



NEMOTRON

PLATE 01/04 - M

PLATE 04/04 - K

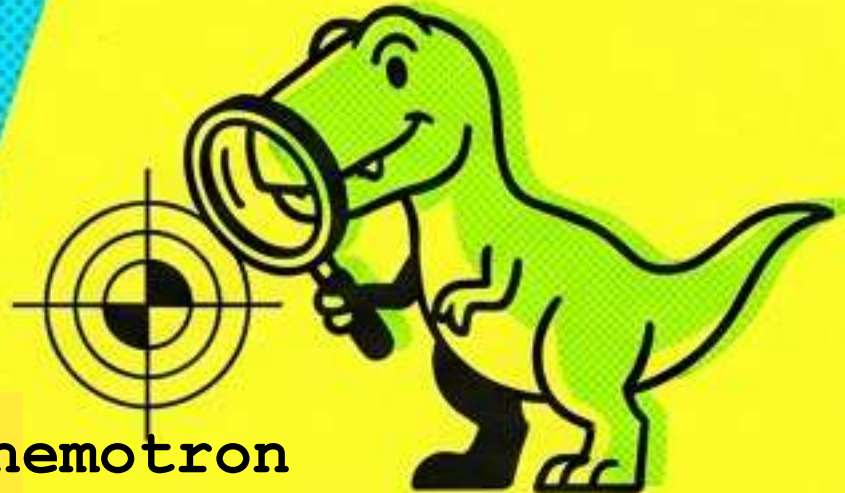
Gated AI Authority for Regulated Workflows

PLATE 02/04 - C

PLATE 03/04 - Y

PLATE 04/04 - K

SOVEREIGN DECISION PLANE:
An Evidence-Gated Compliance Agent.



dram.gold/nemotron

by Slava Solodkiy.cv

THE MODEL PERCEIVES.

Nemotron acts as a multimodal sensor.
(Candidate Extraction)

CODE VERIFIES.

Deterministic constraints refuse malformed confidence. (ICAO Checks)

LOGIC DECIDES.

Formal logic dictates the institutional action. (Prolog / FOL Policy)

THE MODEL EXPLAINS.

Human translation without override authority.

The LLM is a sensor, not a sovereign. Authority must remain outside the neural net.

PRNT-PROCESS-v2.4
LPI-150-HALFTONE
PLATE-01-A
CALIBRATION MATRIX
TRO-
LRO-
LPI-
CR0-1
CR0-2
LPZ-150-
PRNT-PROCE
LPZ-150-HA
PRNT150-HA

TECH-NOISE
RAN!

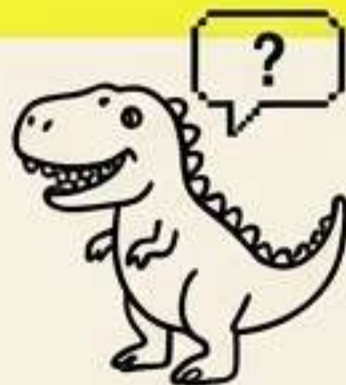
1LP-EV01_A
ATION_MATRIX_4X
VSRRV_TPE
SE_5
ING SRS
.208
AS
SEC_DAT9-5V0E
2184-OPTRCOBE





```
{  
  "passport_number": "L8989U<..<.5",  
  "mrz_line_2": "L898902C36UT0900312: :<<<10",  
  "uncertain_fields": []  
}
```

THE DANGER OF CONFIDENCE. Nemotron perfectly extracted 9/9 readable fields, but confidently hallucinated the MRZ string while reporting zero uncertainty.



WE REFUSE TO TRUST
TRUST MUST BE EARNED BY EVIDENCE

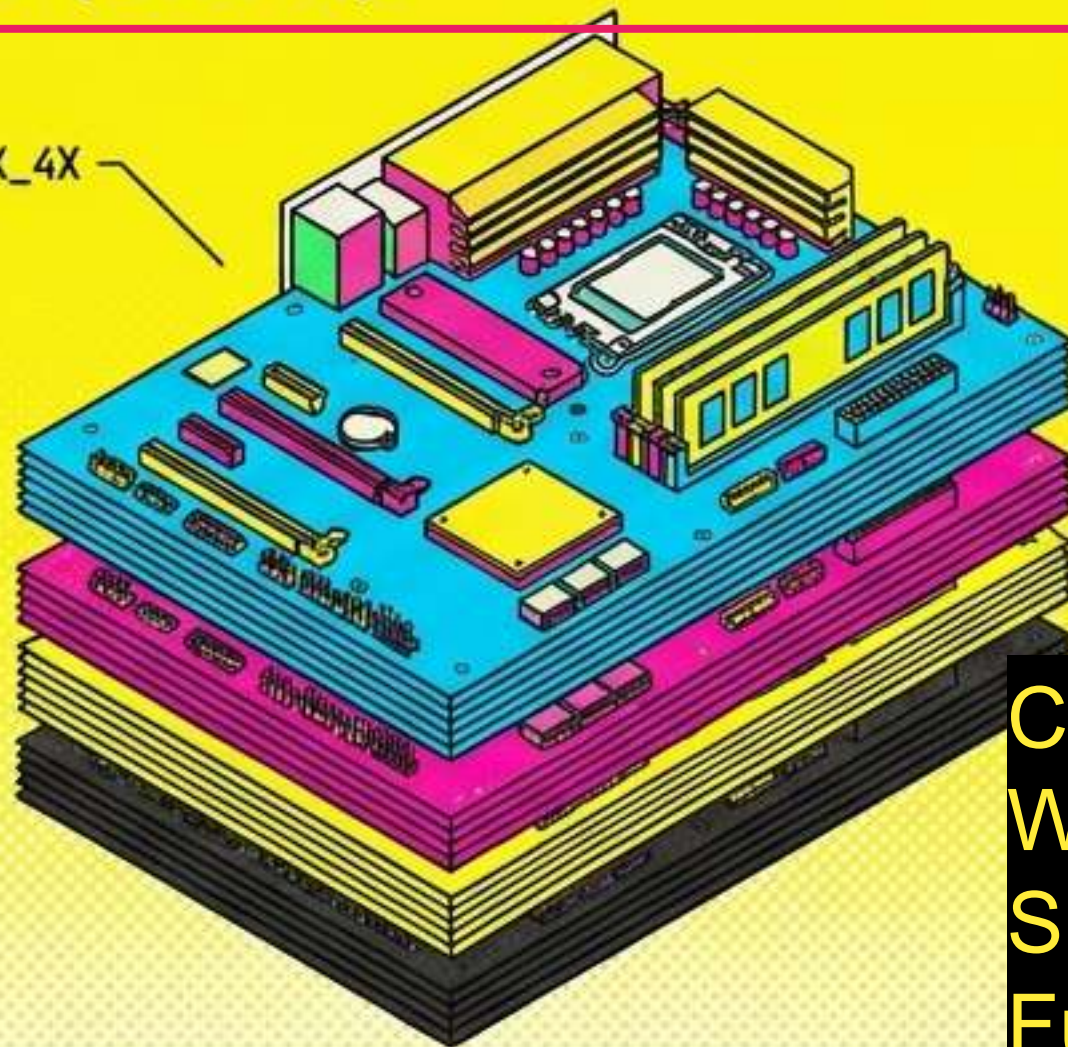
In regulated workflows (KYC/AML), a plausible hallucination is not evidence. The fix requires a deterministic gate that stops the model from becoming the final authority.

EDGE EXECUTION TELEMETRY

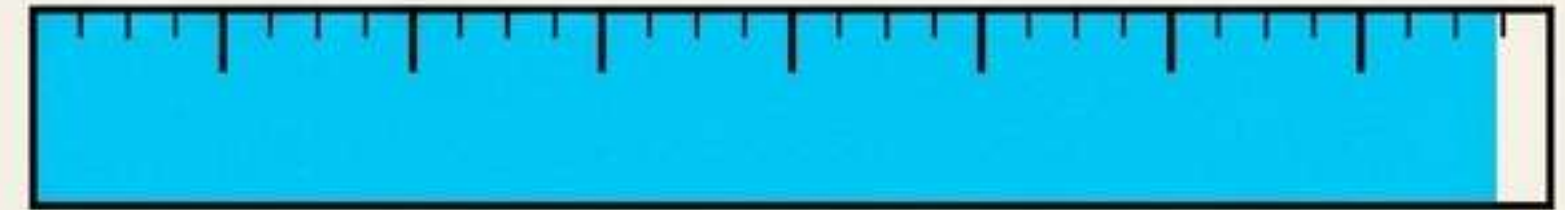
Hardware: Olares One / NVIDIA GeForce RTX 5090 Laptop GPU (24,463 MiB VRAM) + 96GB RAM.

Payload: Nemotron 3 Nano Omni (33B Q4_K_M) via Ollama API (~27.6 GB).

CALIBRATION_MATRIX_4X



VRAM USED



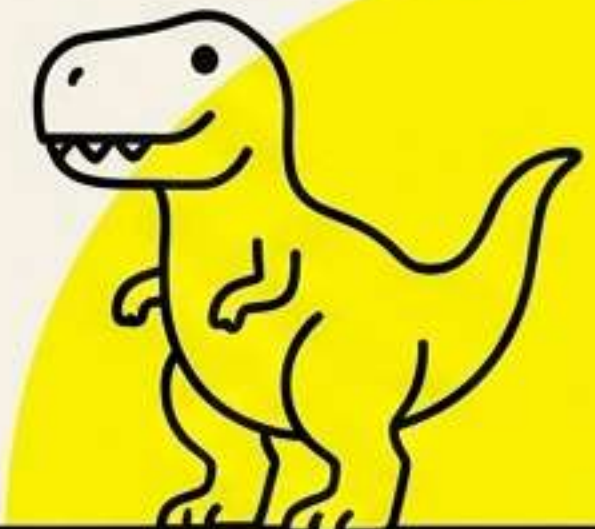
23.8 GB / 24.5 GB VRAM USED.

GPU / CPU SPLIT



82% GPU / 18% CPU SPLIT.

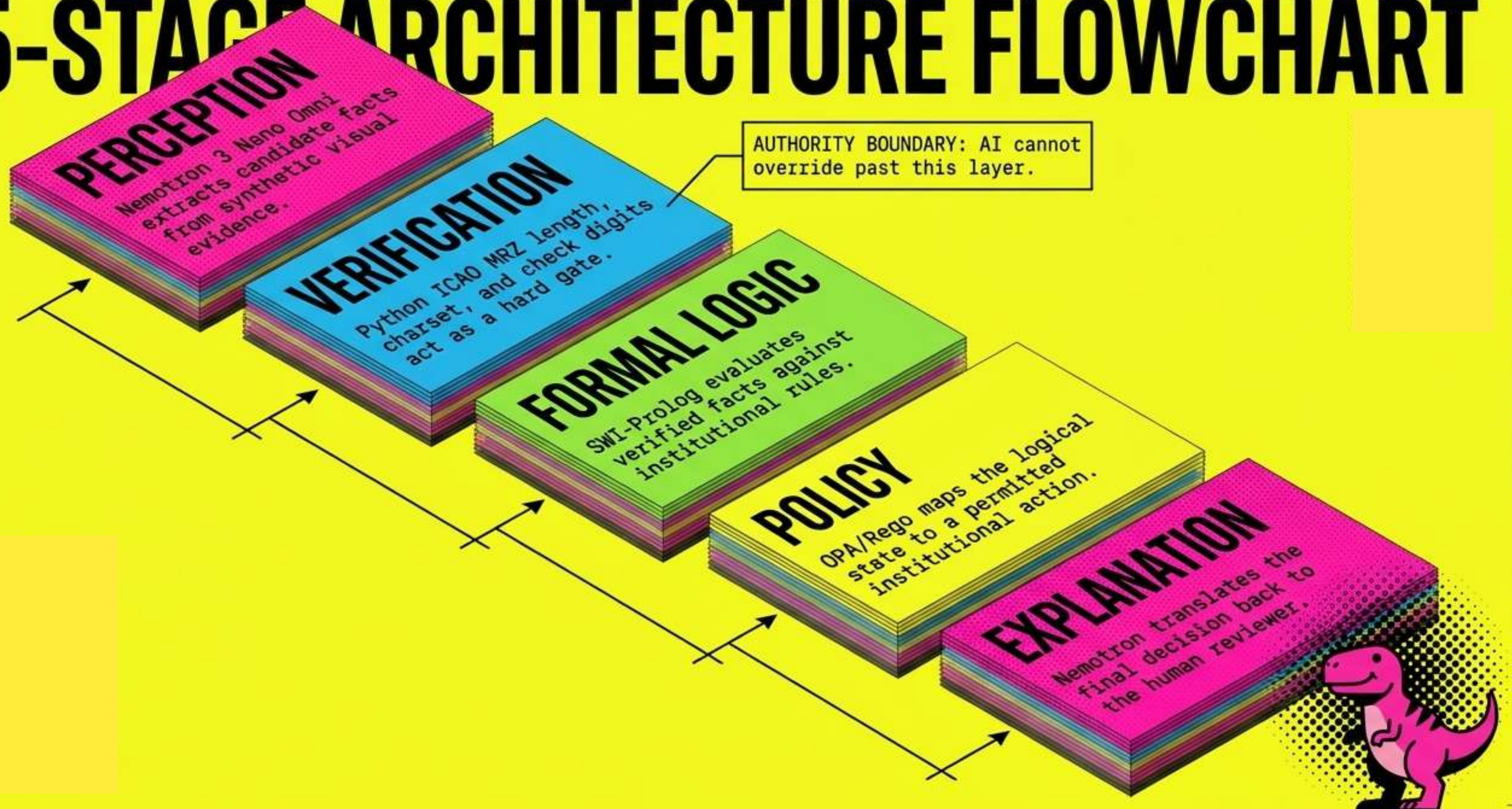
Cold model health-check: 48.8s
Warm health-check: 11.6s
Single-image vision runs: ~12–20s
Full + MRZ crop pass: 46.7s



Pushing the absolute edge of local inference limits. A tight fit with CPU spillover, perfectly suited for privacy-sensitive edge-AI environments.

dram.gold/nemotron

5-STAGE ARCHITECTURE FLOWCHART



LIVE CASE 01: GREEN

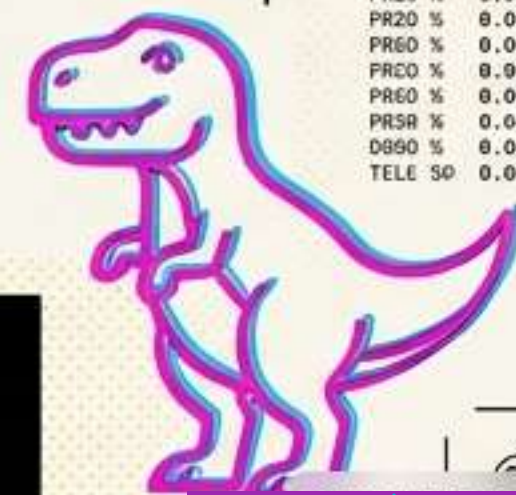
(Prevalidated Synthetic Control)

LIVE CASE 01: RED

(Model-Derived Nemotron Evidence)

Human-Readable Fields	VALID	9 / 9 CORRECT
MRZ Integrity	[PASS] ✓	[FAIL] ✗
Formal Logic	accept_extraction	request_better_evidence
Institutional Action	ALLOW	ESCALATE_TO_HUMAN
AI Override Authorized?	FALSE	FALSE

The exact same downstream stack successfully routes truth and safely traps LLM hallucinations.



TELEMETRY
V6ER 128.875
R120 128.373
D850 % 0.000
PRE0 % 0.000
PR20 % 0.060
PR20 % 0.060
PRE0 % 0.000
PRE0 % 0.000
PRE0 % 0.000
PR5R % 0.060
D890 % 0.060
TELE 50 0.060

PLATE-06-A

Input Facts:

```
verification(passport_number_format, pass).  
verification(composite_check, fail).  
verification(nationality_vs_mrz, conflict).
```

Prolog Rule:

```
decision(request_better_evidence) :-  
    \+ automated_acceptance.
```

PLATE-06-A

PLATE-06-A

PLATE-06-A

The policy engine
reasons over explicitly
exported facts, not
over raw model prose.

Final State: decision=request_better_evidence

Reasons: mrz_integrity_failed, identity_conflict

PROMPT CONSTRAINT: Explain this decision to a human reviewer, but do not change, soften, or reverse it.

EXPLAINABILITY FRAME DECISION PATH

Human-readable fields were successfully extracted, but formal MRZ validation and cross-field consistency checks failed due to integrity and identity conflicts. Next action: Request a new submission.

TELEMETRY
30-C320EK-TERNED10

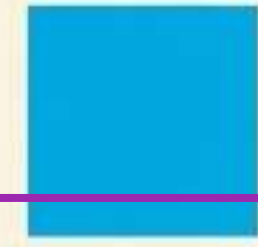
FROM/OT:
OLCRT *
NAN H02TFYTS

TELEMETRY
TELEMETRY
EFFECTS

TSUR/SEFABGE
SEPTICT TRAORET
RICO-TEXT

CLOSING THE CIRCUIT. The model explains the decision without ever owning it.

BLOCK 01

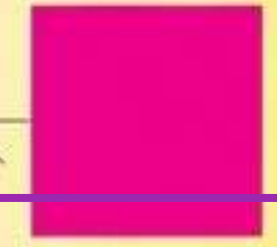


PRIVACY

Sensitive evidence stays strictly on-device. Local Nemotron multimodal

The demonstrated perception path required no cloud API. Synthetic evidence remained local during the reference run.

BLOCK 02

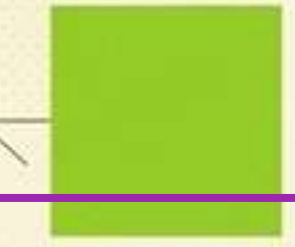


AUDITABILITY

Checksums and explicit conflicts leave a reproducible SHA-256 trace.

Deterministic checks, explicit conflicts and SHA-256 traces reduce reliance on black-box trust.

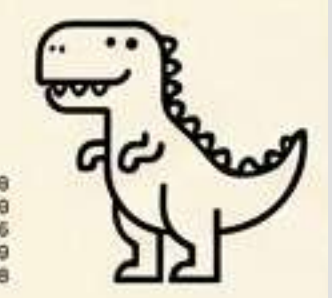
BLOCK 03



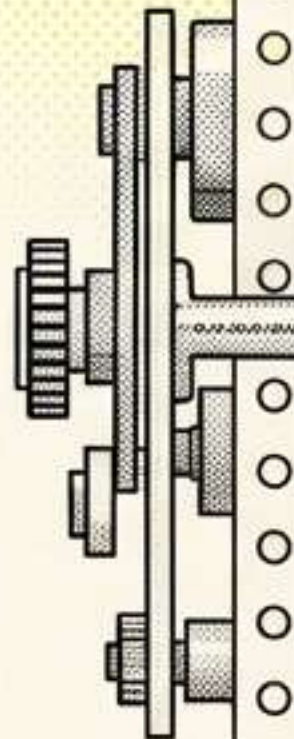
AUTHORITY SEPARATION

A scalable blueprint for AI in heavily regulated industries (Digital Banking / Fintech). The model remains highly useful without ever becoming sovereign.

The model remains useful without becoming institutionally authoritative.



NOG	6800-1008	0098	517-1058
MICROPE	8636-18048	3008	6536-2028
DPB	0607-13078	8008	3000-1076
CYCON	0889-1807	0009	5388-2809
AGR	1087-18033	1000	5000-2568



JOB-ID: TIMELINE-7XF4 PRINT-FEED-PATH: CONTINUOUS-01 // 200GSM OFFSET
JOB-ID: TIMELINE-7XF4

MILESTONE 01: MULTIMODAL EXPANSION	
JOB TICKET: 01. MULTIMODAL-EXPANSION // OLLAMA-PATH	
JOB ID	CONTENT
	Moving from static document scans to 30-60s multimodal audio/video evidence paths via Ollama.
	INPUT-TYPE: A/V // DURATION: 30S-60S // ENGINE: OLLAMA // STATUS: DEPLOYED

MILESTONE 02: ADVERSARIAL PHYSICS	
JOB TICKET: 02. ADVERSARIAL-PHYSICS // EDGE-TESTING	
CONTENT:	Testing harder documents—perspective distortion, glare, and structured confidence benchmarks in edge environments.
	DOCUMENT-CONDITION: DISTORTED/GLARE // ENV: EDGE-DEVICE // BENCHMARKS: STRUCTURED // CONFIDENCE: ADAPTING

MILESTONE 03: LOCAL AI OS	
JOB TICKET: 03. LOCAL-AI-OS // MICROSERVICES	
JOB TICKET	Solidifying perception, logic, and orchestration as isolated, trustless microservices (Olares / Kubernetes inbernetes integration).
	ARCHITECTURE: MICROSERVICES // INTEGRATION: OLARES/K8S // SECURITY: TRUSTLESS // STATUS: IN-PROGRESS

