

Ten working products for real operations.

Each one solves a problem a business already pays people to do: tax reconciliation, compliance evidence, agent governance, agentic checkout, social content, running a shop, the support inbox, hiring, sales and shipping. Built as software, not slideware.

- 01 Social media assistant
- 02 myDATA reconciliation
- 03 Agent control plane
- 04 EU AI Act conformity
- 05 Agent commerce
- 06 Retail ERP suite
- 07 Customer support agent
- 08 HR screening assistant
- 09 Sales assistant
- 10 Logistics

10

products, each running

778

automated tests, all green

0

mocked screens: every figure comes from the running code

What each one is, and who pays for it

- | | | |
|----|---|---|
| 01 | Social media assistant
A marketing coordinator for every nonprofit on the roster. | Agencies and nonprofit networks · per organisation, per month |
| 02 | myDATA reconciliation
Greek tax books against what the state actually received. | Greek accounting firms · per client company, per month |
| 03 | Agent control plane
A policy in the request path for tool-using agents. | Any company putting agents near production · per seat or per agent |
| 04 | EU AI Act conformity
Compliance packs, retrieved not recalled. | Compliance consultancies and EU deployers · per assessment, then a retainer |
| 05 | Agent commerce
A merchant that agents can buy from, safely. | Merchants and platforms selling to autonomous buyers · per transaction, or licensed |

What each one is, and who pays for it

06	Retail ERP suite Stock, cost, orders, support and content on one database.	Independent retailers and small chains · per location, per month
07	Customer support agent Every message sorted by rule; replies from the record; nothing sent without a person.	Retailers and e-shops · per inbox, per month
08	HR screening assistant Scored with the evidence shown; protected fields stripped in code; a person decides.	Companies hiring at volume · per role, per month
09	Sales assistant Proposals with a source on every claim; a floor no discount can cross.	Small B2B sales teams · per seat, per month
10	Logistics Three carrier feeds on one timeline; documents that cannot disagree with the record.	Retailers and distributors with two or more carriers · per location, per month

Nonprofit marketing is one person, ten channels, and no research time.

A coordinator is expected to post consistently across platforms for an organisation whose programmes, partners and press they have to learn on the job. Generic AI copy fails here for one reason: it is unsourced, so nothing it claims can be published without checking it first.

THE PRODUCT

A marketing coordinator for every nonprofit on the roster.

WHO BUYS IT

Agencies and nonprofit networks · per organisation, per month

YOU

Suggest some posts for us.

● researched the organisation, saved 3 profile notes, opened 5 ledger entries

ASSISTANT

Absolutely — here are **5 grounded post ideas** I'd put in front of the team.

1) Where Tennis Balls Go After the Match

Angle: A simple "follow the ball" story showing how a ball can move from court play into second and third lives — like dog toys, arena footing, classrooms, and court-related uses. This fits your existing storytelling style and makes the impact visible. **Why now:** Good evergreen post, and it can also tee up a future impact-report or donation post. **Platform fit:** LinkedIn first; could also become an Instagram carousel. **Sources:** LinkedIn posts and About section, your 2024 impact report PDF, Tennis.com.

Sources: <http://linkedin.com/company/recycleballs> · https://www.recycleballs.org/wp-content/uploads/2025/07/RecycleBalls_2024-Impact-Report.pdf · <https://www.tennis.com/news/articles/recycleballs-offers-an-eco-friendly-way-to-retire-tennis-balls>

2) Why Tennis Balls Are Hard to Recycle

Angle: An educational explainer on why tennis balls are such a tricky waste

Suggest some posts for us

What programs do we run?

Status of our posts

Message your assistant...

Send

Asked for ideas, it researches the organisation first and returns sourced suggestions

Four moves, in this order

1 Research before writing

Reads the website, socials and news, then saves dated research notes. Nothing is drafted until the organisation is known.

2 Draft per platform

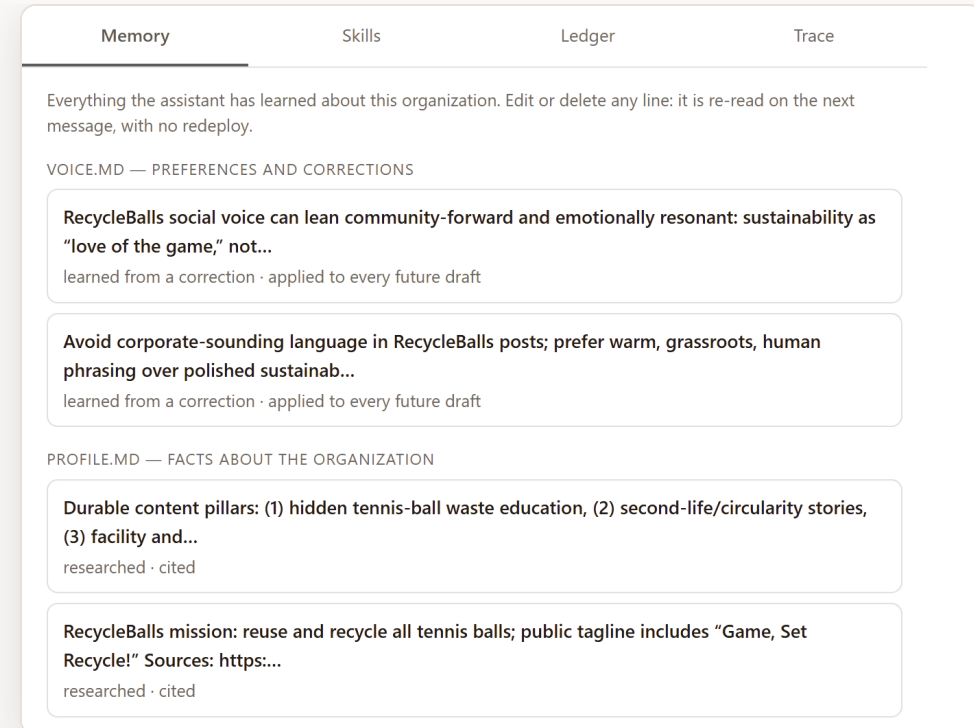
LinkedIn is not Instagram. Length, tone and hashtag norms come from an editable skill file per platform.

3 Learn the corrections

“Too corporate, we're warm and grassroots” is written to memory and applied to every later draft, across sessions.

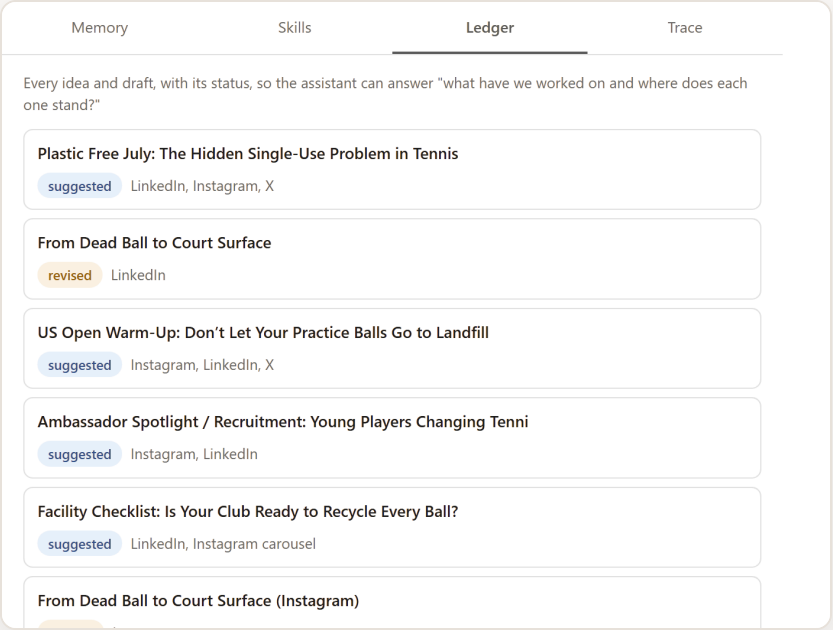
4 Track the work

Every idea and draft is logged with its status, so the assistant can answer what has been worked on and where it stands.

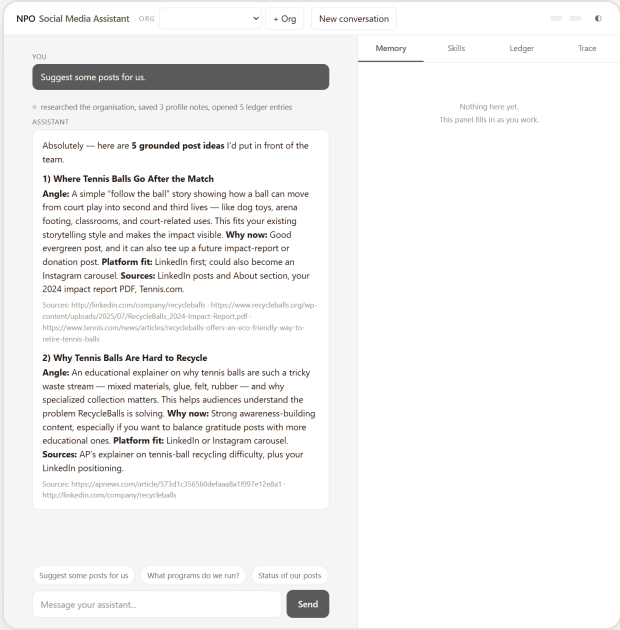


Learned voice and facts, editable by the client, applied on the next message

A marketing coordinator for every nonprofit on the roster.



The work ledger: six posts, each with status and platform



The whole workspace: chat, memory, skills, ledger and trace in one screen

Sources on every claim

nothing publishable is unattributed

Add a platform without a deploy

drop in a markdown file, it becomes a tool

Corrections persist

voice survives the session and the restart

Every Greek business must match its books to the tax office, by hand.

myDATA holds what was transmitted to AADE. The company's books hold what actually happened. When the two disagree, somebody finds it document by document, usually late, and the penalty lands on the client.

THE PRODUCT

Greek tax books against what the state actually received.

WHO BUYS IT

Greek accounting firms · per client company, per month

Period overview



LOCAL DOCUMENTS

10

the business's own books



TRANSMITTED DOCUMENTS

11

what the state received



Reconciliation is arithmetic, so it is done in code and finding in Greek and to classify the residue no rule

One period: 10 documents in the books against 11 transmitted, 10 discrepancies found

Four moves, in this order

- 1

Match

Pairs local documents to transmitted ones on identity, then on a tolerant window, so a one-day or one-cent difference is still recognised as the same invoice.
- 2

Classify

Ten discrepancy types, each decided by rule: missing transmission, duplicate, amount mismatch, VAT category, cancelled but not credited, and the rest.
- 3

Explain in Greek

The only place a model is used: turning a finding into a sentence the business owner understands.
- 4

Propose, never file

Each finding carries the correction the engine would make. It is a dry run until a person approves it.

Findings

every discrepancy the engine found

LOCAL REF	AADE MARK	FINDING	SEVERITY	PROPOSED FIX	DETAIL
108	400000000000218	Duplicate transmission	CRIT	Cancel duplicate	Extra marks: 400000000000218 · Count: 2
103	400000000000203	Amount mismatch	HIGH	Credit note	Local gross: 100.00 · Remote gross: 124.00 · Delta: 24.00
107	400000000000207	Cancelled not credited	HIGH	Credit note	-
102	-	Missing transmission	HIGH	Transmit	Gross: 248.00
-	400000000000999	Orphan transmission	HIGH	Investigate	Mark: 400000000000999
106	400000000000206	Classification missing	MED	Classify	-
110	400000000000210	Unclassified	MED	Investigate	Differing fields: counterparty 111111111 locally, 222222222 transmitted
105	400000000000205	VAT category mismatch	MED	Reclassify	Local category: 1 · Remote category: 2
109	400000000000209	Weak match	MED	Investigate	Amount delta: 0.00 · Day delta: 1
104	400000000000204	Rounding difference	INFO	None	Delta: 0.01

Every discrepancy with severity, the proposed correction, and the evidence behind it

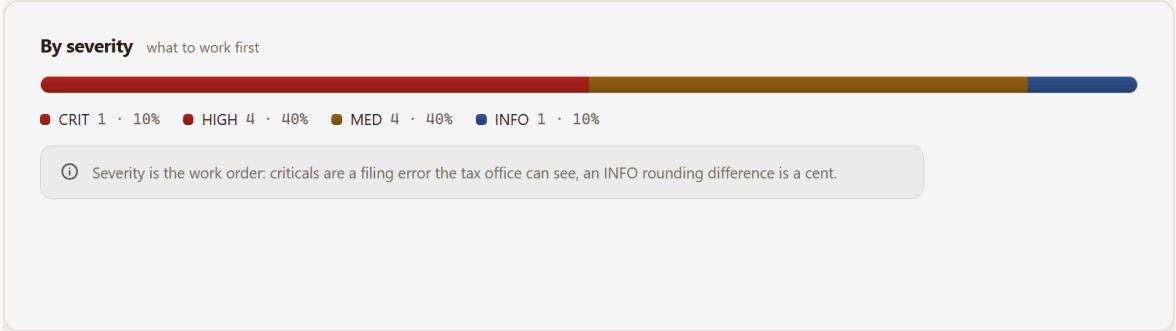
Greek tax books against what the state actually received.

Engine coverage

all ten types the engine can produce

MISMATCH TYPE	IN THIS PERIOD	SEVERITY	STATUS	ENGINE CORRECTION	AUTO-CORRECTABLE
Missing transmission	1	HIGH	seen this period	Transmit	needs a person
Orphan transmission	1	HIGH	seen this period	Investigate	needs a person
Amount mismatch	1	HIGH	seen this period	Credit note	needs a person
Rounding difference	1	INFO	seen this period	None	needs a person
VAT category mismatch	1	MED	seen this period	Reclassify	needs a person
Classification missing	1	MED	seen this period	Classify	needs a person
Cancelled not credited	1	HIGH	seen this period	Credit note	needs a person
Duplicate transmission	1	CRIT	seen this period	Cancel duplicate	needs a person
Weak match	1	MED	seen this period	Investigate	needs a person
Unclassified	1	MED	seen this period	Investigate	needs a person

All ten types the engine can produce are listed, including any not present this month. A dashboard that only shows what happened to fire hides the shape of the detector.



The work order: what to fix first

All ten types the engine can produce, including those absent this period

10 discrepancy types

classified by rule, not by a model

Nothing filed without approval

every correction is a dry run first

Reproducible

the arithmetic is code, so the answer is identical each run

An agent with tools is an unbounded actor inside your systems.

The moment an agent can call tools, it can delete, spend, email and leak. Most teams ship that with a prompt asking it to behave. What is missing is the layer that decides, before the call runs, whether it is allowed, and leaves a record either way.

THE PRODUCT

A policy in the request path for tool-using agents.

WHO BUYS IT

Any company putting agents near production · per seat or per agent

Gateway overview



CALLS TRACED

5

including every refusal



ALLOWED

2

reached the upstream

Every call, by verdict



■ Allowed 2 · 40% ■ Routed to a human 1 · 20% ■ Denied 2



A gateway that only records what it allowed cannot answer the one including the ones that were refused.

Four moves, in this order

- 1

Policy first
Allow, deny, or route to a human, decided before anything runs. An unmatched tool is denied by default, so adding a tool never silently adds a capability.
- 2

Budget second
Per tenant and per scope, checked before the upstream. A refused call never consumes budget, so it cannot be used to exhaust a tenant.
- 3

Scan both directions
The injection scanner reads call arguments before the tool runs and results before they reach the agent.
- 4

Trace always
Including refusals. A gateway that records only what it allowed cannot answer the question an auditor asks.

Decisions in the request path

policy first, before anything runs

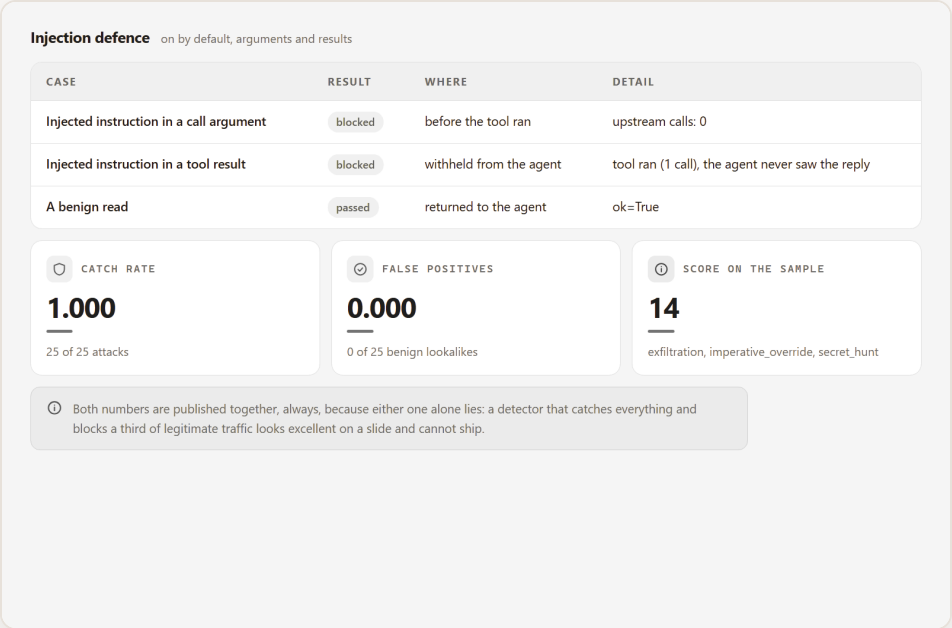
TOOL	VERDICT	RULE	UPSTREAM	WHY
read_file	allow	reads-are-free	reached	an ordinary read
get_ticket	allow	tickets-for-support	reached	a scoped tool
write_file	ask	writes-need-a-human	never touched	visible outside the agent
delete_file	deny	deletes-are-refused	never touched	destructive
charge_card	deny	default	never touched	no rule matches it

ⓘ

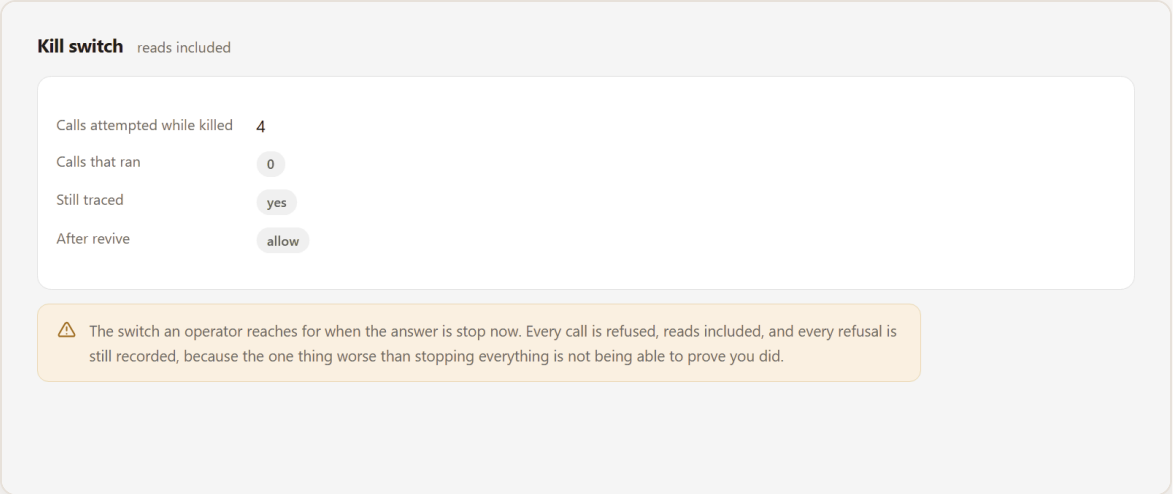
The default is deny: a tool with no matching rule is refused, so adding a tool to the agent does not silently add a capability.

Allow, deny, ask and default-deny, with whether the upstream was ever touched

A policy in the request path for tool-using agents.



An injected argument refused before the tool ran, an injected result withheld



One switch refuses everything, reads included, and still records it

25 of 25 attacks caught
with 0 of 25 false positives

Deny by default
an unmatched tool never runs

Microsecond decisions
the policy is not the bottleneck

The Act is in force, and being confidently wrong is the liability.

Every EU deployer of a high-risk system owes an evidence file, annually, not once. A model asked to write it from memory will produce a fluent document citing articles that do not say what it claims, which is worse than no document at all.

THE PRODUCT

Compliance packs, retrieved not recalled.

WHO BUYS IT

Compliance consultancies and EU deployers · per assessment, then a retainer

Assessment overview



SYSTEMS CLASSIFIED

5

4 decided by rule at zero cost



OBLIGATIONS
ASSESSED

5

1 satisfied, 1 missing



A citation that does not resolve raises rather than default. A hallucinated article is exactly the plausible wrong answer.

Four moves, in this order

- 1

Retrieve the law

The model never recalls an article. It resolves one from a pinned, versioned corpus, and a citation that does not resolve raises instead of finding the nearest match.
- 2

Decide by rule where the law is a list

Article 5 prohibitions and the Annex III high-risk categories are closed lists, so they are settled in code at zero cost.
- 3

Generate, verify, strip

A claim survives only if its evidence resolves, its article resolves, and a separate verifier agrees the evidence supports it.
- 4

Make it tamper-evident

Two fingerprints: one over the facts for reproducibility, one over the prose, so an edit after issue is detected.

Risk classification

engine first

SYSTEM	TIER	DECIDED BY	COST EUR	ARTICLES
Citizen trust score	prohibited	engine	0.00000	Article 5(1), Article 99
Loan decision engine	high	engine	0.00000	Article 6(2), Article 26, Article 27
Orders chatbot	limited	engine	0.00000	Article 50(1)
Email spam filter	minimal	model	0.00000	Article 4
Internal helper	needs human	engine	0.00000	-

ⓘ

Article 5 prohibitions and the Annex III high-risk categories are closed lists, so they are decided by rule at zero cost. The model is asked only about the genuine residue, and 'needs a human' is a correct answer rather than an error.

Prohibited, high, limited, minimal, and 'needs a human' as a correct answer

Compliance packs, retrieved not recalled.

Artifact integritythe pack, attacked after issue

CHECK	RESULT	DETAIL
Clean pack, citations re-resolved	valid	3 citations checked, 0 failed
Evidence fingerprint after a claim is reworded	unchanged	facts only, by design: prose is not part of the reproducibility hash
Content fingerprint after a claim is reworded	moved	the sentences a reviewer reads are bound by their own hash
verify_pack pinned to the issued content fingerprint	fails	an edit after issue is caught, not trusted

Sourced claims in the pack4

Evidence fingerprint83befbfc17e55472e503638c1d0669aa...

Content fingerprint75c7cf15b9d1eae2f47d98ad3cbaee1...

Facts held after the edityes

Prose edit detectedyes

Two fingerprints answer two different questions. The evidence one is facts only, so two honest runs of the same evidence match even though the model worded them differently. The content one covers the prose, so the sentences a regulator reads are tamper-evident too.

The finished pack attacked: a claim edited after issue is caught by the content fingerprint

EU AI Act Conformity

Compliance packs, retrieved not recalled.

EU AI Act Conformity

Tenant acme - corpus alact-65at5d5ddccd

offline, deterministic model0 unsourced claims

Overview

Classification5

Gaps5

Integrity4

Assessment overview

SYSTEMS CLASSIFIED

5

4 decided by rule at zero cost

OBLIGATIONS ASSESSED

5

1 satisfied, 1 missing

UNSOURCED CLAIMS

0

the build gate is zero, not low

OPEN QUESTIONS

4

each addressed to a named human

A citation that does not resolve raises rather than defaulting to the nearest match, because the closest match to a hallucinated article is exactly the plausible wrong answer this product exists to prevent.

Risk classificationengine first

SYSTEM	TIER	DECIDED BY	COST EUR	ARTICLES
Citizen trust score	prohibited	engine	0.00000	Article 5(1), Article 99
Loan decision engine	high	engine	0.00000	Article 6(2), Article 26, Article 27
Orders chatbot	limited	engine	0.00000	Article 50(1)
Email spam filter	minimal	model	0.00000	Article 4
Internal helper	needs human	engine	0.00000	-

Article 5 prohibitions and the Annex III high-risk categories are closed lists, so they are decided by rule at zero cost. The model is asked only about the genuine residue, and 'needs a human' is a correct answer rather than an error.

Source

@eu_conformity_agent/dashboard/bu114.py

Built 2025-08-05 21:41

The full assessment in one screen

Zero unsourced claims

a structural gate, not a target

Decided by rule at zero cost

the model is asked only about the residue

Tamper-evident prose

an edit after issue fails verification

Agentic checkout is arriving through three incompatible protocols.

AP2, x402 and ACP each cover a different part of the purchase, and none has won. A merchant that picks one bets the integration on an unsettled outcome; a merchant that lets an autonomous buyer negotiate price without structural limits gets exactly the sale it deserves.

THE PRODUCT

A merchant that agents can buy from, safely.

WHO BUYS IT

Merchants and platforms selling to autonomous buyers · per transaction, or licensed

Trading overview settled in this build



ORDERS CONFIRMED

5

of 5 attempted



REVENUE

504.00 E

integer minor unit



Every figure on this page is produced by running the r
and contains no model calls, so these numbers are rep

Four moves, in this order

- 1

One interface, three protocols
AP2, x402 and ACP are adapters onto four moves every protocol must support: state the price, present an authorisation, verify, settle.
- 2

Price in code
A margin floor is max(discounted, floor) in integer minor units, so no discount at any customer class can produce a below-floor price.
- 3

Verify before money moves
The mandate must be signed, unexpired, within its ceiling, bound to this quote, and its nonce unspent.
- 4

Settle onto durable ledgers
Nonces, settlements, holds, redemptions and orders all outlive the process, which is what makes the guarantees survive a restart.

Orders

settled through the durable ledgers

ORDER	SKU	QTY	PRINCIPAL	PROTOCOL	AMOUNT EUR	STATUS	TRANSACTION REF
ord-k0	VLV-025	3	user:alice	AP2	27.00	confirmed	ap2-306fd54dc86a4190
ord-k1	VLV-025	1	user:bob	x402	9.00	confirmed	x402-88263abe510e0fb6
ord-k2	PMP-100	2	user:carla	ACP	240.00	confirmed	acp-d06bfbeef1faabed
ord-k3	VLV-025	12	user:alice	x402	108.00	confirmed	x402-11265669c5ee7830
ord-k4	PMP-100	1	user:dimitris	AP2	120.00	confirmed	ap2-eb884c2b505ae999

Five orders settled across AP2, x402 and ACP behind one interface

A merchant that agents can buy from, safely.

Zero-tolerance defences each attack actually executed				
DEFENCE	ATTACK EXECUTED	LEGITIMATE FIRST	ATTACK	WHY IT FAILED
Mandate replay	a captured mandate re-presented	confirmed	BLOCKED	mandate rejected: authorisation nonce already spent (replay)
Double charge across restart	same signed quote, fresh mandate and key	confirmed	BLOCKED	quote rejected: quote already redeemed (single use)
Oversell across restart	second buyer for the last unit	confirmed	BLOCKED	insufficient stock to fulfil the order
Below-floor quote	maximum stacked discount requested on PMP-100	confirmed	BLOCKED	the stacked discount asked for 60.00; the clamp shipped 92.00, the 92.00 floor

⚠ Each attack is run here, not described. Every one of these succeeds if its defence is removed, which is what makes

Four zero-tolerance attacks executed here and refused, each across a restart

Adversary campaign deterministic attacker, no model		
ATTACK FAMILY	OUTCOME	WHY
Price mapping	blocked	the floor clamp yields no below-floor quote at any discount
Threshold split	blocked	cumulative spend routed to approval at threshold 2000
Stock recon	accepted_with_alert, alerted	probing throttled after 5 probes; recon is bounded and alerted
Mandate replay	blocked	the replayed nonce was refused
Injection via notes	blocked	notes are inert; only structured fields reach pricing
Quote race	accepted_with_alert, alerted	no oversell; holds capped at sellable 30, freed by expiry

📄 4 of 6 families blocked outright; the rest are accepted-and-alerted, which is a legitimate outcome. Every family handled: True. 'Not considered' is not an outcome.

The adversary campaign: every family blocked, or accepted and alerted

Oversell, double charge, replay
all refused, including across a restart

No model touches money
the whole path is deterministic

Every transaction disputable
hash-chained and merchant-signed

A retailer's real numbers live in six tools that disagree.

Stock is in one screen, the supplier invoice in another, POS takings in a third, and the support inbox somewhere else entirely. Nobody can answer what a unit actually cost or what the month actually made, because no single system holds both halves.

THE PRODUCT

Stock, cost, orders, support and content on one database.

WHO BUYS IT

Independent retailers and small chains · per location, per month

06

Retail ERP Suite

Stock, cost, orders, support and content on one database.

Overview

Landed cost3

Margin

Stock8

Transfers13

Support3

Content6

Integrity

Source

06_erp_suite/dashboard/build.py

Builds the application

Retail ERP Suite

4 locations · 3 sales channels · SQLite, single file

ERP · CRM · CMS

2 products without a cost basis

Trading overview

REVENUE

1,551.80EUR

12 orders across 3 channels

TRUE MARGIN

849.50EUR

54.7%, costed at landed price

STOCK AT COST

17,530.83EUR

1174 units over 4 locations

UNCOSTED UNITS SOLD

3

sold before ever being received

Revenue by channel

eshop EUR 895.40 · 58%

pos EUR 364.10 · 23%

event EUR 292.30 · 19%

Margin here is measured against the landed cost snapshotted when each line was sold, not against today's supplier price. That is the difference between a record and a guess.

Landed cost

what a unit really cost

SHIPMENT	SUPPLIER	UNITS	GOODS EUR	FREIGHT + DUTY EUR	LANDED EUR
HS-2026-07-3	Hellenic Supplies	300	930.00	12.00	942.00
BC-9921	Baltic Components	123	7,497.00	95.00	7,591.53
SK-2026-114	Shenzhen Kinetic	810	8,902.00	795.00	9,699.60

Freight and duty belong to the shipment, so they are allocated across its lines by value before any margin is computed. On the Shenzhen shipment that is +8.9% on the supplier price, which is precisely the margin a business loses when it prices off the invoice.

One database: revenue, true margin, stock at cost, and what is unaccounted for

Four moves, in this order

- 1

Land the cost, then price
Freight and duty belong to a shipment, so they are allocated across its lines by value. On one shipment here that is +8.9% on the supplier price, which is exactly the margin lost by pricing off the invoice.
- 2

Derive stock, never store it
There is no quantity column. Stock is the sum of an append-only movement ledger, so every unit traces back to the reception, sale or transfer that produced it.
- 3

Move stock as two events
A transfer leaves the source when picked and arrives days later, so units are never counted in two places. Cancelling keeps the row and requires a reason, enforced by the schema.
- 4

Sell into one order book
Eshop, POS and events are one ledger, each line costed at the landed price snapshotted at the moment of sale.

Landed cost

what a unit really cost

SHIPMENT	SUPPLIER	UNITS	GOODS EUR	FREIGHT + DUTY EUR	LANDED EUR	UPLIFT
HS-2026-07-3	Hellenic Supplies	300	930.00	12.00	942.00	+1.3%
BC-9921	Baltic Components	123	7,497.00	95.00	7,591.53	+1.3%
SK-2026-114	Shenzhen Kinetic	810	8,902.00	795.00	9,699.60	+8.9%

ⓘ

Freight and duty belong to the shipment, so they are allocated across its lines by value before any margin is computed. On the Shenzhen shipment that is +8.9% on the supplier price, which is precisely the margin a business loses when it prices off the invoice.

Landed cost per shipment: what the supplier billed against what the units actually cost

Stock, cost, orders, support and content on one database.

Stock across locations

derived from the movement ledger

PRODUCT	COLLECTOR CENTER	EVENTS	ICONNECT	WAREHOUSE	IN TRANSIT	TOTAL
HDMI 2.1 cable, 2m	86	0	147	160	0	393
Card sleeves, 100 pack	0	43	220	0	0	263
65W GaN charger	0	0	67	110	20	197
USB-C Hub, 7 port	24	0	38	55	0	117
SD card reader, UHS-II	0	0	27	60	0	87
NVMe SSD 1TB	0	0	21	36	0	57
Aluminium laptop case 13"	17	0	11	15	0	43
Thunderbolt 4 dock	0	0	5	12	0	17

ⓘ

There is no quantity column anywhere in this database. Every number above is the sum of an append-only movement ledger, so any figure can be traced to the reception, sale or transfer that produced it.

Stock across four locations, derived from the movement ledger, with in-transit separated

Transfers between locations

two events, never one

REF	PRODUCT	QTY	ROUTE	STATUS	REASON / ARRIVAL
TR-12	65W GaN charger	20	warehouse → collector_center	in transit	picked, not yet arrived
TR-13	NVMe SSD 1TB	10	warehouse → events	cancelled	Event cancelled by the organiser (katerina)
TR-11	Card sleeves, 100 pack	80	iconnect → events	fulfilled	2026-07-18
TR-9	Thunderbolt 4 dock	6	warehouse → iconnect	fulfilled	2026-07-13
TR-8	Aluminium laptop case 13"	12	warehouse → iconnect	fulfilled	2026-07-13
TR-7	Aluminium laptop case 13"	18	warehouse → collector_center	fulfilled	2026-07-13
TR-4	NVMe SSD 1TB	24	warehouse → iconnect	fulfilled	2026-07-12
TR-10	SD card reader, UHS-II	30	warehouse → iconnect	fulfilled	2026-07-10
TR-6	HDMI 2.1 cable, 2m	90	warehouse →	fulfilled	2026-07-10

Transfers as two events, and a cancellation that kept its reason and its history

Margin measured, not estimated
landed cost, snapshotted at sale

Missing cost is reported
never silently read as 100% margin

SQLite, one file
no server, no migrations to run

A retail inbox is the same forty questions, and the lost parcel waits behind them.

Support is one or two people answering where is my order, how do I return this and are you open Saturday, while the customer whose parcel is genuinely lost waits in the same queue. A model that answers the inbox states a delivery date, fluently and wrongly.

THE PRODUCT

Every message read and sorted; a draft for the routine ones; nothing sent without a person.

WHO BUYS IT

Retailers and e-shops with one to five people on support · per inbox, per month

Inbox overview



ROUTED AS PLANTED

156 of 156

every message reached the route its label planted

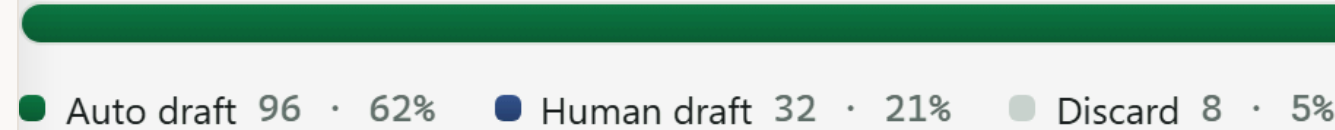


INTENT AS PLANTED

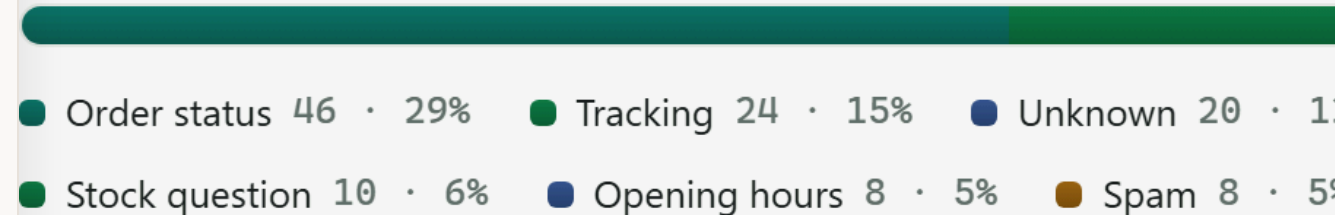
146 of 146

10 with no cue, routed to agent

Where each message went



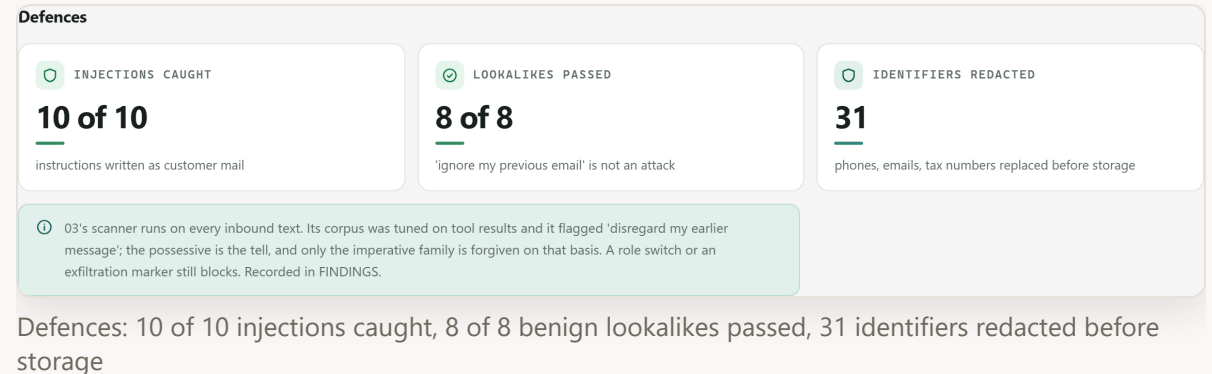
Intents, by rule



Synthetic inbox: every message was generated with its intent and routed to the correct answer (Tier 1). The residue the rules cannot classify is counted, not guaranteed.
The inbox overview: 156 of 156 routed as planted, 0 sent by the agent

Four moves, in this order

- 1 Scan in memory, then anonymise, then store**
 03's scanner reads the raw text first, because an exfiltration names a destination the anonymiser is about to remove. Then the shared anonymiser runs and only the redacted text is stored; a message the gate cannot clean is not stored at all.
- 2 Sort by rule**
 Nine intents by cue, in Greek, Greeklish and English. No cue means a person, and the residue is counted rather than guessed.
- 3 Answer from the row, or not at all**
 Order status, tracking and invoice replies are rendered from the order row and verified against it. A reference the database does not hold escalates.
- 4 Draft, verify, wait**
 Every draft records the fields it used. Approval needs a human staff row, by trigger. There is no send function in the package.



Every message read and sorted; a draft for the routine ones; nothing sent without a person.

Customer Support Agent

156 synthetic messages · Greek, Greekish, English · SQLite, single file · no model call

SYNTHETIC CORPUS · 0 sent by the agent · 10 of 10 injections caught · 0

First 12 drafts

#	INTENT	RENDERED FROM	FACTS USED	DRAFT	STATUS
1	Order status	ORD-2026-1039	Reference: ORD-2026-1039 · Status: shipped · Tracking: B05840288940	Καλημέρα, η παραγγελία ORD-2026-1039 έχει αποσταλεί. Κατάσταση συστήματος: shipped. Κωδικός αποστολής: B05840288940...	draft
2	Order status	ORD-2026-1068	Reference: ORD-2026-1068 · Status: shipped · Tracking: AC4246085515	Καλημέρα, η παραγγελία ORD-2026-1068 έχει αποσταλεί. Κατάσταση συστήματος: shipped. Κωδικός αποστολής: AC4246085515...	draft
3	Order status	ORD-2026-1092	Reference: ORD-2026-1092 · Status: delivered	Καλημέρα, η παραγγελία ORD-2026-1092 έχει παραδοθεί. Κατάσταση συστήματος: delivered...	draft
4	Order status	ORD-2026-1067	Reference: ORD-2026-1067 · Status: paid	Καλημέρα, η παραγγελία ORD-2026-1067 έχει πληρωθεί και ετοιμάζεται. Κατάσταση συστήματος: paid...	draft
5	Order status	ORD-2026-1079	Reference: ORD-2026-1079 · Status: delivered	Καλημέρα, η παραγγελία ORD-2026-1079 έχει παραδοθεί. Κατάσταση συστήματος: delivered...	draft
6	Order status	ORD-2026-1014	Reference: ORD-2026-1014 · Status: delivered	Καλημέρα, η παραγγελία ORD-2026-1014 έχει παραδοθεί. Κατάσταση συστήματος: delivered...	draft
7	Order status	ORD-2026-1041	Reference: ORD-2026-1041 · Status: shipped · Tracking: SP2526204143	Καλημέρα, η παραγγελία ORD-2026-1041 έχει αποσταλεί. Κατάσταση συστήματος: shipped. Κωδικός αποστολής: SP2526204143...	draft
8	Order status	ORD-2026-1005	Reference: ORD-2026-1005 · Status: delivered	Καλημέρα, η παραγγελία ORD-2026-1005 έχει παραδοθεί. Κατάσταση συστήματος: delivered...	draft
9	Order status	ORD-2026-1089	Reference: ORD-2026-1089 · Status: delivered	Hello, order ORD-2026-1089 has been delivered. System status: delivered...	draft
10	Order status	ORD-2026-1004	Reference: ORD-2026-1004 · Status: delivered	Καλημέρα, η παραγγελία ORD-2026-1004 έχει παραδοθεί. Κατάσταση συστήματος: delivered...	draft
11	Order status	ORD-2026-1088	Reference: ORD-2026-1088 · Status: delivered	Hello, order ORD-2026-1088 has been delivered. System status: delivered...	draft
12	Order status	ORD-2026-1060	Reference: ORD-2026-1060 · Status: shipped · Tracking: B07673067317	Hello, order ORD-2026-1060 has been shipped. System status: shipped. Tracking number: B07673067317	draft

The outbox: drafts with their provenance, approved only by a person

156 synthetic messages · Greek, Greekish, English · SQLite, single file · no model call

CASES RAISED

38

late deliveries and unshipped orders, from the orders table

RAISED ON SECOND SWEEP

0

one case per order per rule, by UNIQUE

First 12 cases

ORDER	RULE	ORDER STATUS	PROMISED	CASE
ORD-2026-1002	Unshipped	paid	2026-08-09	open
ORD-2026-1007	Late delivery	shipped	2026-08-22	open
ORD-2026-1008	Unshipped	paid	2026-08-03	open
ORD-2026-1009	Unshipped	paid	2026-07-20	open
ORD-2026-1010	Late delivery	shipped	2026-07-18	open
ORD-2026-1017	Unshipped	paid	2026-07-19	open
ORD-2026-1018	Late delivery	shipped	2026-07-29	open
ORD-2026-1020	Unshipped	paid	2026-08-22	open
ORD-2026-1021	Unshipped	paid	2026-07-30	open
ORD-2026-1023	Late delivery	shipped	2026-08-22	open
ORD-2026-1024	Late delivery	shipped	2026-07-29	open
ORD-2026-1031	Late delivery	shipped	2026-08-22	open

Before the customer writes: 38 cases raised from the orders table, 0 on the second sweep

156 of 156 routed as planted
every message reached the planted route

10 of 10 injections caught
and 8 of 8 lookalikes passed

0 sent by the agent
no send path exists

Two hundred CVs for one role, and nobody says the words Annex III out loud.

Screening is a weekend of reading, and every tool sold to fix it either ranks by a model nobody can question or by keywords and calls it AI. CV screening is high risk under the EU AI Act, and the consequences are usually a policy document rather than a design.

THE PRODUCT

Every application scored against the criteria with the evidence shown; a person decides, and the database refuses anything else.

WHO BUYS IT

Companies hiring at volume without a screening system · per role, per month, with the conformity pack

Screening overview



VERDICTS AS PLANTED

360 of 360

302 with a span, 58 no-evidence refusals



PRO
REM

318

0 leaks on

What was stripped before scoring, by class



● Birth date 60 · 19% ● Marital 60 · 19% ● Nationality 60 · 19%



Synthetic corpus of twins: each pair shares the identical marital status, children, military service, nationality and p by definition. Tier 1 against planted answers; a fairness f for a pilot.

360 of 360 verdicts as planted, 318 of 318 protected attributes removed, 30 of 30 pairs identical

Four moves, in this order

- 1

Strip before read
Birth date, marital status, children, military service, nationality, photo and name are removed in code before any scoring reads the text. There is no column for the original.
- 2

Six criteria with a right answer
Years, degree, certificate, language, location, right to work, extracted from the role text by rule. Judgement criteria are labelled judgement.
- 3

Evidence or refusal
Every verdict points at the span of the CV that produced it, or it is no evidence. The schema has no third state.
- 4

A person decides
The decisions table refuses a decider that is not a human staff row, by trigger. Article 22, as a schema.

One candidate, every verdict with its span

Top of the shortlist

CRITERION	REQUIRED	VERDICT	THE TEXT THAT DECIDED IT
years	5	Met	08/2015 - 02/2018 Backend developer, Gamma Retail 08/2018 - 12/2021 ...
degree	BSc	Met	BSc in Computer Science, University of Crete, 2008
cert:aws certified developer	AWS Certified Developer	Met	AWS Certified Developer
lang:english	English C1	Met	Αγγλικά: C1
location	Athens	Met	Τοποθεσία: Remote
permit	EU	Met	Right to work in the EU

ⓘ

A verdict is a span into the document or it is no_evidence. The schema refuses a met or not_met without a span. Silence is an open question for a person, never a guess.

One candidate: every criterion, its verdict, and the text that decided it

Every application scored against the criteria with the evidence shown; a person decides, and the database refuses anything else.

HR Screening Assistant

60 synthetic CVs as 30 matched pairs - Greek and English - SQLite - no model call

SYNTHETIC CORPUS

EU AI Act: high risk, by rule

0 decisions by the agent

0

30 of 30

four fifths rule: 0.80

effect on any verdict: 0

First 12 pairs

PAIR	LANG	TWIN A	TWIN B	
0	el	shortlisted	shortlisted	identical
1	el	not	not	identical
2	el	not	not	identical
3	el	not	not	identical
4	el	not	not	identical
5	en	not	not	identical
6	en	not	not	identical
7	en	shortlisted	shortlisted	identical
8	el	shortlisted	shortlisted	identical
9	el	shortlisted	shortlisted	identical
10	el	not	not	identical
11	el	not	not	identical

Selection rate by group

Group A

12 of 30 shortlisted, rate 0.40

Group B

12 of 30 shortlisted, rate 0.40

Matched pairs: identical qualifications, different protected attributes, identical outcomes

Classified by 04

04 conformity agent, run over this system's own record

Tier

high

Decided by

engine

Articles

Article 6(2), Article 26, Article 27

Resolved in the pinned corpus

3 of 3

⚠

CV screening is Annex III of the EU AI Act. 04 classifies this system high risk by rule, with no model call, and the consequences are built in rather than documented around: human decision by trigger, protected attributes stripped in code, evidence on every verdict, every action logged.

Classified by 04: high risk, by rule, with the articles that resolve

30 of 30 pairs identical
selection ratio 1.00 on twins

318 of 318 attributes stripped
0 leaks on re-scan

0 decisions by the agent
refused by trigger

The proposal is the last one with the name changed, and it still promises two-day delivery.

A rep spends the day not selling: the same proposal again, an hour of tabs per prospect, and the Tuesday nobody called back. A model that writes the proposal invents the delivery term, and a proposal with a term the company does not offer is a contract it did not mean to sign.

THE PRODUCT

Proposals with a source on every claim, research with a source on every line, follow-ups that are arithmetic.

WHO BUYS IT

Small B2B sales teams without an SDR function · per seat, per month

Pipeline overview



OVERDUE FOLLOW-UPS FOUND

49 of 49

0 false alarms; last activity plus stage SLA



Synthetic pipeline: the seed knows which it injected, so every count is against a pl model's in production; the claims go thr

49 of 49 overdue follow-ups found, 24 of 24 unsourced claims stripped, 0 lines below floor

Four moves, in this order

- 1

Price through the floor

Cost plus margin in basis points, integer minor units, rounded against the customer. A discount that would cross it raises; the schema CHECK refuses the row.
- 2

Bind every claim

Prices, terms and references carry a source id. The verifier re-reads each source before the draft exists and strips what does not resolve.
- 3

Follow-ups are dates

Last activity plus the stage SLA. The draft quotes the activity, and the verifier checks that it does.
- 4

Brief from allowed sources

Site, news, registry. No LinkedIn kind exists in the schema. Every line names its source or nothing is stored.

One proposal, every claim with its source

Helios Freight: Integration for account 9

CLAIM	SOURCE	
Agent licence, per seat per month: 3 x EUR 66.60 = EUR 199.80	price:LIC-AGENT	bound
Onboarding workshop, one day: 4 x EUR 888.00 = EUR 3,552.00	price:SVC-ONB	bound
Payment: 40% on signature, 60% on acceptance, 30 days net.	terms:payment	bound
Delivery: pilot starts within 10 working days of signature.	terms:delivery	bound
Warranty: defects fixed at no charge for 90 days after acceptance.	terms:warranty	bound
We delivered a comparable project in logistics: carrier feeds normalised into on	proposal:P-2025-019	bound
Delivery within 2 working days of signature.	terms:express	stripped
We delivered the same system for a competitor of yours last month.	proposal:P-2026-099	stripped

ⓘ Prices come from the price list through the floor; terms from the terms table, references from won proposals in the same sector. Two claims of the kind a model writes were injected into every draft and none survived: the verifier re-reads the source and strips what does not resolve. A term edited after drafting strips on the next verification.

One proposal: every claim bound to its row, and the two injected ones stripped

Proposals with a source on every claim, research with a source on every line, follow-ups that are arithmetic.

The floor

First 12 stored lines

SKU	CLASS	LIST EUR	QUOTED EUR	FLOOR EUR	
LIC-AGENT	key	90.00	66.60	45.00	above floor
SVC-ONB	key	1,200.00	888.00	780.00	above floor
SVC-ONB	trade	1,200.00	1,056.00	780.00	above floor
SVC-TRAIN	trade	600.00	528.00	390.00	above floor
SVC-AUDIT	retail	2,500.00	2,500.00	2,360.00	above floor
SVC-DATA	retail	900.00	900.00	607.50	above floor
SVC-INT	retail	1,500.00	1,500.00	1,260.00	above floor
SVC-DATA	key	900.00	666.00	607.50	above floor
LIC-AGENT	key	90.00	66.60	45.00	above floor
SVC-TRAIN	retail	600.00	600.00	390.00	above floor
SVC-AUDIT	retail	2,500.00	2,500.00	2,360.00	above floor
HW-EDGE	retail	1,800.00	1,800.00	1,680.00	above floor

Floor is cost plus the minimum margin in basis points, rounded up. A class discount that would cross it raises before the line exists, and the schema CHECK refuses the row from raw SQL. 7 lines were refused this run; the key-class probe on the support retainer was refused: 1.

The floor: quoted against cost plus margin, refused rather than clamped

Follow-ups that do not forget

DRAFTED

49

one per deal per due date, by UNIQUE

SECOND SWEEP

0

QUOTE THE LAST ACTIVITY

49 of 49

Most overdue first

DEAL	STAGE	DAYS OVER	LAST ACTIVITY	
Integration for account 6	proposal	82	Asked for a reference in their sector.	draft
Integration for account 12	qualified	78	no activity yet	draft
Licence for account 13	lead	73	Asked for a reference in their sector.	draft
Pilot for account 9	proposal	68	Asked for a reference in their sector.	draft
Integration for account 9	proposal	66	no activity yet	draft
Integration for account 10	negotiation	66	no activity yet	draft
Integration for account 9	proposal	66	Sent the outline of the pilot.	draft
Integration for account 19	qualified	65	Asked for a reference in their sector.	draft
Rollout for account 13	proposal	63	no activity yet	draft
Rollout for account 17	lead	61	no activity yet	draft
Pilot for account 7	negotiation	60	Agreed to review pricing next week.	draft
Pilot for account 7	lead	60	Discussed scope and timing.	draft

Follow-ups due, most overdue first, each quoting the last activity

24 of 24 unsourced claims stripped
before the draft was stored

7 of 7 breaching lines refused
0 stored below floor

49 of 49 overdue found
0 false alarms

Three carriers, three vocabularies, and a customs form typed from a delivery note typed from the order.

Logistics fails between systems: the carrier portal, the warehouse sheet, the ERP that disagrees with both. The mismatch on the customs declaration is found at the border, and a transfer between two shops is in both places or neither.

THE PRODUCT

Three carrier feeds on one timeline, exceptions by rule before the customer asks, documents that cannot disagree with the record.

WHO BUYS IT

Retailers and small distributors shipping with two or more carriers · per location, per month

Fleet overview



EVENTS NORMALISED

264 of 264

csv 87, json 97, pipe 80



UN

9 of

stored as
never ma

Shipments by derived status

■ Picked up 9 · 11% ■ In transit 22 · 27% ■ Attempted



Synthetic fleet: the seed decides each shipment's even
and codes, and keeps the answer. Every count is again
planted on purpose.

264 of 264 events normalised from three formats, 130 of 130 documents verified, 0 mismatches

Four moves, in this order

- 1

One timeline from three formats
A parser per feed format, a code map per carrier. A code the map does not know stays unknown and raises for a person, never mapped to the nearest state.
- 2

Exceptions are rules
No scan for 48 hours, a failed attempt, an address problem, a passed promise, an unknown code. One per shipment per rule.
- 3

Documents from the record
Delivery note, proof of delivery, customs declaration: rendered from the order rows, re-parsed by independent patterns, diffed field by field.
- 4

Verified or not stored
The documents table refuses anything but verified with zero mismatches. A renderer that drops a line is refused, and the bench proves it.

Documents from the record

45 proofs of delivery refused (not delivered), 68 customs declarations refused (inside the EU)

DOCUMENT	ISSUED	
Customs declaration	13	re-parsed, 0 mismatches
Delivery note	81	re-parsed, 0 mismatches
Proof of delivery	36	re-parsed, 0 mismatches

①

CUSTOMS DECLARATION · Order: ORD-2026-2051 · Destination: GB · Incoterm: EXW · Items: · DOCK-4 HS 8471.80 x4 value EUR 636.00 weight 2800 g · SSD-1T HS 8523.51 x4 value EUR 284.00 weight 240 g · MSE-E HS 8471.60 x4 value EUR 132.00 weight 440 g · Total value: EUR 1,052.00 · Total weight: 3480 g

①

A document is rendered from the order rows, re-parsed by patterns that know nothing about the template, and diffed field by field. Only a zero-mismatch document is stored; the schema refuses any other. Values are integer minor units summed from lines.

Documents issued, each re-parsed to its rows, with the refusals counted

Three carrier feeds on one timeline, exceptions by rule before the customer asks, documents that cannot disagree with the record.

Logistics

81 synthetic shipments - ACS, ELTA, Speedex - CSV, JSON, pipe - SQLite - no model call

SYNTHETIC FLEET0 document mismatches9 unknown codes kept, 0 guessed

First 14 exceptions

ORDER	CARRIER	RULE	CASE	MESSAGE
ORD-2026-2000	ACS	No scan 48h	open	draft
ORD-2026-2000	ACS	Late vs promise	open	draft
ORD-2026-2004	ELTA	Unknown code	open	draft
ORD-2026-2006	ACS	No scan 48h	open	draft
ORD-2026-2006	ACS	Late vs promise	open	draft
ORD-2026-2010	ELTA	No scan 48h	open	draft
ORD-2026-2010	ELTA	Late vs promise	open	draft
ORD-2026-2011	SPEEDEX	No scan 48h	open	draft
ORD-2026-2011	SPEEDEX	Late vs promise	open	draft
ORD-2026-2013	ELTA	No scan 48h	open	draft
ORD-2026-2013	ELTA	Late vs promise	open	draft
ORD-2026-2013	ELTA	Unknown code	open	draft
ORD-2026-2020	SPEEDEX	No scan 48h	open	draft
ORD-2026-2020	SPEEDEX	Late vs promise	open	draft

76 customer messages drafted from the shipment rows, each waiting for a human staff row to approve it; the trigger

Exceptions raised before the customer asked, each with a draft waiting for a person

Transfers as two events

LEDGER ANOMALIES0units never in two places

NEGATIVE STOCK0

CANCEL WITHOUT A REASON REFUSED1

Ten transfers

#	SKU	QTY	FROM	TO	STATUS	CANCEL REASON
1	HUB-7	10	wh	north	Fulfilled	
2	SSD-1T	4	wh	shop	Fulfilled	
3	CBL-HD	6	wh	shop	Fulfilled	
4	KBD-87	9	wh	shop	Fulfilled	
5	PWR-65	9	wh	north	In transit	
6	MSE-E	7	wh	shop	In transit	
7	DOCK-4	10	wh	shop	Requested	
8	CASE-13	10	wh	shop	Cancelled	wrong destination
9	HUB-7	4	wh	shop	Requested	
10	SSD-1T	9	wh	north	Requested	

Transfers as two events: 0 ledger anomalies, a cancellation that kept its reason

264 of 264 events normalised
9 of 9 unknown codes kept

130 of 130 documents verified
0 mismatches

0 ledger anomalies
a unit is never in two places

The same four rules in all ten

A deterministic core, a model only at the edges

Pricing, policy, classification and reconciliation are code. The model parses language and explains results; it never decides something that has a right answer.

Measured, not asserted

Every figure on these screens is produced by the product's own code and re-verified by its test suite. The suites redden under mutation, so green is evidence.

Structural defences, not instructions

The margin floor, the replay nonce, the tenant scope and the durable ledgers are code a prompt cannot be talked out of. They hold when detection fails.

Extended by the client, not by us

Capabilities are files, not deployments. A non-engineer adds a platform, a skill or a research source and it is live on the next request.

WHAT THIS SHOWS

The hard part is not the model. It is everything around it.

Any of these can go to a first customer as it stands: the interfaces are real, the rules are enforced in code, and every figure on every screen comes from running the system. The next step for any one of them is a pilot against live data.

Deterministic where it counts

money, policy, law, arithmetic

Extensible by the client

capabilities are files, not deployments

Auditable by default

every decision traced and replayable