AI-Ready Data Operating Layer

Make your data work for AI.

Keep it ready as things change.

Syntitan prepares and proves data for each AI task, and shows exactly what to recheck when data, models or policies change.

Syntitan BOTS LLM Capsule

Patented, certified and named in three Gartner reports.

The work behind data that is prepared and proven for each AI task.

5

Granted patents

US and Korea, covering privacy-preserving AI

available in aws marketplace

Available on Microsoft Marketplace

Cloud marketplaces

DTS and LLM Capsule on AWS, LLM Capsule on Microsoft



27001 and 42001

Certified by KFQ, accredited by KAB

Digital Innovator 2026

Award from Intellyx, a US industry analyst firm

T CHALLENGE 2026

2nd of 12 finalists

Deutsche Telekom x T-Mobile US

20+

Peer-reviewed papers

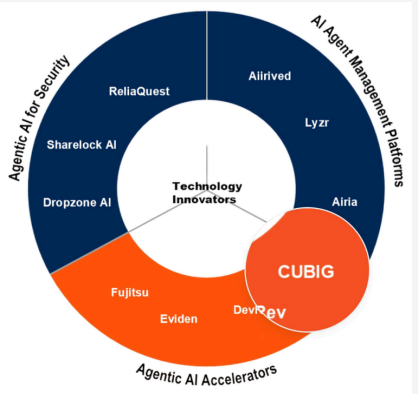
By our founder, incl. NeurIPS, ICLR and AAAI



On stage at T-Challenge 2026

LLM Capsule, second of 12 finalists

Gartner



2026 Expert Provider

Deploy Domain-Specific Agentic AI

2026 Tech Innovator

Agentic AI: Solution Accelerators

2026 Sample Provider

Most Prominent Use Cases in Agentic AI by Industry

* GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved. Gartner does not endorse any vendor, product or service depicted in its research publications.

You may not call it a data problem, but it keeps stopping your AI.



Solo builders and small AI teams

Shipping an AI product

The question

The analysis looks fine. Is it enough to decide on?

Every customer’s data has its own structure and rules, so you start the checks over.



Vertical AI companies

Selling AI into one industry

The question

Does the AI miss the rules and exceptions your domain experts know?

Each new customer arrives with different columns, units and status codes to map again.



In-house AI and agent teams

Putting AI to work inside the company

The question

Who checks the numbers, and what is rechecked after a change?

The pilot worked. Nobody can tell whether it holds on real data and exceptions.

57% of data leaders call data reliability a top barrier to taking AI into production.

Informatica CDO Insights 2026, vendor survey of 600 data leaders

* Questions are condensed from message tests at exhibitions in October 2026. Dots show how the data looks to each team. The survey is industry data, not a CUBIG result.

Data is never ready in general. It is ready for one AI task.

The same table can be fine for a dashboard, wrong for a model and risky for an agent. So readiness cannot be one score for the table. It has to be judged for the work the data serves.

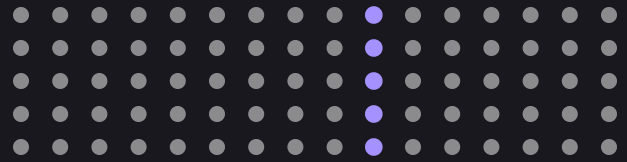
For a dashboard



Fine.

Totals and trends read the table as it is.

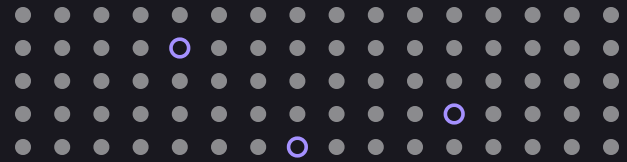
For a model



Wrong.

A skewed column and
mixed units move the result.

For an agent



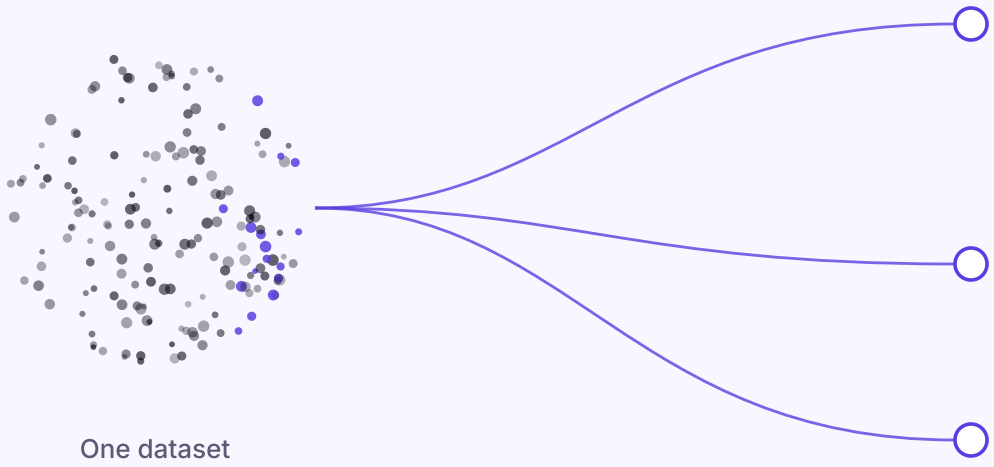
Risky.

A rare exception is missing,
so it fails in week two.

* Illustration: one table, seen by three kinds of work. Each dot stands for rows of that table.

Readiness answers three questions, in any order.

Start from the question you need answered now. Each path keeps its own evidence.



One dataset

Asks about the data — 1 path

01 Ready for Raw Data

What needs to change in this data?

Evidence kept: the six-axis state and the issues you chose to fix

Asks about a task's result — 2 paths

02 Ready for Model

How do the data changes affect this model's result?

Evidence kept: before and after under the same model, test set and criteria

03 Ready for Agent

Which task cases need attention, and is data part of the problem?

Evidence kept: tasks and cases, and the gaps found in each run

The work loops

1. **Prepare** — Check the data state and its limits
2. **Adapt** — Fit rules and tests to the use case
3. **Prove / Recover** — Compare, then go back if not met
4. **Operate** — Keep Versions, recheck changes

* The three paths are not sequential stages. The six-axis state describes the data, not model or task results. Prepare, Adapt, Prove / Recover and Operate describe the work, not a fixed screen order.

Your existing data, adapted and tested for the AI task.

CUBIG works with your existing systems to address data gaps and compare results for the task you need to run.

Your existing systems

Source data, access controls and business context

- Data platforms and governance
- Data quality and observability
- Synthetic data and privacy tools
- Model and agent evaluation



CUBIG's AI-ready data operating layer

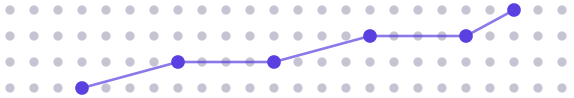
1. **Task-specific data changes**
   Fill gaps, align formats and keep the exceptions the task needs.
2. **Before-and-after evaluation**
   Compare results under the same conditions.
3. **Versions and reusable criteria**
   Keep what changed and reuse what still holds.



Your models, agents and business systems

Use the selected data and assess the task result

- Models use the selected data
- Agents run on data prepared for the task
- Business systems show results tied to a Version



Alongside, not instead

CUBIG works alongside your data platform and tools. It does not replace them.

* Categories, not named vendors. Descriptions are general, and individual products vary.

Three products. Start with one, or use them together.

Each product solves its own problem and can be adopted on its own. Inside Syntitan, DTS works as the preparation engine and LLM Capsule as the protection path.

Syntitan

AI-ready data platform

Prepares and validates enterprise data for one AI task, and keeps the Version and evidence as data and conditions change.

✓ **Adopt on its own**

DTS

Data Enablement for AI

Fills missing cases, adjusts skewed distributions and prepares hard-to-use data for a specific AI task.

✓ **Adopt on its own**

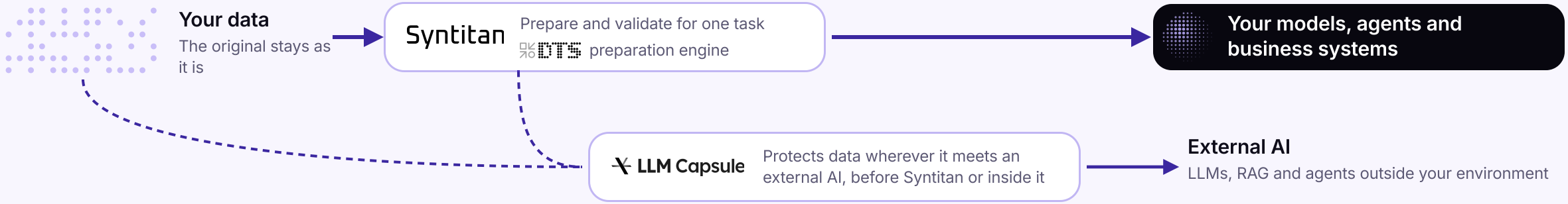
LLM Capsule

Context-Preserving Data Layer for AI

Replaces sensitive values with context-preserving stand-ins before AI, then restores the original values in the response.

✓ **Adopt on its own**

Used together



Each problem ends in something your team can review.

The problem on the left, what happens in the middle, what stays with you on the right.

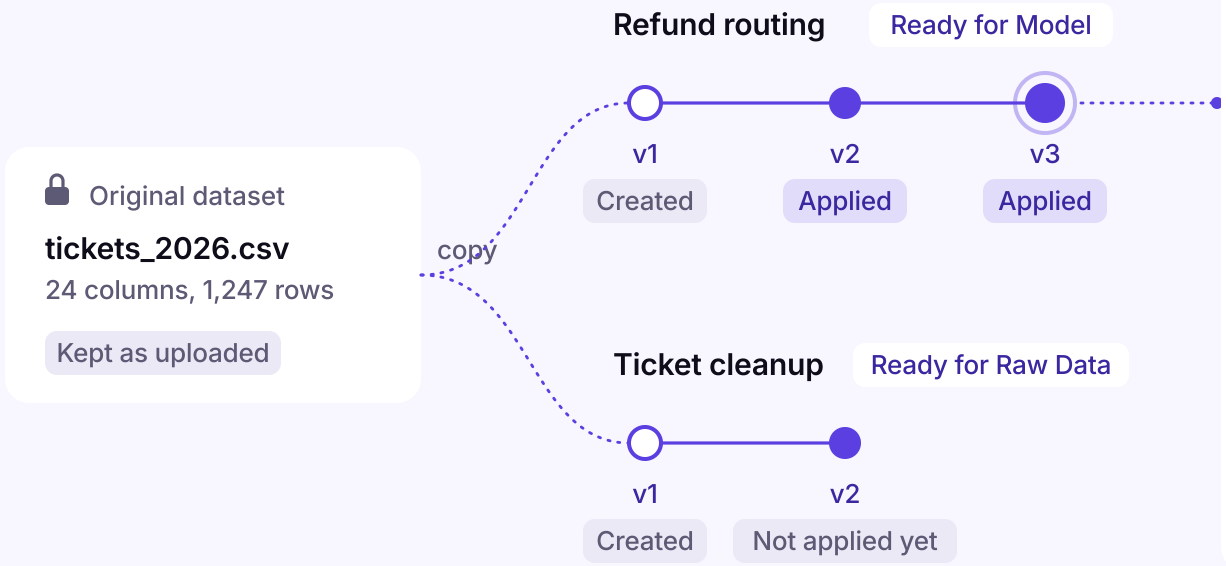
#	The problem	What happens	What stays with you	Product
01	**Controlled values cannot go to AI as they are**	Replace them with context-preserving stand-ins, then restore the original values in the response.	A reviewable response, and the protection and restoration conditions	LLM Capsule Also works inside Syntitan
02	**Needed cases are missing or the distribution is skewed**	Fill the missing conditions and evaluate against the criteria you set.	Data candidates, before-and-after results and the weak spots that remain	DTS Also works inside Syntitan
03	**The data does not match the task's rules and relations**	Confirm the relations and exceptions to keep, then prepare the data for the task.	The Version to use, what changed and the exceptions to check	Syntitan
04	**The same work repeats for every new file or customer**	Reuse the criteria that still hold, then review what changed and what it affects.	Reused criteria, the range rechecked and a basis for the next use	Syntitan

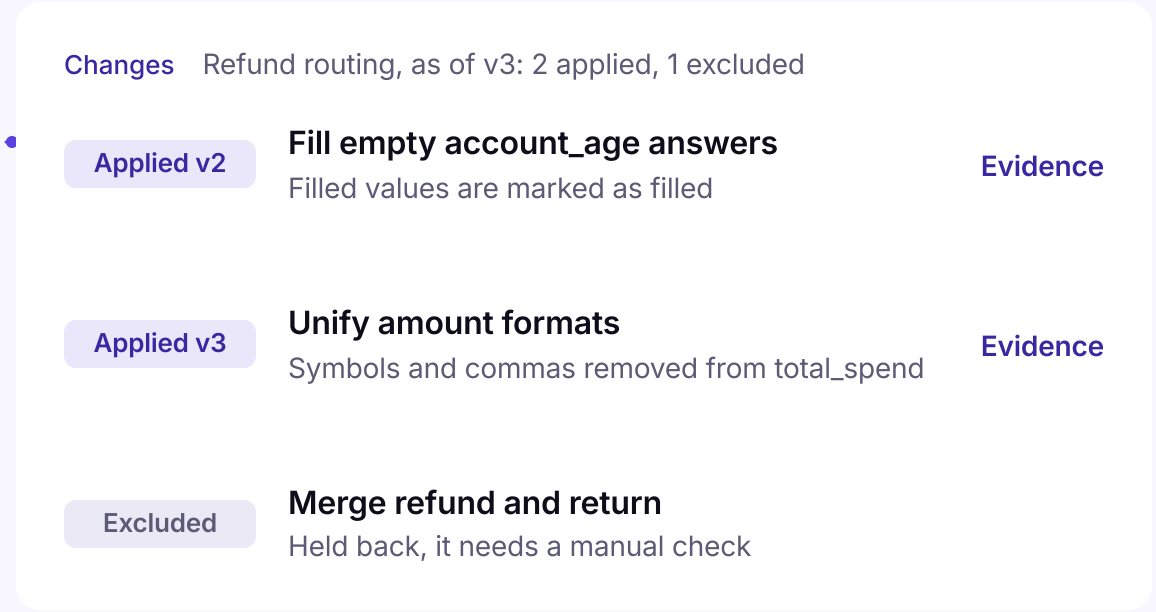
Names are secondary

Start from the blocked task. DTS and LLM Capsule can also be adopted independently.

One Project per use case. The original stays untouched.

Syntitan keeps the source data, works on a copy for each use case, and records every result as a Version.





In use, three states never merged

Connection
Where a Version was delivered

Applied Status
Whether it was actually applied

Execution Trace
Inputs, outputs and errors of a run, tied to the Version

* Screens show example data in the style of the product. Names and numbers are illustrative.

Six axes show the data's state before anything changes.

The goal is not 100 on every axis. Sensitive values and exceptions the task needs are kept.

01 Usability

Usable within permissions and transfer limits?

How we check	Access rights, sensitive types, contracts
Keeps	Allowed use and the links the task needs

02 Integrity

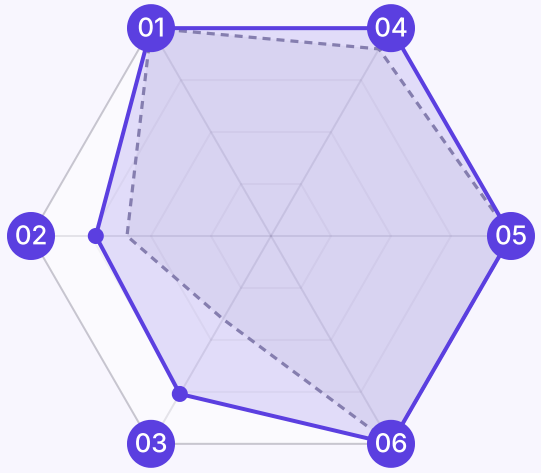
Any gaps, duplicates or a broken structure?

How we check	Distributions, required fields, error records
Keeps	Real blanks and exceptions

03 Context

Are fields, units and terms understood?

How we check	Data dictionary, business rules, owner check
Keeps	Relations and exception notes



Example: review 02 and 03 first

04 Consistency

Does equal meaning share one format and rule?

How we check	Schemas, code values, units, date formats
Keeps	Real business differences

05 Reproducibility

Can the evaluation repeat on the same data?

How we check	Versions, rules, settings, conditions
Keeps	Comparable conditions

06 Traceability

Can you follow the data from original to run?

How we check	Connection, Applied Status, Execution Trace
Keeps	Preparation and run history, kept apart

* The radar is an example profile, not a customer result. Each finding becomes a diagnosis card with what we saw, why it matters and the evidence.

Syntitan proposes. Your team confirms what matters.

Each diagnosis comes with a treatment you can apply, hold or exclude. People judge the business rules and key exceptions.

Diagnosis

The amount column mixes in symbols

What we saw

131 of 2,100 cells (6.24%) in total_spend carry a currency symbol or a comma.

Why it matters

Amounts cannot be added, compared or sorted as numbers.

Evidence

$1,669.91 $1,760.71 1,197

Treatment, recommended

Strip symbols and commas from amounts

What it keeps

Values that cannot be read as numbers stay as they are. No currency conversion.

Before and after

$1,669.91 → 1669.91

Your choice

Apply Hold Exclude

11 treatments, each with its own card

8 Recommended by default ● not reversible

- Remove duplicate rows ●
- Clean blanks written as text
- Normalize number formats
- Unify date formats
- Strip control characters and tags
- Cap outlier values
- Reduce skew with a log scale ●
- Fill empty cells, marking where they were

3 Needs your confirmation

- Unify case and spacing
- Group rare categories ●
- Drop columns with no information ●

Left out by meaning

ID, age and free-text columns skip treatments that would change what they mean.

Checked, no issue

All eleven checks are listed, including the ones that found nothing.

* The diagnosis and treatment cards show example data in the style of the product, not a customer result.

Simple guides the task. Expert opens the evidence.

Same Project and Version at two depths. You apply, hold or exclude each change, and the original stays untouched.

Simple View

Guides the task to completion

- Fill empty answers
- Match date formats
- Merge duplicate inquiries

Creating v2. The original stays as it is.

Expert View

Check and adjust conditions and evidence

Processing	Axis	Cells	Evidence
Unify amount units	Consistency	2	column amount, rule u-03
Fill missing account age	Integrity	1	column account_age_days
Merge refund and return	Consistency	held	Model does not use type

You decide Which changes to apply, and which Version to use.

01 **Ask**

02 **Diagnose**

03 **Apply**

69 → 84

Data readiness

04 **Compare**

71.08 → 78.41%

F1, same model and test set

05 **Recheck**

58 → 41

Missed fraud cases

* Figures are mockup examples, not customer results. Each belongs to the step above it.

The same business question, tested before and after the data changes.

Business question

What are net sales for the agreed reporting period?

Agreed first

Cancelled orders follow the agreed rule

Returns count in the agreed period

Currencies stay separate until a conversion policy is confirmed

Before preparation

Order	Status	Amount	Cur.
A-1042	Cancelled	1,200	USD
A-1043	Delivered	3,400	USD
A-1044	Returned	2,100	EUR

Net sales

$1.28M

Counts cancelled orders and adds currencies.

What Syntitan changed

Applied	Cancelled orders separated by the rule
Applied	Return dates matched to the agreed period
Preserved	Order and return links, original currency
Preserved	The original dataset, kept as v1
Left open	14 orders without a return date

After preparation

Net sales, same question

$1.12M

Same question, same evaluation conditions.

- Cancelled orders excluded
- Returns counted in the agreed period
- Currency conversion: your decision

The team can see whether the result follows the agreed business rules, and which exceptions still need review.

* Illustrative example with sample values, not a customer result. This case does not represent all of Syntitan. Ask us for more use cases by industry.

Same task, same conditions: four models, one goal.

Syntitan trains each task model on the original and on the prepared Dataset, then compares them.

We compared 4 models under the same conditions. 4 met the F1 75% goal.

Task model	F1 before → after	Change	Preprocessing applied	Verdict
XGBoost	80.00% → 87.50%	+7.50pp	1	Goal met
LightGBM	82.76% → 86.67%	+3.91pp	2	Goal met
Random Forest Selected	80.00% → 90.91%	+10.91pp	2	Goal met
Logistic Regression	11.11% → 83.33%	+72.22pp	5	Goal met

Interpretation **Random Forest already met the goal, and preprocessing raised it further.**

Missed targets

5 0 cases

Missed targets decreased by 5.

False alarms

0 3 cases

Non-targets flagged as targets rose by 3.

Correctly found

10 15 cases

Real targets found correctly rose by 5.

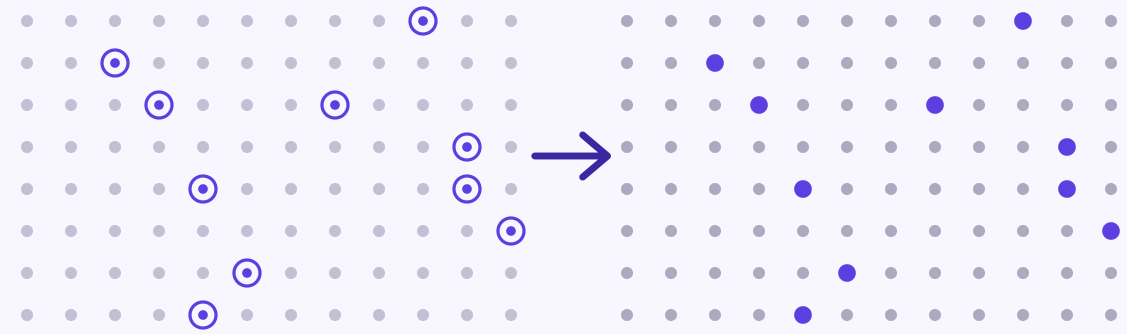
* Development screen with built-in example results, rebuilt for this page. Not a customer result.

Find the cases an agent misses, then decide what to fix.

Which task cases need attention, and is data part of the problem?

Run 1, gaps found

Same cases, after fixes



Before

An account takeover report arrives after the refund window. The agent treats it as a simple return.

After

The missing case is added and run again. It is flagged and handed to the right team.

Case passed · Gap found · Gap closed

01 **Build Tasks and Cases**

From the real work the agent must handle, including the unusual cases.

02 **Find the gaps** Planned

Read execution results for missing situations and the cause of each failure.

03 **Decide and fix**

Your team decides what to apply and fixes data, prompts, logic or tools in its own environment, then registers the change.

04 **Test again** Planned

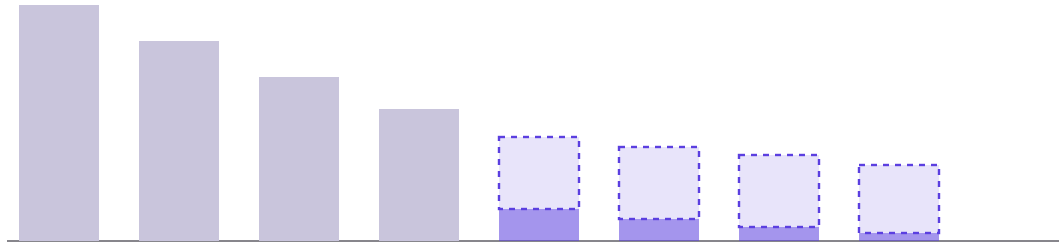
Run the same cases and confirm each gap is closed.

* Illustrative scenario. Each dot is one test case. Steps marked Planned are not generally available yet, and scope is confirmed per customer.

DTS fills the cases your data is missing.

For one defined AI task: rare cases and skewed groups are filled in, then checked on the same criteria.

Class distribution, illustrative



Where AI misses business exceptions

01 Missing rare events or failure cases
02 Skew across conditions and groups
03 Limits on source data use or sharing

Each output is judged on quality, protection and task usefulness, separately.

Insurance consultation classifier

88.55% weighted F1, up from 58.55%



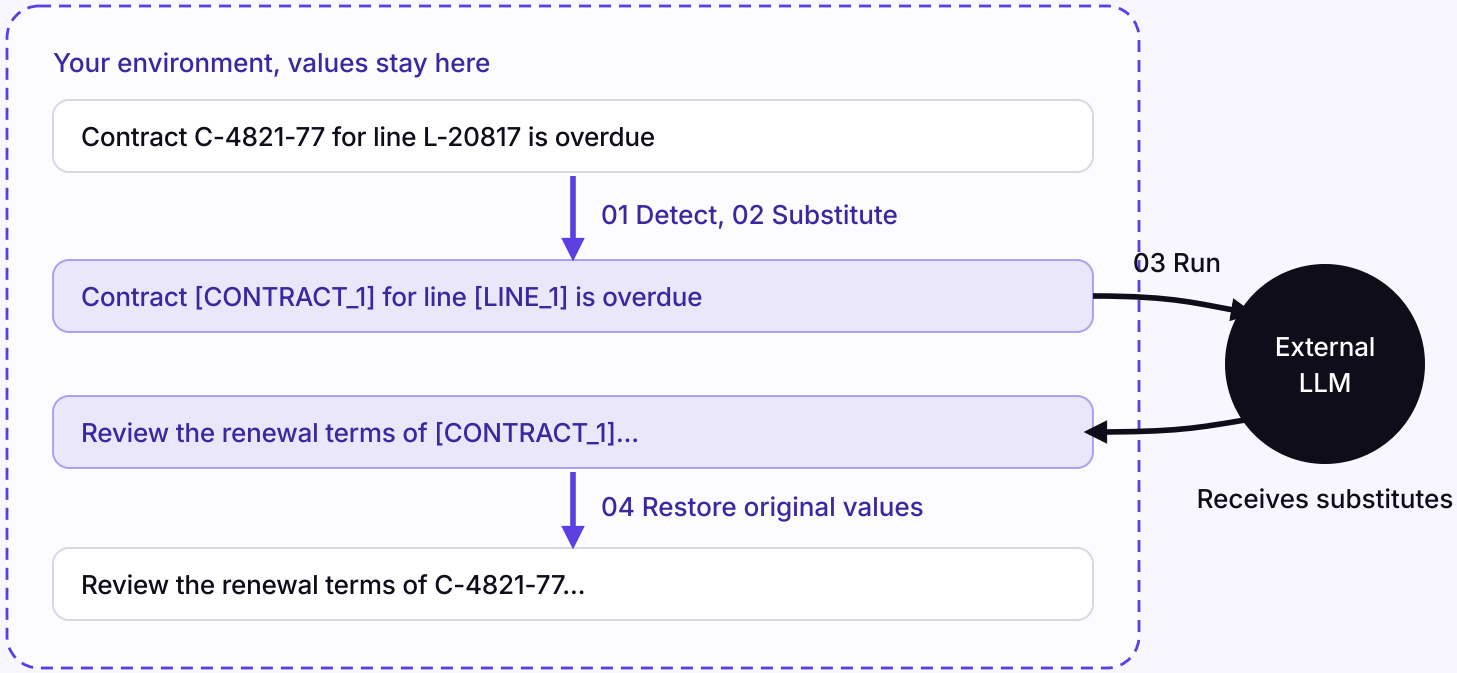
- 55 of 64 categories improved
- 6,874 cleaned records, 687 original records held out for testing
- Categories with under 50 records were supplemented
- Same classifier, same original test data

DTS provides data preparation and remediation inside Syntitan and can also be adopted independently.

* Measured by CUBIG on an insurance consultation PoC, 8 Jan 2025, on original test data. Macro F1 rose from 27.9% to 70.7%. Self-measured, not a guarantee.

The AI gets the context. Your team keeps the values.

Protection and restoration happen inside the system your team already uses.



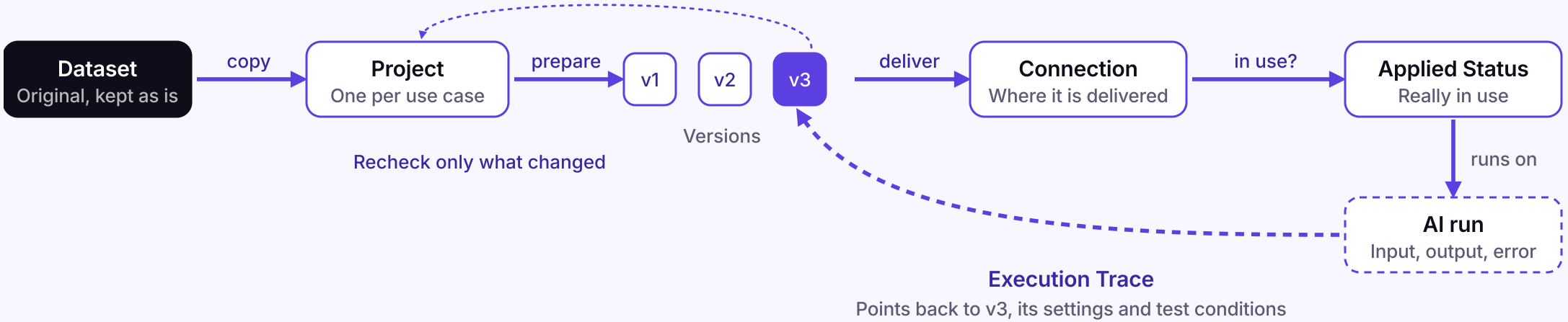
1. **Detect controlled values**
   Line IDs, contract references and other controlled values are found.
2. **Substitute them**
   Stand-ins keep the context and business relationships.
3. **Run the approved AI**
   The AI receives the substituted input.
4. **Restore original values**
   Your team reviews the result in its own system.

LLM Capsule LLM Capsule supports protected AI use within Syntitan and can also be adopted independently in an existing workflow.

* Example values are fictional. Where protection and restoration run depends on the deployment.

When a result changes, Syntitan shows what to recheck.

Delivery, application and execution are recorded as separate evidence.



Connection

The delivery route for a Version

Applied Status

Whether it was actually applied

Execution Trace

Inputs, outputs and errors of a run, tied to the Version and its conditions

Why it matters

Finding a problem is half the job.
Syntitan points you back to the data and conditions behind it, helps prepare it again and rechecks only what changed.

* Execution Trace connects a run to the data and conditions behind it. It is not just a log. Shown for supported connections.

The next task starts from criteria already reviewed.

Review existing rules and mappings for reuse. Recheck customer-specific values, units, permissions and evaluation conditions.

Criteria from the first task

- Standard data structure
- Column and code mapping approach
- Preparation and preservation rules
- Task evaluation criteria

Customer records and specific mapping values do not carry over.

Situation	Reuse	Recheck
01 **Same customer, new data**	Valid mappings and criteria	New fields, codes and affected results
02 **Another customer's data**	Structure and mapping template	Their values, units, permissions and evaluation conditions
03 **A different AI task**	Data and preparation rules	Task criteria, in a new Project

Why it matters The next task can start from reviewed criteria, so the team can focus on changes and unresolved decisions.

In Europe, AI teams are asked to explain their data.

The EU AI Act and GDPR shape what buyers ask about the data behind an AI system.
These are the questions we hear first, and the ones our products are built around.



	The questions	Where they come from
01	**Where did the data come from, and how was it prepared?**	EU AI Act, Art. 10: origin, labelling, cleaning, enrichment and known gaps
02	**Does the data reflect the setting where the AI is used?**	EU AI Act, Art. 10(3) and (4): relevance, representativeness and completeness
03	**Can you trace what ran, on which data and conditions?**	EU AI Act, Art. 12 and Annex IV: automatic logs and dated test records
04	**Do personal values have to leave your environment?**	GDPR, Art. 5, Art. 25 and Chapter V: minimisation, pseudonymisation and transfers outside the EEA

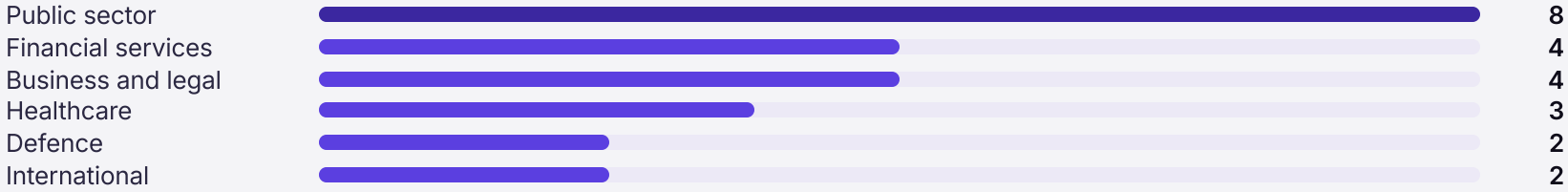
ISO/IEC 27001 and 42001 certified. CUBIG LTD, our UK entity, since 2025.

* A summary for context, not legal advice. Fit with each rule is assessed per customer and deployment. This page makes no compliance claim. Photo: Guillaume Périgois, Unsplash.

23 enterprises and institutions across six sectors.

Contracts, projects, PoCs, PoVs and programmes: each is a different stage of engagement.

References
23
Enterprise and institutional references



Insurance company, Korea

Life insurance

Customer service classifier

58.55 → 88.55%

Weighted F1, CUBIG-measured, Jan 2025

Samsung Securities

Securities, Samsung Group

Churn prediction

Excellence award, 2025

Syntitan

CJ 프레시웨이

Food service, CJ Group

Quote drafting

Product-level quotes from one dataset

Syntitan

DB

Non-life insurance

External LLM use

External LLM in use, values restored inside

LLM Capsule

Telecom · Insurance · Banking · Securities · Food service · Payments

SK telecom · KYOBO 교보생명 · DB · Industrial Bank of Korea · WOORI BANK · NH농협은행 · Samsung Securities · CJ 프레시웨이 · Danal

* One case does not represent all of CUBIG. Ask us for more use cases by industry. Customer names are shown with approval.



Ho Bae
Co-founder and CEO, CUBIG
Professor, Ewha Womans University

Why enterprise AI stalls after the demo

“CUBIG exists so AI work can be checked, reused, and redone only where something changed.”

From security research to AI-ready data

2007 to 2009
UCL
MSc Information Security

2010 to 2012
Samsung
Network security research

2012 to 2016
SAP Labs Korea
In-memory database

2016 to 2021
Seoul National University
PhD

2021 to now
Ewha Womans University
Professor, CUBIG founded

20+
Peer-reviewed papers, incl. NeurIPS, ICLR and AAAI

2018
First-authored survey, Security and Privacy Issues in Deep Learning

Since 2007
Across security, data systems and AI

Selected papers

- NeurIPS 2023 — PUCA: Patch-Unshuffle and Channel Attention for Enhanced Self-Supervised Image Denoising
- ICLR 2023 — New Insights for the Stability-Plasticity Dilemma in Online Continual Learning
- ICLR 2024 — DAFA: Distance-Aware Fair Adversarial Training
- PSB 2020 — AnomiGAN: Generative Adversarial Networks for Anonymizing Private Medical Data

CUBIG

See what changes in your AI task when the data changes.

Start with one task and an agreed comparison. Review the prepared data, the result and the issues that remain.

Request a Syntitan demo ↗

