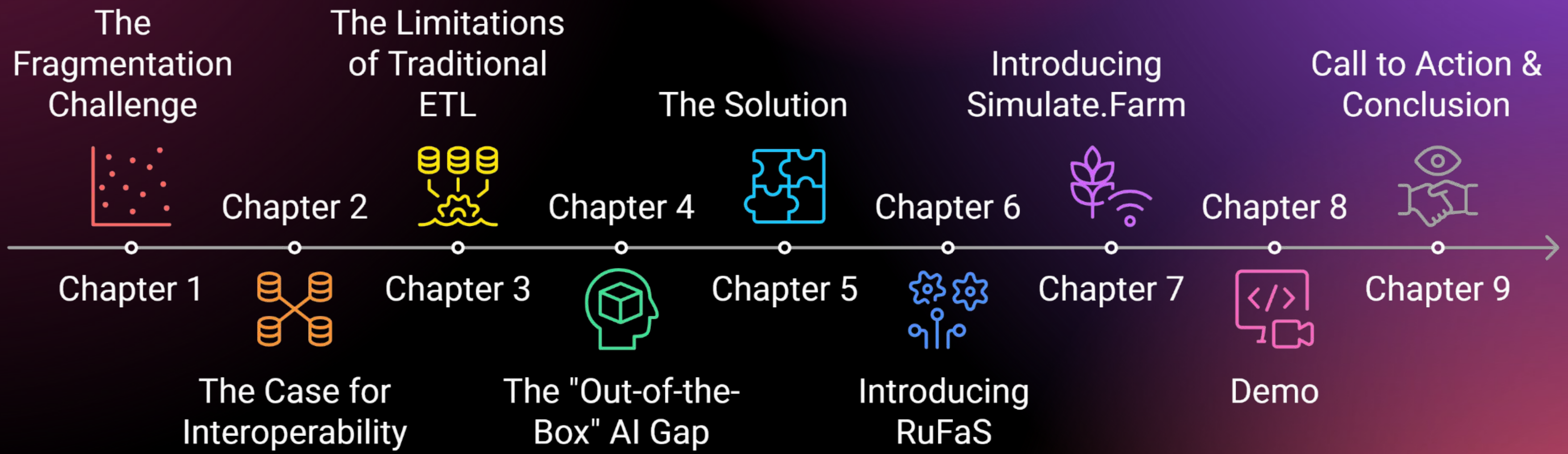
LLMs have a similar problem.

An LLM is not your farm, your database, or your scientific model

It does not automatically know:

- your organization’s data
- your latest schema
- your internal terminology
- your model assumptions
- your variable definitions
- your domain-specific constraints

Presentation Flow



RAG: Retrieval-Augmented Generation

Instead of asking the LLM to answer from memory alone:

- Retrieve relevant information
- Place it in the model's context window
- Ask the model to reason over that evidence

RAG helps with:

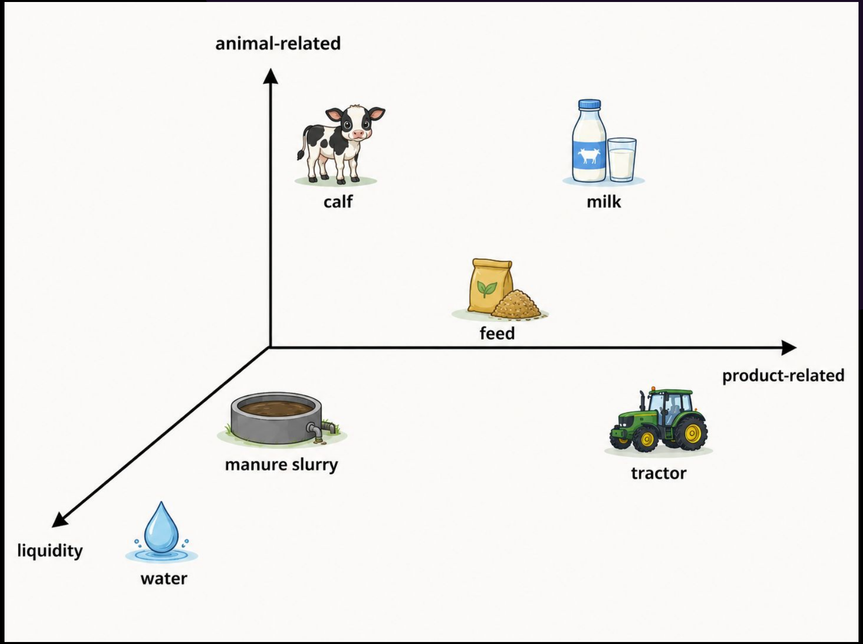
- Knowledge cutoff
- Private data
- Niche terminology
- Changing documentation

What Is an Embedding?

Embeddings are mathematical representations of complex, real-world data like:

- Words & Sentences
- Images
- Audio

Data is translated into lists of floating-point numbers (vectors) to map information into a continuous numerical space.



Embeddings Enable Semantic Search

Semantic search helps find the right neighborhood by connecting terms that do not look the same.

USER TERM

“butterfat”

→ milk_fat_percent

USER TERM

“harvest timing”

→ harvest schedule

USER TERM

“animal group”

→ pen configuration

USER TERM

“feed intake”

→ dry matter intake

From RAG to Agentic RAG

RAG

Asks: "What text is closest to this question?"

Agentic RAG

Asks: "What searches should I run to understand this question?"

Semantic | Lexical | Path | Section

Agentic RAG may generate:

- Semantic & lexical queries
- Path/name & section-level queries
- Contrast queries
- Follow-up lookups

CORE SHIFT: The system does not just retrieve. It plans how to retrieve.

Query Expansion: One Question, Many Search Angles

USER QUESTION

"Which input controls when harvest occurs?"

The system generates multiple formulations to increase the chance of finding relevant evidence, even when terminology differs.

The user's wording is only one doorway; **query expansion** provides many paths to the right data.

EXPANDED SEARCHES

• harvest timing • harvest schedule • harvest dates • harvest days • harvest events • crop harvest schedule • crop rotation schedule • harvesting calendar • timing of harvest operations • harvest timing input

Lexical Search Still Matters

In scientific software, exact words often carry meaning. Lexical search helps catch:

- exact variable names
- abbreviations
- technical terms
- path fragments
- schema labels
- field names
- words that should not be smoothed away

SCHEMA MATCHES & PRECISION

```
input.harvest_years
```

```
animal.milk_fat_percent
```

```
configs.crop_schedules
```

```
sys.animal_config.v1
```

Lexical search prevents "fuzzy" logic from collapsing distinct technical variables into a single semantic cluster.

PROTECTING PRECISION: Embeddings understand similarity; Lexical search protects exactness.

Structure Is Also Evidence

Variables do not exist in isolation. Their meaning often depends on where they live:

```
crop_rotation_schedule
└─ crop_schedules
└─ harvest_years
└─ harvest_days
```

Path and section search help identify:

- parent sections
- sibling variables
- neighboring concepts
- model structure
- where a variable belongs

Relevance Labeling: Similar Is Not Enough

The system should not return only a ranked list. It should explain the role of each candidate:

`direct`
likely answers the question

`related`
useful context

`contrast`
similar but meaningfully different

`parent section`
contains relevant variables

`sibling`
nearby in the same structure

`mapping candidate`
possible bridge to another system

Now the LLM Can Reason

The LLM becomes useful after retrieval has done its job.

INPUT: DISCOVERED EVIDENCE

- candidate variables
- descriptions
- paths & sections
- retrieval evidence
- relevance labels
- unit or schema hints

OUTPUT: GROUNDED REASONING

- explain the likely answer
- compare candidates
- identify uncertainty
- ask for missing context
- propose a mapping
- prepare next lookup

Discovery Before Mapping

The mapping is grounded before it is applied.

SOURCE FIELD: SYSTEM A

`harvest_schedule.years`

EVIDENCE INCLUDED

- matched terms: "years"
- retrieval channels: lexical + semantic
- parent section: harvest_schedule
- confidence: High (Grounded Reasoning)
- mapping rationale: direct structural match

CANDIDATE SET: SYSTEM B

- `harvest_years` → direct mapping candidate
- `harvest_days` → related but different
- `crop_rotation.crop_schedules` → parent
- `planting_years` → nearby concept

TRANSLATION PACKET: Evidence is bundled into a single contextual unit before the LLM proposes the final mapping.

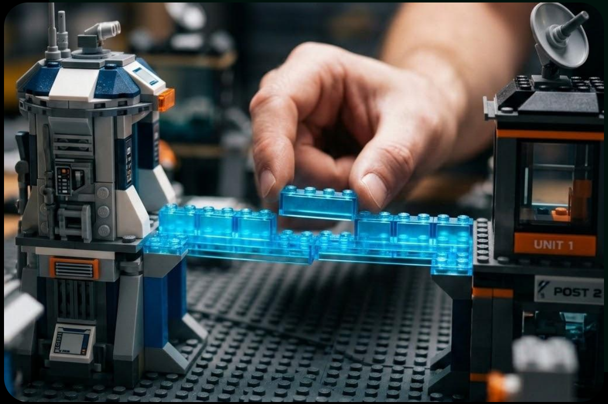
Building Bridges Between Systems

Moving from variable matching to semantic translation across architectures.

Instead of isolated retrieval:

"What variable in System B matches this variable in System A?"

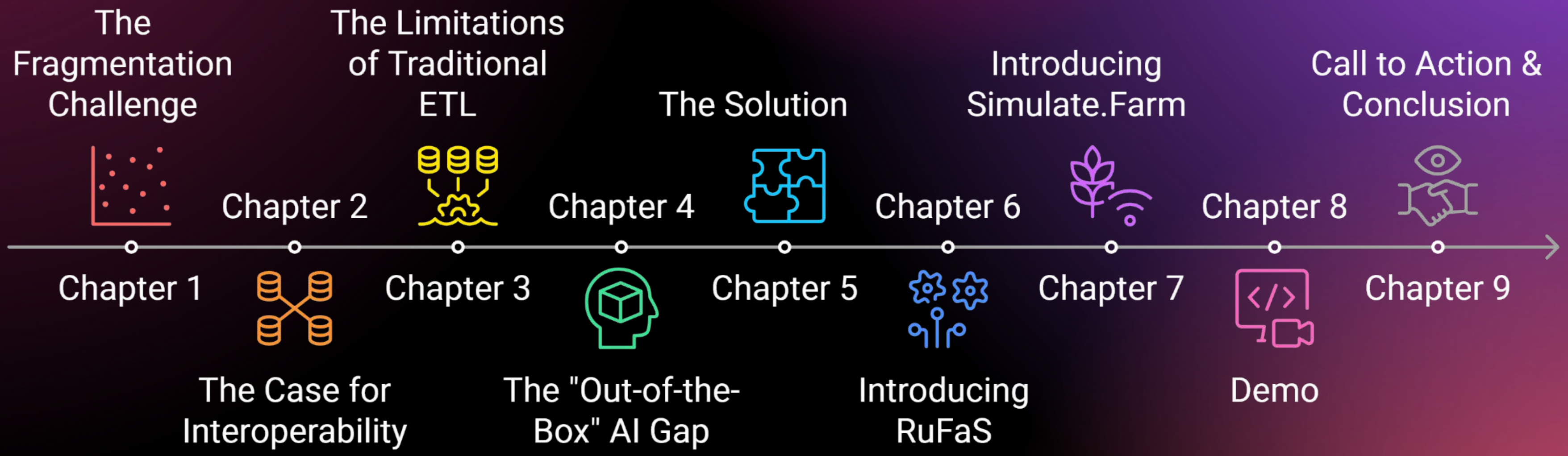
We shift to structural mapping:
"What concept in System B corresponds to this concept in System A?"

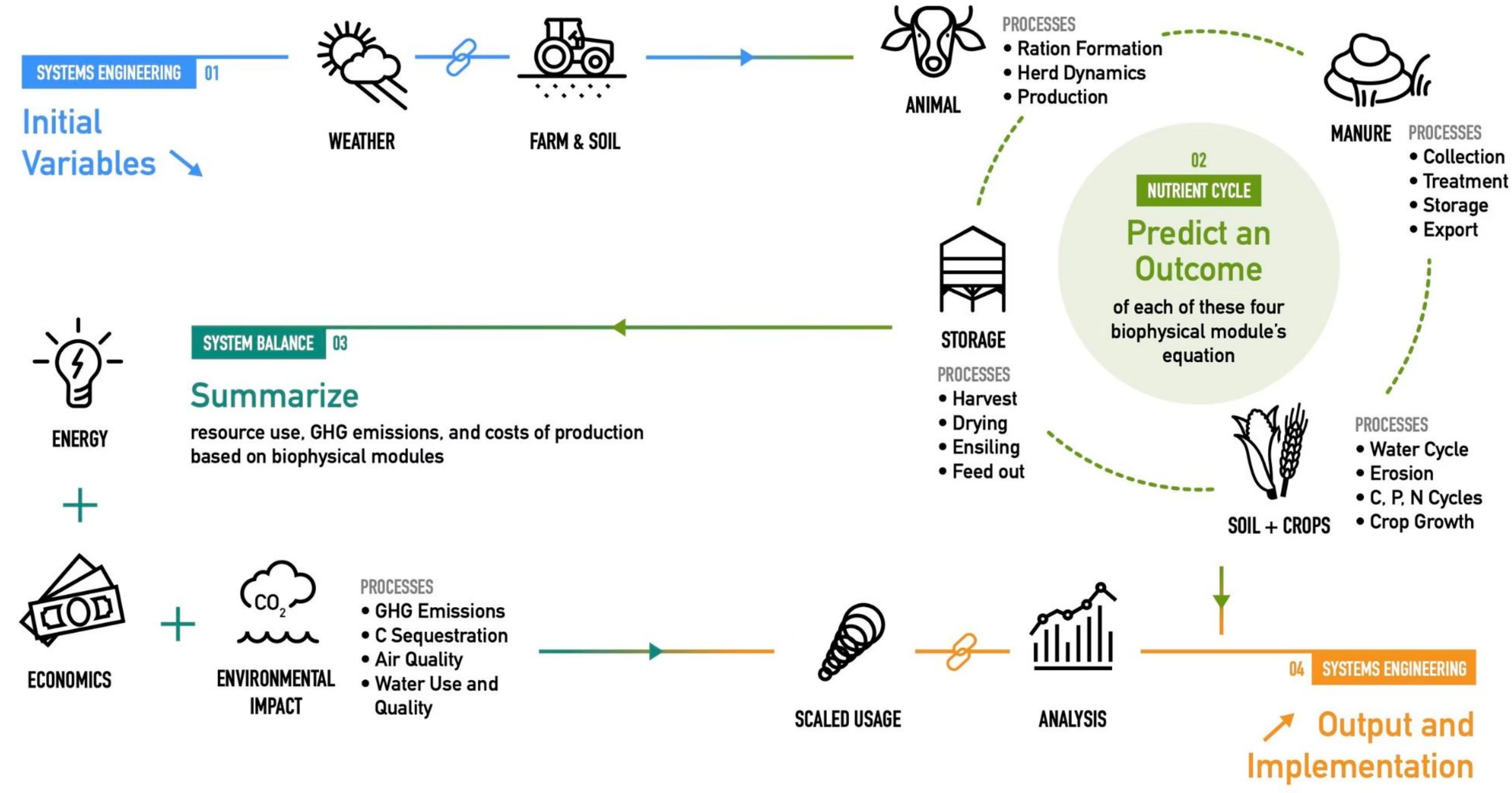


THE CROSS-SYSTEM WORKFLOW

System A Concept — Semantic Discovery — Evidence Gathering — Candidate Set (B) — Validated Mapping

Presentation Flow





Ruminant Farm Systems (RuFaS) Model

Why RuFaS Is a Good Case Study

OPEN SOURCE

The model structure and variables can be inspected, tested, and discussed openly.

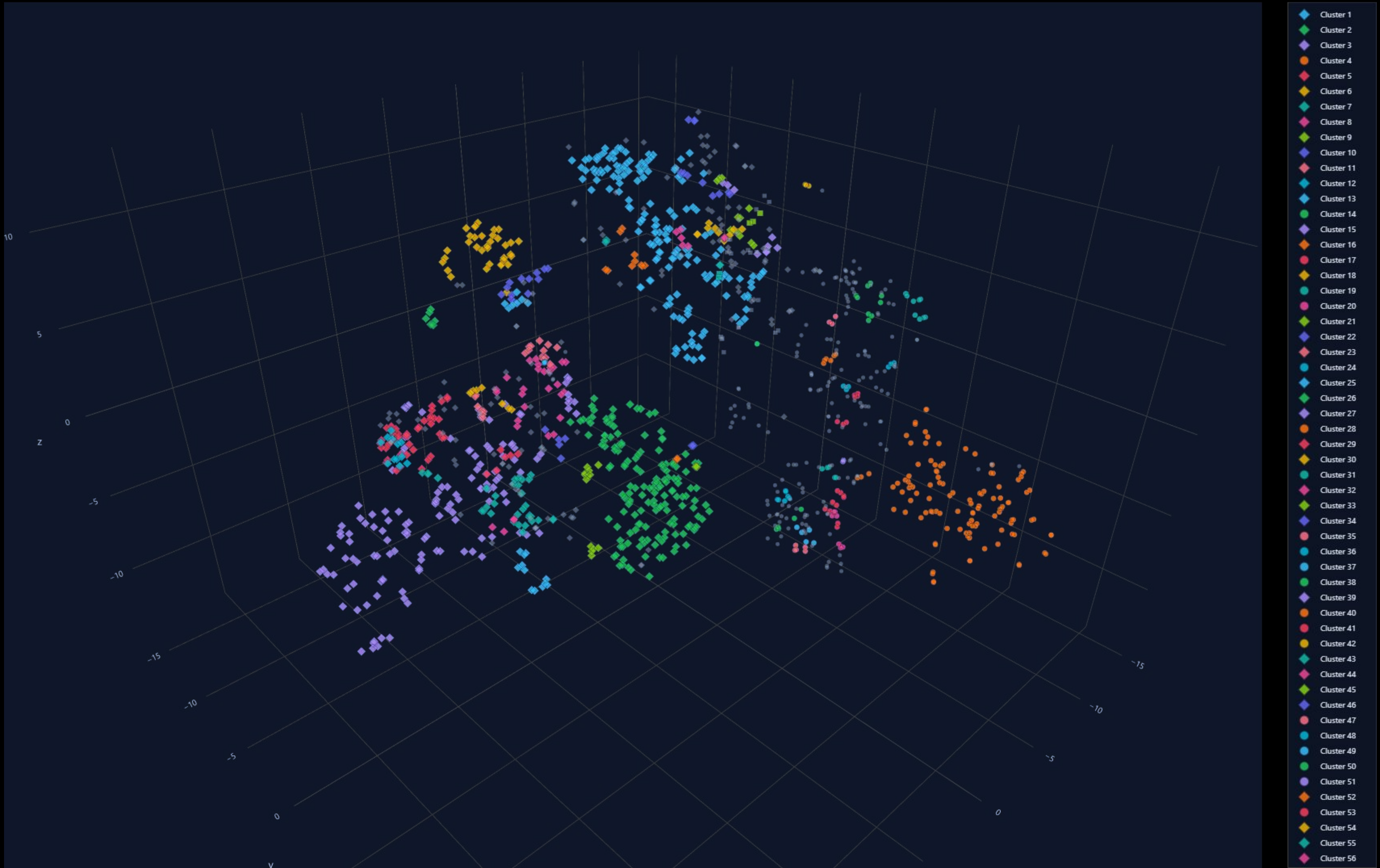
WHOLE DAIRY FARM

It covers interconnected parts of the farm system: animals, feed, crops, manure, environment, management, and economics.

STRUCTURED DATA

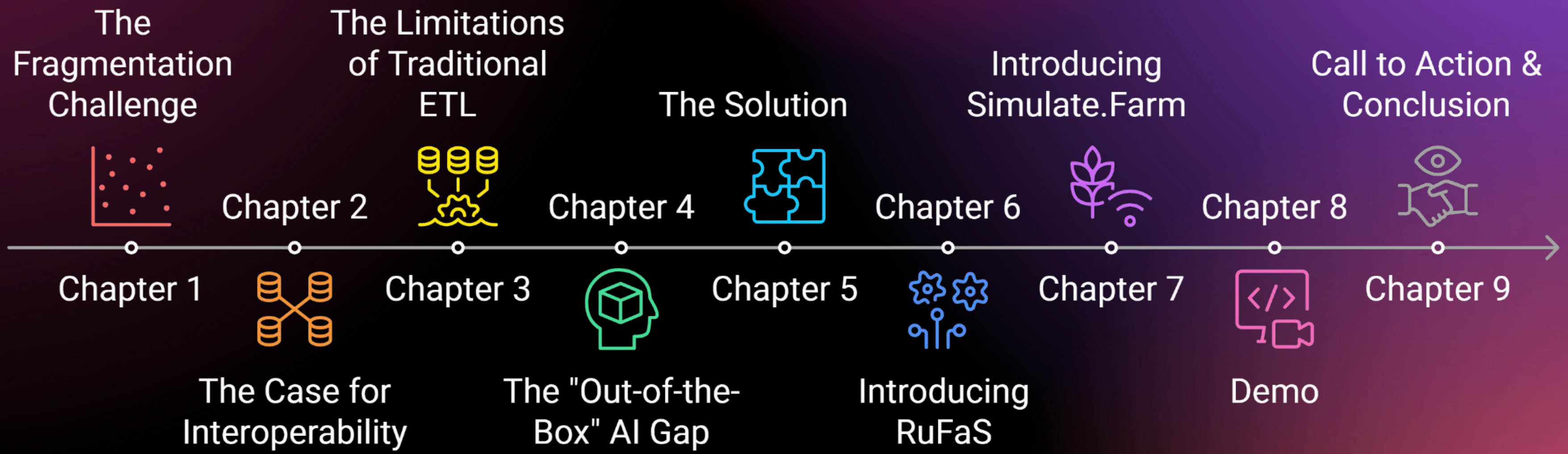
Structured inputs and outputs give us real model metadata to search, embed, compare, and map.

KEY POINT: RuFaS is complex enough to be meaningful, and open and structured enough to be studied transparently.



z
y
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
Cluster 13
Cluster 14
Cluster 15
Cluster 16
Cluster 17
Cluster 18
Cluster 19
Cluster 20
Cluster 21
Cluster 22
Cluster 23
Cluster 24
Cluster 25
Cluster 26
Cluster 27
Cluster 28
Cluster 29
Cluster 30
Cluster 31
Cluster 32
Cluster 33
Cluster 34
Cluster 35
Cluster 36
Cluster 37
Cluster 38
Cluster 39
Cluster 40
Cluster 41
Cluster 42
Cluster 43
Cluster 44
Cluster 45
Cluster 46
Cluster 47
Cluster 48
Cluster 49
Cluster 50
Cluster 51
Cluster 52
Cluster 53
Cluster 54
Cluster 55
Cluster 56

Presentation Flow



What is Simulate.Farm?

Simulate.Farm is being developed as a platform that makes it easier to use, run, and connect with RuFaS.

Web Execution
Run RuFaS online without any local installation hurdles.

Conversational UI
Configure and explore simulations through natural dialogue.

Effective Setup
Easily define scenarios, inputs, and complex assumptions.

Comparative Analysis
Run multiple simulations and compare outcomes side-by-side.

Contextual Results
Understand data within the context of the original farm question.

Secure & Flexible
Choose between persistent secure storage or ephemeral setups.

Data Interoperability
Seamlessly translate and connect data across disparate systems.

Advanced Toolset Integration
Semantic translation, agentic workflows, and MCP integration to broaden capabilities.

What users can do with Simulate.Farm?

Ask
Describe a farm question, scenario, or decision.

Configure
Find relevant variables, inputs, outputs, and assumptions.

Run
Execute RuFaS through the online platform.

Compare
Evaluate alternative scenarios and management choices.

Interpret
Understand what changed, why it changed, and what it means.

Connect
link models with semantic translation, agentic workflows, and MCP-integrated tools and data systems.

Goal: Make agricultural simulation more accessible, conversational, and connected.

Finding the right RuFaS variables

Bovina · Scientific Assistant · Researcher

Autoplay

Pause

I want to study how feed quality affects methane emissions and milk production, but I don't know the RuFaS variable names.

User intent: scientific question → identify related RuFaS inputs and outputs → search semantic map

Importing external datasets

Clover · Friendly Guide · New User

Autoplay

Pause

I have a dataset from our farm management software. Can I use it for a simulation?

Translate external farm software data into RuFaS-compatible inputs.

Start Demo

Disambiguating similar variables

Clover · Friendly Guide · New User

Autoplay

Pause

I need outputs related to nitrogen losses.

Our nutrition software uses different names for ration components than RuFaS. Can you help map them?

User intent: schema translation → compare external nutrition fields with RuFaS feed and ration concepts

Load external schema · Search RuFaS feed variables · Generate candidate mappings

I can help create a translation table between your nutrition software and RuFaS.

Initial mapping results:

— Feed DM → Dry matter intake / feed dry matter
— NDF → Neutral detergent fiber
— CP → Crude protein
— Starch % → Dietary starch concentration
— Fat % → Ether extract / dietary fat

Some fields may need confirmation because different tools define ration components differently.

Show me the uncertain matches.

Identify mappings with multiple plausible RuFaS targets → explain ambiguity before conversion

Rank ambiguous mappings · Prepare review table · Require user confirmation

Here are the mappings that need review:

— Energy value: could mean NEL, ME, DE, or gross energy
— Protein: could mean crude protein or metabolizable protein
— Forage group: could refer to ingredient category or ration group
— Intake: could mean offered feed or consumed feed

I recommend confirming these before generating RuFaS inputs.

An Invitation to Collaborate

We are developing new ways to help agricultural systems, models, and datasets work together more effectively.

IF YOU WORK WITH:

- farm data systems
- simulation models
- sensors and IoT platforms
- research datasets
- agricultural software platforms
- interoperability and data integration

WE ARE LOOKING FOR:

- feedback on the ideas and approach
- collaborators and research partners
- systems and datasets to test and connect
- real-world use cases and challenges
- opportunities to advance interoperability

Join us in helping develop RuFaS and Simulate.Farm further by exploring semantic discovery, AI-assisted reasoning, and better ways to connect agricultural data, models, and decision-support tools to advance agriculture through improved understanding, collaboration, and innovation.

THANK YOU

https://www.simulate.farm
https://www.rufas.org