People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Kasdi Merbah University – Ouargla
Faculty of New Technologies of Information and Communication
Department of Electronics and Communications

Master's Thesis

Specialty: Instrumentation

Multi-Agent System for Smart Factory Management

Presented by: **Bilal AISSAOUI**

President:	Dr. Dahraoui Nadia
Examiner:	Dr. Fares Bennaceur
Supervisor:	Dr. Youcefa Abdelmadjid
Incubator Representative:	Dr. Delhoum Khalida
Economic Partner:	Eng. Boukrara Karim

Academic Year: 2025–2026

Abstract

Manual retrieval of engineering knowledge from Piping and Instrumentation Diagrams (P&IDs), diagnostic manuals, and technical procedures remains a critical bottleneck in industrial plant operations: every minute of diagnostic delay carries direct costs in production and safety. Existing approaches either require extensive manual digitisation or rely on general-purpose AI systems that produce plausible but plant-unverifiable outputs.

This work presents MAESTRO, a governed Multi-Agent System that transforms P&ID drawings and industrial documents into structured, audit-traceable plant intelligence through a seven-layer architecture spanning OCR-led extraction, ISA-5.1 semantic enrichment, human-governed knowledge promotion, and agentic diagnostic reasoning. Governance is enforced architecturally: the agent's tool contracts structurally prevent control-action execution, independent of any prompt or instruction.

Validated on a refrigeration-loop P&ID, MAESTRO extracted complete instrument and equipment candidates with zero false positives – eliminating engineering time spent on phantom candidates – and executed multiple diagnostic scenarios with response times between 8 and 18 seconds, combining topology traversal, live telemetry retrieval, and 85-chunk expert knowledge retrieval within a single governed pipeline.

These results demonstrate that plant-aware agentic reasoning is feasible at Technology Readiness Level 4–5 without sacrificing governance or auditability, establishing a reproducible pathway toward deployable industrial AI.

Keywords: Multi-Agent Systems, P&ID Digitisation, Industrial Knowledge Graph, Human-in-the-Loop Governance, Agentic RAG, ISA-5.1.

Contents

Abstract **i**

List of Abbreviations **ix**

1 Industrial Intelligence: From Silent Knowledge to Governed Agentic Reasoning **1**

1.1 General Context: The Silent Knowledge Problem in Industrial Environments 1
1.2 From Industry 4.0 to Industry 5.0: A Necessary Evolution 2
1.3 State of the Art and Research Gap 2
1.3.1 Industrial Knowledge: Structure, Preservation, and Loss 2
1.3.2 P&ID Digitisation: From Pixels to Semantics 3
1.3.3 Industrial Knowledge Graphs and the Hybrid Knowledge Layer . . 5
1.3.4 Agentic Systems and Retrieval-Augmented Generation 5
1.3.5 Human-in-the-Loop Governance in Industrial AI 7
1.3.6 Research Gap: A Field in Productive Fragments 7
1.4 Motivation and Intellectual Origin 8
1.5 A Central Conceptual Contribution: Confidence Is Not Trust 9
1.6 Research Problem Statement 9
1.7 Research Hypothesis 9
1.8 Research Objectives 9
1.9 Research Questions 10
1.10 Main Contributions 11
1.11 Scope and Limitations 11
1.12 Organisation of the Thesis 12

2 The MAESTRO Architecture and Governance Framework **14**

2.1 Conceptual Overview: Governance by Design 14
2.2 Architectural Governance and Safety Contracts 14
2.2.1 The Vulnerability of Prompt-Based Guardrails 14
2.2.2 Structural Tool Contracts: Read-Only by Construction 16
2.2.3 The Single Governed Write Gate 16
2.3 Architecture of Record and Infrastructure 16
2.3.1 Vision API 16
2.3.2 Platform 17
2.3.3 Knowledge Layer Infrastructure 17

2.4 The Governed Methodological Pathway . . . 18
2.5 The Seven Architectural Layers . . . 18
2.5.1 Vision Layer: OCR-Led, Geometry-Aware, ISA-5.1-Informed . . . 20
2.5.2 Topology Layer . . . 21
2.5.3 Calibration Layer . . . 24
2.5.4 Semantics Layer . . . 24
2.5.5 Governed Graph Layer and the Hybrid Knowledge Layer . . . 24
2.5.6 Agentic Reasoning Layer . . . 26
2.5.7 Human-in-the-Loop Governance Sublayer . . . 29
2.5.8 Closed-Loop Decision Support Architecture . . . 29
2.6 Architectural Evolution as Method . . . 30
2.7 Validation Scope . . . 32
2.8 MAESTRO Architectural Principles . . . 32
2.9 Mathematical Formalisation . . . 34
2.9.1 Embedding and Retrieval . . . 35
2.9.2 ISA-5.1 Entity Validity Test . . . 35
2.9.3 Vector Store Governance Constraint . . . 35
2.9.4 Governed Knowledge Graph . . . 36
2.9.5 Governance State Machine . . . 37
2.10 Chapter Conclusion . . . 38

3 Implementation and Experimental Validation 40
3.1 Introduction . . . 40
3.2 Experimental Implementation and Testbed . . . 40
3.2.1 Containerised Software Stack . . . 40
3.2.2 The Validation Dataset . . . 40
3.3 Vision Layer Results . . . 40
3.3.1 YOLO Experiment: Training Evidence and Structural Limitations . 42
3.4 Topology Layer Results . . . 49
3.5 Calibration and Semantics Results . . . 50
3.6 The Governed Graph Layer: Below the Surface . . . 50
3.7 Agentic Reasoning Layer Results . . . 52
3.7.1 Cross-Scenario Governance Analysis . . . 61
3.8 Ablation Analysis . . . 64
3.9 Latency and Efficiency Benchmarks . . . 65
3.10 Synthesis and Discussion . . . 67
3.10.1 A Key Finding: Legacy Knowledge in Governed Systems . . . 67
3.10.2 Cross-Layer Architectural Patterns . . . 68

3.10.3 Industrial Threat Model and Security Analysis 69
3.11 Diagnostic Scenarios in Dialogue with the Literature 69
3.12 Formal Answers to the Research Questions 71
3.13 MAESTRO vs. Closest Alternatives: Direct Comparison 71
3.14 Research Contributions with Evidence 73
3.15 Scope and Inference Boundaries 73
3.16 Chapter Conclusion 73

4 Conclusion and Future Directions **76**
4.1 Critical Discussion 76
4.1.1 Agent Autonomy versus Architectural Governance 76
4.1.2 Scalability to Multi-Unit Industrial Facilities 76
4.2 Conclusion to the Research Problem 77
4.3 Answer to the Main Research Question 77
4.4 Scientific and Industrial Contributions 77
4.5 Technology Readiness Level Assessment 78
4.6 Limitations of the Current Work 79
4.7 Future Research Directions 79

List of Figures

1.1 Refrigeration loop P&ID with ISA-5.1 reading key. Solid lines: process pipes. Dashed lines: control signals. 4
1.2 Why MAESTRO uses both Graph Search and Vector Search. A single engineer query activates Neo4j (topology, governance state, cascade impact) and Supabase/pgvector (expert documents, tacit knowledge, operating limits) simultaneously. Neither strategy alone can answer both the structural and semantic dimensions of a plant diagnostic query. 6

2.1 MAESTRO runtime architecture: four layers separated by a single governed write gate. All agent tools are read-only; `approve-edge` is the only write path, requiring human-verified identity. 15
2.2 The n8n diagnostic composer node. Governance flags (`read_only`, `candidate_only`, `approved`, `graph_truth`) are enforced in the data structure before the model is invoked. The system-message prompt only restates a constraint the architecture already guarantees. 17
2.3 MAESTRO governed architecture: nine-stage pathway from engineering evidence to industrial decision support, across seven core layers. 19
2.4 YOLO versus ISA-5.1 rule-based extraction. Identical circle symbols for all instrument types make vision-only detection structurally unreliable. . . 20
2.5 Vision Layer Stage 1: OCR-led geometry binding pipeline. From raw P&ID image to 15 labelled candidate objects via ISA-5.1 FrozenSet validation, distance-based circle contour binding, and SharedDisplay channel expansion. 22
2.6 ANSI/ISA-5.1 engineering knowledge translated into deterministic extraction rules – no training data required. 23
2.7 ISA-5.1 tag validation workflow. Non-compliant tags are rejected at extraction time. 24
2.8 Vision Layer Stage 2: 3-pass candidate deduplication. IoU, tag-identity, and IoS containment passes reduce 15 raw candidates to 12 clean ones; all output candidates carry `approved=false` and `review_required=true`. . . 25
2.9 OCR grouping overlay on the validated P&ID. Green bounding boxes show ISA-5.1-tagged candidates; all 12 extracted with zero false positives after three-pass deduplication. 26
2.10 Structural topology overlay on the validated P&ID. Red lines: process flows; blue dotted lines: signal paths. 27

2.11 MAESTRO Agentic Reasoning Loop (Agent = Model + Tools + Knowledge). Control-action requests are intercepted before LLM invocation; allowed queries traverse five governed tools. 28
2.12 The five governed tool nodes in the n8n workflow. All contracts are read-only; no write or control-action interface is exposed to the agent. 29
2.13 VectorStore ingestion workflow. Documents are chunked, embedded via HuggingFace, and stored in Supabase with full provenance metadata. Validated scenarios use the 104-chunk corpus (13 sources); 737-chunk total includes all development batches. 29
2.14 MAESTRO Plant Knowledge tab showing the Semantic Evidence Upload interface and governance readiness panels. Both Hybrid Readiness and Simulation Readiness carry explicit governance notices: neither approves truth, neither triggers topology mutation, and neither result bypasses the HITL gate. 30
2.15 Governance validation pipeline. Only HITL-approved entities are accessible to the agent layer. 30
2.16 MAESTRO Draft Editor in pre-approval state. Banner confirms DRAFT REVIEW / NOT APPROVED / NOT GRAPH TRUTH / HUMAN REVIEW REQUIRED. Green dashed edges are candidate relations awaiting HITL verification. 32
2.17 Interactive topology graph during draft review. Candidate nodes (FIC-42, PDT-42, FV-42, TT-41 and others) and their candidate relations are visible in Draft Review state; no element carries approved status until the engineer completes HITL verification. 34
2.18 MAESTRO Semantic Knowledge Injection and Retrieval Pipeline. Phase 1 chunks and embeds industrial documents into Supabase pgvector; Phase 2 retrieves top-k evidence chunks via cosine similarity for agent reasoning. . . 36
2.19 Governed semantic retrieval pipeline (schematic). Queries are matched against the vector store; retrieved chunks carry candidate-only provenance flags. 37
2.20 Labelled property graph $G = (V, E_G, P_V, P_E)$ instantiated on the validated refrigeration-loop case. Dark-blue nodes carry $P_V(v)[\texttt{approved}] = \texttt{true}$ and are reachable via agent tool contracts; the dashed node (`approved=false`) is structurally excluded by the agent visibility invariant. 37
2.21 Governed knowledge lifecycle. Human authorization is the only valid path to the approved state. 38

3.1 MAESTRO Architectural Dashboard: Deterministic Vision and Governed Hybrid RAG. 42
3.2 MAESTRO Containerised Runtime — Validation Run. 43
3.3 Algorithmic Flow and Functional Blocks of the MAESTRO Vision Pipeline. 44
3.4 YOLO Training Dataset Configuration (`data.yaml`). 44
3.5 Three Structural Limitations of Standalone YOLOv8 in P&ID Digitisation. 45
3.6 YOLOv8n Detection Results and Training Class Distribution. 45
3.7 YOLOv8n Training Curves (30 Epochs, Dual T4 GPU). 47
3.8 Extraction Evidence Panel — PDT-42 (Confidence 99.8%). 48
3.9 MAESTRO Vision Pipeline: P&ID to Structured Topology. 49
3.10 Validated MAESTRO Knowledge Graph in Neo4j (12 Nodes, 22 Relationships). 51
3.11 MAESTRO Platform Topology After HITL Review. 52
3.12 FV-42 Golden Scenario — n8n Execution Trace (ID#220, 17.951 s). 54
3.13 FV-42 Golden Scenario — Agent Diagnostic Output (Arabic, 17.951 s). . . 54
3.14 FV-42 Golden Scenario — Topology Context. 55
3.15 FV-42 Golden Scenario — Cascade Impact Traversal. 56
3.16 TT-41 Governance Refusal — Agent Console View. 57
3.17 TT-41 governance refusal – n8n workflow Editor. Blocked Request Router activated; no LLM or tool invoked. 57
3.18 TT-41 governance refusal – MAESTRO Agent Console (English, 8.19 s). Setpoint modification refused; governance flags confirmed. 57
3.19 Governance enforcement banner. Footer: ADVISORY CANDIDATE ONLY / NO CONTROL ACTION TAKEN / HITL REQUIRED. 58
3.20 FT-42 Oscillation Scenario — n8n Execution Trace (ID#229, 11.103 s). . . 58
3.21 FT-42 LOW-LOW Alert Scenario — n8n Execution Trace (ID#245, 15.623 s). 58
3.22 FT-42 LOW-LOW Alert — Agent Response (Arabic). 59
3.23 FIC-42 Expert Knowledge Scenario — n8n Execution Trace (ID#249, 10.975 s). 59
3.24 FIC-42 Evidence Hub (VECTORSTORE / READ-ONLY). 85 chunks from 9 source documents across 7 evidence categories. 60
3.25 FV-42 Prohibited Actions Scenario — n8n Execution Log (ID#301, 9.917 s). 60
3.26 FV-42 Prohibited Actions — Agent Console (ID#301, Arabic). 61
3.27 MERGE Key Refactoring: Deduplication to 8 Canonical Topology Edges. . 62
3.28 Governed Diagnostic Response Time by Scenario. 67
3.29 Knowledge Graph Workspace — Provenance Mode. 68

List of Tables

1.1 Research Gap Mapping – Five-Layer Coverage. — = outside stated scope. Final row: requirements for full industrial deployment. 7
1.2 Extraction Cost Comparison – Training Requirements and Operational Constraints. N/M = not measurable on the same metric. FPR row: internal controlled comparison on the same P&ID. 8

2.1 Eight-Stage Governed Pathway from P&ID to Agent Response 18
2.2 Architectural Decisions Emerging from Observed Failure Modes 31
2.3 MAESTRO Architectural Governance Principles 33

3.1 MAESTRO Implementation Environment and Technology Stack (Unified Database Version) 41
3.2 YOLOv8n Training Dataset: Class Distribution. 43
3.3 YOLO-Primary Architecture: Structural Obstacles and MAESTRO's Resolutions. 46
3.4 Extraction Approach Comparison. 47
3.5 Physical Entity Layer – Validated Neo4j Node Inventory 50
3.6 Eight APPROVED_TOPOLOGY_EDGE Relationships (Neo4j Query). . 51
3.7 Validated Diagnostic Scenario Summary. 53
3.8 Governance Operating Modes Across Six Validated Scenarios. 62
3.9 Quantitative Results per Architectural Layer — Validated Case. 63
3.10 Validated Results with Evidence and Scope. 64
3.11 Ablation Analysis: estimated impact of removing each architectural component on validated results. 65
3.12 MAESTRO Operational Results Against Documented Baselines. 66
3.13 Industrial Threat Model: Eight Threats and MAESTRO Defences. 70
3.14 FV-42 Golden Scenario — Workflow Trace (ID#220, 10 Stages). 71
3.15 MAESTRO vs. Closest Alternatives — Validated Results. 72

4.1 TRL Development Roadmap – Three-Phase Architecture. 78

List of Abbreviations

AI	Artificial Intelligence
BFS	Breadth-First Search
CRUD	Create, Read, Update, Delete
DCS	Distributed Control System
DEXPI	Data Exchange in the Process Industry
DWSIM	Dynamic and Steady-State Simulator
FDRKG	Fault Diagnosis and Reasoning Knowledge Graph
FPR	False Positive Rate
GKL	Governed Graph Layer
HKL	Hybrid Knowledge Layer
HITL	Human-in-the-Loop
IoU	Intersection over Union
ISA	International Society of Automation
ISA-5.1	ANSI/ISA-5.1: Instrumentation Symbols & Identification
JWT	JSON Web Token
LLM	Large Language Model
MAESTRO	Project name: Multi-Agent System for Smart Factory Management
MAS for SFM	Multi-Agent System for Smart Factory Management
MAS	Multi-Agent System
MQTT	Message Queuing Telemetry Transport
OCR	Optical Character Recognition
OPC-UA	OPC Unified Architecture (industrial protocol)
P&ID	Piping and Instrumentation Diagram
PLC	Programmable Logic Controller
RAG	Retrieval-Augmented Generation
RCA	Root Cause Analysis
RQ	Research Question
TRL	Technology Readiness Level
VectorStore	Supabase/pgvector semantic retrieval layer
YOLO	You Only Look Once – real-time object detection architecture

Chapter 1

Industrial Intelligence: From Silent Knowledge to Governed Agentic Reasoning

1.1 General Context: The Silent Knowledge Problem in Industrial Environments

Industrial plants are, by their nature, dense accumulations of engineering knowledge. Every valve position, every control loop, every instrument tag on a Piping and Instrumentation Diagram (P&ID) encodes a decision made by an engineer at some point in the plant's history. This knowledge is not incidental. It is, arguably, the most operationally critical asset a facility possesses.

And yet, for most industrial facilities today, this knowledge remains practically inaccessible to automated systems.

The P&ID is perhaps the clearest illustration of this paradox. According to ANSI/ISA-5.1 standards [5], these diagrams encode equipment identity, instrumentation tags, control logic, and process topology in a highly structured symbolic language that any trained engineer can read fluently. But to a computational reasoning system, the raw P&ID is an opaque image – a collection of pixels that carries no semantic meaning without the full pipeline of interpretation.

The gap between what the diagram means to a human engineer and what it represents to a machine is the core reason why intelligent systems in industrial environments so frequently fail to perform at the level their theoretical capabilities suggest.

This work began from an attempt to close that gap.

The broader industrial context makes the challenge urgent. The vision articulated under the Industry 4.0 framework – self-organising, adaptive, data-driven manufacturing – depends on the availability of structured, machine-accessible plant knowledge [6, 24, 25]. Predictive maintenance, fault propagation analysis, intelligent decision support: all of these require systems that understand not just sensor readings, but the engineering meaning behind those readings. A temperature anomaly in isolation is ambiguous. The same anomaly, interpreted in the context of a compressor–knockout drum–evaporator refrigeration loop, with knowledge of which temperature transmitter is connected to which shared display and which control valve it governs – that is actionable information.

MAESTRO – Multi-Agent System for Smart Factory Management – is the response to this structural problem.

1.2 From Industry 4.0 to Industry 5.0: A Necessary Evolution

The Industry 4.0 vision brought with it a powerful promise: automation, cyber-physical integration, and data-driven decision-making at industrial scale. What it did not fully anticipate was the reliability problem that emerges when artificial intelligence systems operate without adequate grounding in verified, plant-specific knowledge.

Large language models are the most visible current example of this problem. They are capable of generating fluent, apparently knowledgeable responses to industrial queries. But they do not know the plant. They cannot verify whether a detected relationship between two instruments is structurally plausible within the actual facility topology. And crucially, they cannot stop when evidence is insufficient.

The emerging Industry 5.0 paradigm addresses this directly. The European Commission's 2021 report [7] reframes the human role as a collaborative authority who governs the outputs of intelligent systems. MAESTRO is explicitly designed at this intersection: it automates P&ID knowledge construction (Industry 4.0) while enforcing mandatory human governance over unverified knowledge (Industry 5.0).

1.3 State of the Art and Research Gap

Having established the industrial problem and the Industry 5.0 imperative, this section surveys the technical landscape across the four domains that MAESTRO integrates: P&ID digitisation, industrial knowledge graphs, agentic systems with retrieval-augmented generation, and human-in-the-loop governance. Each domain has matured in isolation; the research gap is the absence of any work that unifies all four into a single governed pathway. The review concludes by mapping this gap systematically.

1.3.1 Industrial Knowledge: Structure, Preservation, and Loss

The Five Layers of Industrial Knowledge Representation

Abonyi, Nagy, and Ruppert [9] argue that effective industrial knowledge management requires a layered semantic framework. Industrial knowledge in process-plant contexts decomposes into five layers:

Ontology defines the formal conceptual schema: entity classes, properties, and permissible relationships. ANSI/ISA-5.1 [5] constitutes the foundational ontology for process instrumentation.

Topology describes actual physical and structural connectivity in the specific plant as encoded in the P&ID.

Graph is the computational representation layer enabling formal query, traversal, and inference.

Semantics enriches raw structural connections with engineering meaning: `SENSES_PROCESS_FROM`, `SIGNALS`, `DISPLAYS_IN`.

Evidence and Provenance records where each piece of knowledge originated and what level of human verification it has received.

MAESTRO's seven-layer architecture addresses all five dimensions directly.

The Preservation Problem: Tacit Knowledge at Risk

Industrial plants accumulate two fundamentally different categories of knowledge. Explicit knowledge is in principle recordable and transferable. Tacit knowledge – the accumulated experience of knowing which valve tends to stick in winter, which sensor reading to distrust after a calibration drift – resides primarily in the minds of experienced engineers.

Industry analyses estimate significant annual knowledge attrition costs in large industrial organisations, with the challenge projected to intensify as the workforce ages [3].

Cervera [10] proposes Expert Mind, a RAG-based architecture for capturing tacit expertise in the energy sector, operating on textual expert testimony. The Ontology, Topology, Graph, and Provenance layers remain outside its scope.

Ma et al. [11] address the reasoning end of the problem with FDRKG-LLM. MAESTRO operates in a complementary space: where FDRKG-LLM takes an existing knowledge graph as given, MAESTRO constructs that knowledge graph from raw engineering documents.

1.3.2 P&ID Digitisation: From Pixels to Semantics

Fletcher and Kasturi [12] established the classical basis for separating text from engineering graphics in mixed documents; modern transformer-based document-understanding architectures have extended this paradigm to multimodal end-to-end processing [29].

Jamieson et al. [13] make the critical observation that engineering drawing digitisation is not reducible to object detection. MAESTRO's own experimentation confirmed this: YOLOv8n achieved 1.00 recall but a 0.19 background false positive rate.

Azangoo et al. [14] at Aalto University demonstrate a semi-automated methodology for generating digital twins from P&ID-derived graph models. MAESTRO diverges fundamentally in purpose: where Azangoo et al. target simulation model generation, MAESTRO targets governed agentic reasoning.

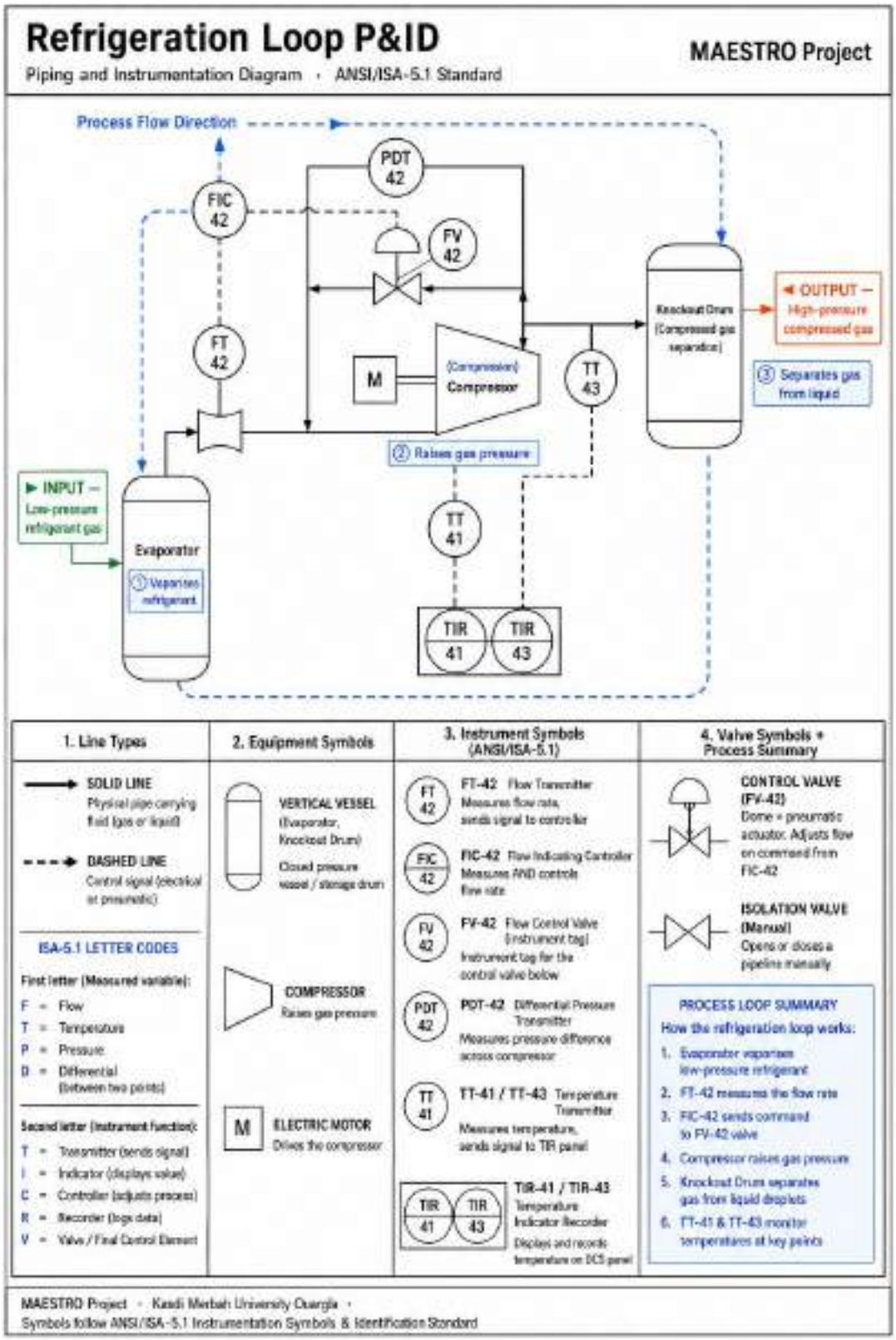


Figure 1.1: Refrigeration loop P&ID with ISA-5.1 reading key. Solid lines: process pipes. Dashed lines: control signals.

Stürmer et al. [15] apply the Relationformer architecture to extract graphs directly from P&ID images. PID2Graph operates on visual and relational learning; MAESTRO encodes ISA-5.1 function code conventions as algorithmic logic.

The scale of data required by vision-based approaches is substantial. Kim et al. [2] constructed a dataset of 75,031 symbols, 10,073 text instances, and 90,054 line annotations to train their detection networks – an effort that presupposes institutional resources unavailable in most industrial settings. Rahul et al. [1] acknowledge explicitly that deep learning P&ID extraction "requires a large amount of data and the user must label it." Even advanced commercial solutions do not exceed 80% asset and connection coverage, and manual digitisation typically requires three to six months per plant document [13]. As Fig. 1.1 illustrates, the visual ambiguity inherent in ISA-5.1 notation – where identical circle symbols represent functionally distinct instruments – makes this data dependency a structural limitation rather than an engineering problem solvable by scale.

1.3.3 Industrial Knowledge Graphs and the Hybrid Knowledge Layer

The DEXPI initiative [16] demonstrates how graph databases can serve as living plant documentation systems [30]. Edge et al. [17] demonstrate that graph-structured retrieval enables more accurate multi-hop synthesis than vector-only retrieval; Pan et al. [31] extend this into a broader roadmap for unifying LLMs and knowledge graphs.

MAESTRO addresses the graph/vector complementarity through the Hybrid Knowledge Layer: `Neo4j` for engineering structure and governance state; `Supabase` for documentary and semantic context. Fig. 1.2 illustrates why both retrieval strategies are necessary and how a single engineer query activates both simultaneously.

1.3.4 Agentic Systems and Retrieval-Augmented Generation

Lewis et al. [18] established the RAG paradigm. Yao et al. [19] extended this with ReAct, interleaving reasoning and action through explicit tool invocation. Benchmarking has demonstrated, however, that LLMs fail systematically at sequential multi-step planning without tool-enforced grounding [32, 33] – a structural limitation that motivates MAESTRO's tool-constrained architecture.

Singh et al. [20] provide the most comprehensive recent survey of Agentic RAG; Wang et al. [34] survey the broader landscape of LLM-based autonomous agent architectures. MAESTRO's Agentic Reasoning Layer implements this pattern through `n8n` orchestration, with agents constrained to query approved plant knowledge.

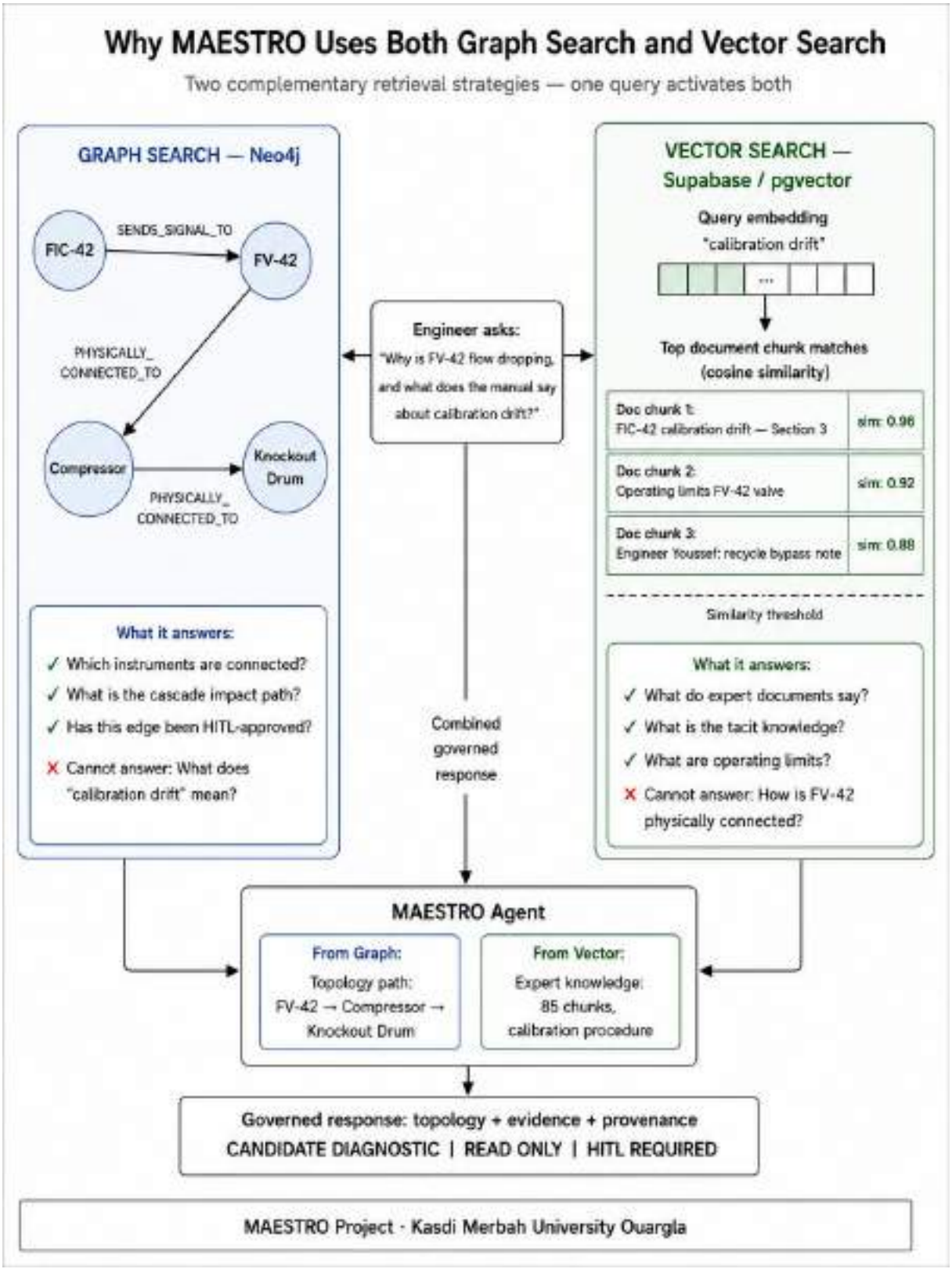


Figure 1.2: Why MAESTRO uses both Graph Search and Vector Search. A single engineer query activates Neo4j (topology, governance state, cascade impact) and Supabase/pgvector (expert documents, tacit knowledge, operating limits) simultaneously. Neither strategy alone can answer both the structural and semantic dimensions of a plant diagnostic query.

1.3.5 Human-in-the-Loop Governance in Industrial AI

Mosqueira-Rey et al. [21] distinguish between annotation-oriented and governance-oriented HITL. In governance-oriented HITL, machine outputs are held in a provisional state until a qualified human confirms them as operationally trustworthy.

IEC 61511-1 [8] mandates human verification for safety-relevant decisions in process automation.

1.3.6 Research Gap: A Field in Productive Fragments

The research gap is structural: the field has developed its components in productive isolation, and no existing work integrates all five layers of industrial knowledge representation into a single pipeline in which each transition between industrial truth states is explicitly controlled and traceable. Within Layer 1 (P&ID digitisation), foundational work by Rahul et al. [1] established the CNN-based extraction pipeline and Kim et al. [2] advanced it to an industrially applicable end-to-end system; both remain confined to the digitisation layer with no claim to agentic reasoning or governance. The knowledge loss risk that motivates the tacit knowledge layer is well-documented in the oil and gas sector [3], confirming that formalisation of expert knowledge is an operational necessity, not a research novelty. Tables 1.1 and 1.2 map these findings systematically.

Table 1.1: Research Gap Mapping – Five-Layer Coverage. — = outside stated scope. Final row: requirements for full industrial deployment.

Work	Onto.	Topo.+Graph	Sem.	Prov.	HITL	Agent	Full Path
Rahul et al. [1]	—	**Part.**	—	—	—	—	—
Kim et al. [2]	—	**Impl.**	—	—	—	—	—
Azangoo [14]	—	**Impl.**	—	—	—	—	—
Stürmer [15]	—	**Part.**	—	—	—	—	—
Jamieson [13]	—	—	—	—	—	—	—
Edge [17]	—	**Impl.**	**Part.**	—	—	**Impl.**	—
Ma [11]	**Part.**	**Impl.**	**Part.**	—	—	**Impl.**	—
Lewis [18]	—	—	—	—	—	**Part.**	—
Singh [20]	—	—	—	—	—	**Impl.**	—
Mosqueira-Rey [21]	—	—	—	**Part.**	**Impl.**	—	—
Cervera [10]	—	—	—	**Part.**	—	**Part.**	—
Abonyi [9]	**Impl.**	**Impl.**	**Impl.**	—	—	—	—
DEXPI [16]	**Part.**	**Impl.**	—	—	—	—	—
Required: full governed pipeline	*Req.*	*Req.*	*Req.*	*Req.*	*Req.*	*Req.*	*Req.*

Onto. = ontology / semantic standard grounding; Topo.+Graph = topology extraction and graph construction; Sem. = semantic enrichment beyond tag recognition; Prov. = formal provenance tracing; HITL = human-in-the-loop governance; Agent = agentic reasoning layer; Full Path = end-to-end governed pipeline from raw P&ID to agent response. — entries reflect scope boundaries of the cited work, not deficiencies.

Table 1.2: Extraction Cost Comparison – Training Requirements and Operational Constraints. N/M = not measurable on the same metric. FPR row: internal controlled comparison on the same P&ID.

Dimension	**Rahul et al. [1]**	**Kim et al. [2]**	**MAESTRO (this work)**
FPR – instrument class [1]	N/M	N/M	**0%**
Training data required [2]	"large amount; user must label it"	**75,031** annotated symbols	**Zero**
GPU required	**Impl.**	**Impl.**	—
Works without annotation	—	—	**Impl.**
P&IDs tested	4	5	1
Scope of output	Digitisation only	Digitisation only	Extraction + governance + reasoning

Notes. (1) MAESTRO FPR = 0%: see Table 3.4. The comparison is controlled: YOLOv8n achieves FPR 0.19 on the instrument class, while the ISA-5.1 OCR-led strategy achieves FPR 0.00 on the same P&ID – see Fig. 2.4 and Fig. 2.6. (2) Rahul et al. quote from §1.3.2: *"requires a large amount of data and the user must label it"* [1]. Kim et al. dataset from §1.3.2: *"75,031 symbols, 10,073 text instances, and 90,054 line annotations"* [2]. MAESTRO zero training: see Fig. 2.4 caption – *"no training data required"*.

1.4 Motivation and Intellectual Origin

The intellectual trajectory of this project did not begin with a clean theoretical framework. It began with a concrete observation about Maintenance 4.0: why do predictive maintenance systems so frequently underperform? The answer, in most cases, is not a failure of the prediction algorithm but of the knowledge substrate on which the algorithm depends.

Early experimentation with YOLOv8-based symbol detection demonstrated that visual priors alone are insufficient for reliable instrument identification. The instrument class showed a perfect recall score but an unacceptably high false positive rate (0.19), reflecting the visual ambiguity inherent in ISA-5.1: the same circular symbol is used for multiple instrument types, differentiated only through the text tag. This practical failure led to the architecture's defining decision: a transition toward an OCR-led, geometry-aware, ISA-5.1-informed interpretation strategy.

1.5 A Central Conceptual Contribution: Confidence Is Not Trust

The distinction at the centre of the architecture is the difference between confidence and trust. Confidence refers to the probabilistic output of perception and semantic enrichment processes. Trust – in the sense relevant to operational industrial reasoning – is established only through explicit human governance: the HITL sublayer in which a qualified engineer reviews machine-generated draft elements and explicitly approves their promotion to verified plant knowledge.

This three-level distinction – perception confidence, semantic confidence, and validation status – is formalised throughout the MAESTRO architecture. The Agentic Reasoning Layer reports and exposes these states but does not generate trust autonomously.

1.6 Research Problem Statement

Industrial AI agents cannot be trusted for plant management support unless they are grounded in verified, governed, plant-specific knowledge. The challenge is to transform raw P&ID engineering evidence into structured, governed, and agent-consumable industrial knowledge through a coherent end-to-end architecture – where each transition between knowledge states is explicitly controlled and traceable.

1.7 Research Hypothesis

Agentic reasoning in smart factory management becomes reliably grounded when industrial agents operate over a knowledge substrate constructed through an end-to-end governed pathway – one that begins from P&ID engineering evidence, passes through OCR-led interpretation, ISA-5.1-informed semantic enrichment, and explicit human governance, and terminates in a hybrid knowledge layer supporting both graph-structured topological queries and documentary retrieval.

1.8 Research Objectives

The general objective is to design, implement, and validate a Multi-Agent System for Smart Factory Management enabling plant-aware agentic reasoning by transforming P&ID engineering evidence into structured, governed industrial knowledge. The specific objectives are:

O1. Extract industrial entities and instrument tags through an OCR-led, ISA-5.1-informed pipeline.

O2. Construct plant topology using engineering-constrained traversal with a passthrough guard.

O3. Apply adaptive scale calibration using a symbol-geometry-derived base unit.

O4. Enrich structural graph relations with ISA-5.1-informed semantic meaning.

O5. Organise the enriched plant representation in a governed `Neo4j` knowledge graph with explicit draft and approved states.

O6. Support Agentic RAG through a hybrid knowledge layer combining `Neo4j` and `Supabase`.

O7. Equip reasoning agents with five plant-aware tools for graph querying, evidence retrieval, existence verification, and live telemetry access.

O8. Implement the HITL governance sublayer as a controlled validation boundary aligned with IEC 61511-1 [8].

O9. Validate the full governed pathway through a deep case study at TRL 4–5 scope.

1.9 Research Questions

Main RQ: How can a Multi-Agent System for Smart Factory Management achieve reliable agentic reasoning by grounding industrial agents in governed, plant-aware knowledge extracted from P&ID engineering documents?

RQ1: How can raw P&ID diagrams be transformed into industrial entities and topological structures through an OCR-led, ISA-5.1-informed interpretation pipeline?

RQ2: How does semantic enrichment, grounded in ISA-5.1 conventions, improve the quality of industrial reasoning beyond raw structural extraction?

RQ3: How does a hybrid knowledge layer combining graph-structured data with documentary retrieval strengthen environment-aware Agentic RAG?

RQ4: How can the HITL governance sublayer function as a controlled validation boundary rather than a post-hoc annotation interface?

RQ5: How can the governed pathway be organised as a coherent, end-to-end architecture in which each state transition is explicitly controlled?

RQ6: To what extent does the current implementation demonstrate feasibility for plant-aware agentic reasoning within the validated scope?

1.10 Main Contributions

C1. A governed pathway closure – end-to-end bridge from silent P&IDs to governed industrial AI reasoning.

C2. A seven-layer governed architecture addressing each layer's distinct technical challenge.

C3. An OCR-led industrial interpretation pipeline clarifying OCR's dual role in P&ID processing.

C4. Operationalisation of the five-layer industrial knowledge framework [9] across all seven architectural layers of MAESTRO, instantiating each layer with governed, traceable implementation artefacts.

C5. An explicit confidence-versus-trust architecture formalising the distinction between probabilistic and verified knowledge.

C6. An architecturally grounded epistemic restraint mechanism that structurally constrains hallucination propagation through governance-separated boundaries.

C7. A hybrid knowledge layer combining `Neo4j` topological truth and `Supabase` semantic context.

C8. A validated tacit knowledge preservation and retrieval pipeline bridging expert experience with machine-accessible reasoning.

1.11 Scope and Limitations

The validation is grounded in a deep case study of a single validated P&ID – a refrigeration loop comprising a Compressor, Knockout Drum, Evaporator, and eight instrumentation elements (FIC-42, FT-42, PDT-42, FV-42, TT-41, TT-43, TIR-41, TIR-43) per ANSI/ISA-5.1 conventions, yielding 12 validated candidates in total (8 instrument nodes and 4 equipment nodes). The scope corresponds to Technology Readiness Level 4–5 as defined by the European Commission's Horizon Europe framework (TRL 4: validated in laboratory environment; TRL 5: validated in relevant environment).

1.12 Organisation of the Thesis

This thesis is organised in four chapters. The present chapter establishes the industrial problem, surveys the state of the art across four technical domains, and states the research questions, hypothesis, and contributions. Chapter 2 details the MAESTRO architecture and its governance-by-design framework: the runtime topology, the structural safety contracts, the seven architectural layers, and the mathematical formalisation. Chapter 3 reports the implementation and experimental validation, layer by layer, including the containerised testbed, the diagnostic scenarios, latency benchmarks, the threat model, and the ablation analysis. Chapter 4 critically discusses the findings, assesses the Technology Readiness Level, and concludes with future directions.

Chapter 2

The MAESTRO Architecture and Governance Framework

2.1 Conceptual Overview: Governance by Design

MAESTRO is a governed multi-agent system that transforms raw P&ID drawings and industrial documents into structured, audit-traceable plant intelligence. Its defining philosophy is *governance by design*: every machine-generated element remains a non-authoritative candidate until a qualified engineer promotes it through an explicit human decision. This principle is not enforced by instructing the reasoning model to behave cautiously; it is enforced structurally, in the architecture itself, so that no prompt, instruction, or adversarial input can bypass it.

Fig. 2.1 presents the runtime topology of the system as actually deployed. The engineer interacts only through a browser client over HTTPS; the Next.js platform acts as a same-origin proxy so that the browser never contacts the backend services directly. The core services run as containerised components on an on-premise Docker network: the Vision API (perception and extraction), the n8n orchestrator (agentic reasoning over five read-only tools), the Neo4j governed graph layer, and the Supabase pgvector semantic-evidence layer. External model APIs are reached only as outbound calls that cross the container boundary. A single governed write gate (`approve-edge`) is the only path by which any candidate becomes approved plant truth, and it requires a human-verified identity.

The architecture presented in this chapter was not designed upfront and then implemented. It was discovered through a sequence of principled failures: each failure revealed the correct engineering principle; each principle forced a structural decision that determined the next layer's form. The scope of validation – a single, deeply analysed case study – is deliberate, not a concession.

2.2 Architectural Governance and Safety Contracts

2.2.1 The Vulnerability of Prompt-Based Guardrails

Many AI systems attempt to constrain agent behaviour through the system prompt: the model is instructed, in natural language, to refuse dangerous operations. This approach is fragile. A natural-language instruction can be overridden by prompt injection (an input that says, in effect, "ignore your previous instructions"), circumvented by jailbreak phrasing, or simply misinterpreted by the model under distribution shift. A guardrail that depends on the model choosing to obey is not a guarantee; it is a request.

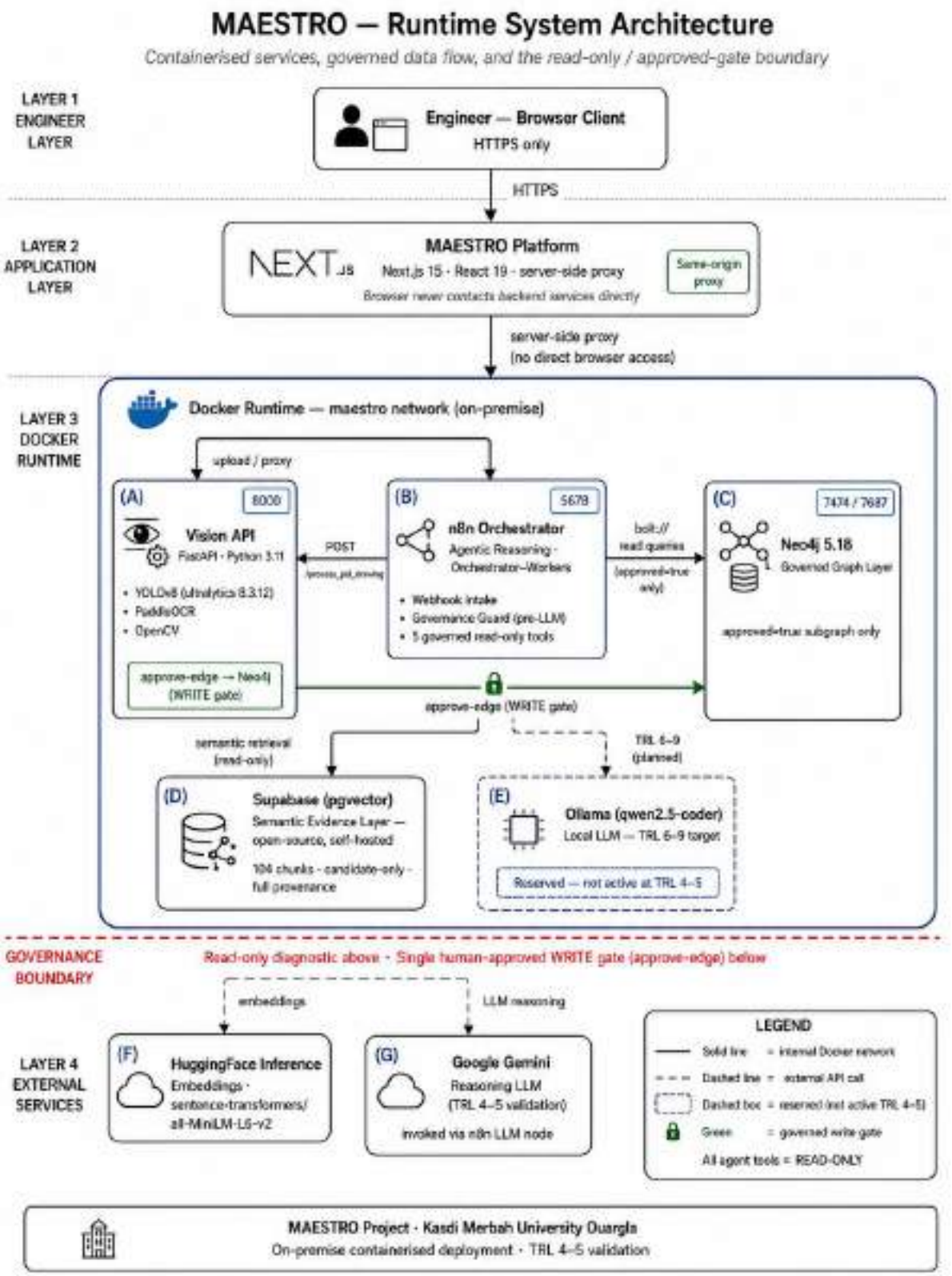


Figure 2.1: MAESTRO runtime architecture: four layers separated by a single governed write gate. All agent tools are read-only; `approve-edge` is the only write path, requiring human-verified identity.

2.2.2 Structural Tool Contracts: Read-Only by Construction

MAESTRO does not rely on the prompt for governance. The reasoning agent is exposed to exactly five tools, every one of which is read-only: evidence retrieval, entity-existence verification, topology context, impact hypothesis, and live telemetry. No tool in the contract can write, approve, promote, or issue a control action – such an operation is not absent by instruction, it is absent by construction. Even if every line of the system prompt were deleted, the agent still could not mutate plant state, because the capability does not exist in its tool set. Each candidate object additionally carries governance flags enforced in the data layer itself (`read_only`, `candidate_only`, `approved=false`, `graph_truth=false`), independent of any model behaviour.

MAESTRO does also include a system-level instruction declaring the agent read-only – it is visible in the diagnostic composer node of the n8n workflow. This instruction is deliberately retained, but it is not the governance mechanism: it is a redundant surface layer that merely restates, in natural language, a constraint already enforced architecturally. Governance is distributed structurally across the seven layers; the prompt only echoes it. This is defence in depth, not dependence on the prompt.

2.2.3 The Single Governed Write Gate

The only mutation path in the entire system is the `approve-edge` gate in the Vision API, which writes an `APPROVED_TOPOLOGY_EDGE` into Neo4j with full audit provenance. The actor identity for this write is derived exclusively from a verified JWT and is never accepted from the request body, so it cannot be injected through a prompt or a crafted payload. Until a human engineer passes through this gate, every candidate remains advisory.

2.3 Architecture of Record and Infrastructure

MAESTRO runs as a set of containerised services orchestrated via Docker Compose. Local deployment avoids cloud handling of sensitive plant diagrams.

2.3.1 Vision API

The Vision API is a FastAPI application responsible for all P&ID perception and extraction. Three components operate in sequence: YOLOv8n provides spatial priors for equipment-scale symbols; PaddleOCR (Docker canonical OCR runtime) recovers all text regions; OpenCV handles geometry – contour detection, skeleton traversal, bounding box operations. The combination is not a parallel stack – it is a pipeline where OCR evidence governs entity determination and geometry governs connectivity.

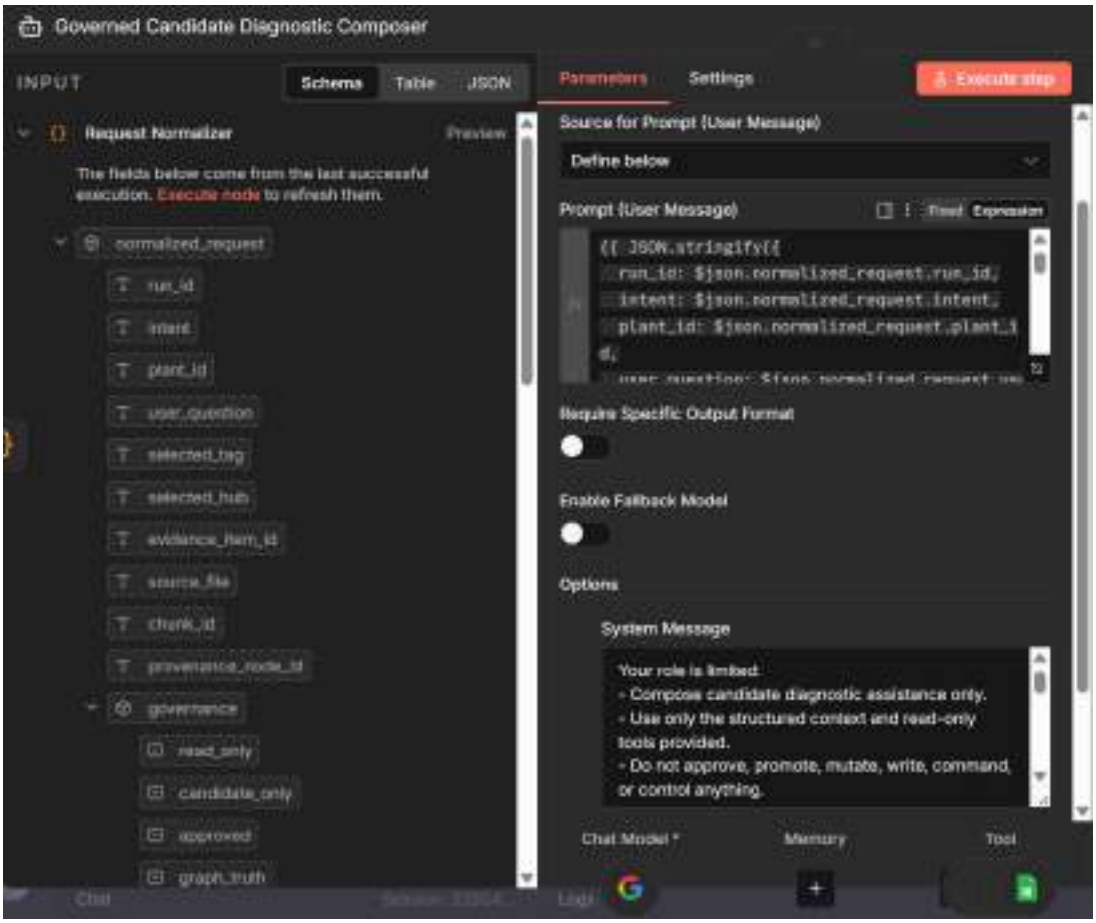


Figure 2.2: The n8n diagnostic composer node. Governance flags (`read_only`, `candidate_only`, `approved`, `graph_truth`) are enforced in the data structure before the model is invoked. The system-message prompt only restates a constraint the architecture already guarantees.

The Approved Gate (`approve-edge`) writes `APPROVED_TOPOLOGY_EDGE` relationships to `Neo4j`. Actor identity is derived from JWT via the Platform proxy and is never accepted from the request body.

2.3.2 Platform

The MAESTRO Platform is a Next.js App Router application in TypeScript. The frontend stack comprises React Flow with ELK.js 0.11.1 for graph visualisation; Zustand for state management; and a server-side Next.js proxy – the browser never contacts the Vision API directly and the `n8n` webhook URL is never exposed.

2.3.3 Knowledge Layer Infrastructure

`Neo4j` (driver `neo4j>=5.18,<6.0`) serves as the Governed Graph Layer. `Supabase` with `pgvector` serves as the documentary retrieval layer. HuggingFace embeddings (`sentence`

-transformers/all-MiniLM-L6-v2) are computed within the n8n workflow. The browser never writes directly to Supabase.

n8n orchestrates the Agentic Reasoning Layer following the Orchestrator-Workers pattern [22]. For TRL 4–5 validation, the reasoning model is Google Gemini 2.5 Flash – a deliberate choice to reduce local computational overhead. This substitution does not alter any governance property; all governance decisions (blocking, tool filtering, approved-state enforcement) execute before the LLM is invoked. The architecture is model-agnostic by design. For TRL 6–9 deployment, the architecture transitions to Ollama-hosted qwen2.5-coder, which restores full local execution and eliminates the cloud data-handling tension inherent in the TRL 4–5 prototype. This tension – that a system designed for local data sovereignty uses a cloud-hosted model during validation – is acknowledged as a prototype limitation: the validation P&ID is not confidential plant data, and no production deployment should use a cloud LLM without explicit data governance controls.

2.4 The Governed Methodological Pathway

Table 2.1: Eight-Stage Governed Pathway from P&ID to Agent Response

Stage	Node	Description
1	P&ID Ingestion	Engineering evidence enters the Vision Layer as a raw image
2	Draft Extraction	OCR-led, ISA-5.1-informed interpretation produces a CandidateEnvelope
3	Graph Construction	Candidates converted to DraftNode and DraftEdge in Neo4j
4	Semantic Enrichment	ISA-5.1 role interpretation assigns engineering meaning
5	HITL Verification	Engineer reviews, corrects, approves or rejects via Draft Canvas
6	Approved Gate	approve-edge writes APPROVED_TOPOLOGY_EDGE
7	Parameterisation	Engineering units and operational limits assigned
8	Agentic Reasoning	n8n-based layer queries the Hybrid Knowledge Layer via governed tools

Governing constraint: approved-state knowledge is unreachable from draft-state without an explicit human decision.

2.5 The Seven Architectural Layers

Fig. 2.3 presents the complete MAESTRO architecture as a governed nine-stage pathway from raw engineering evidence to industrial decision support. The seven core layers (Vision through Agentic Reasoning) are flanked by the engineering evidence input and the

industrial decision support output, making explicit that the system's purpose is to connect plant documentation to actionable, governed intelligence.

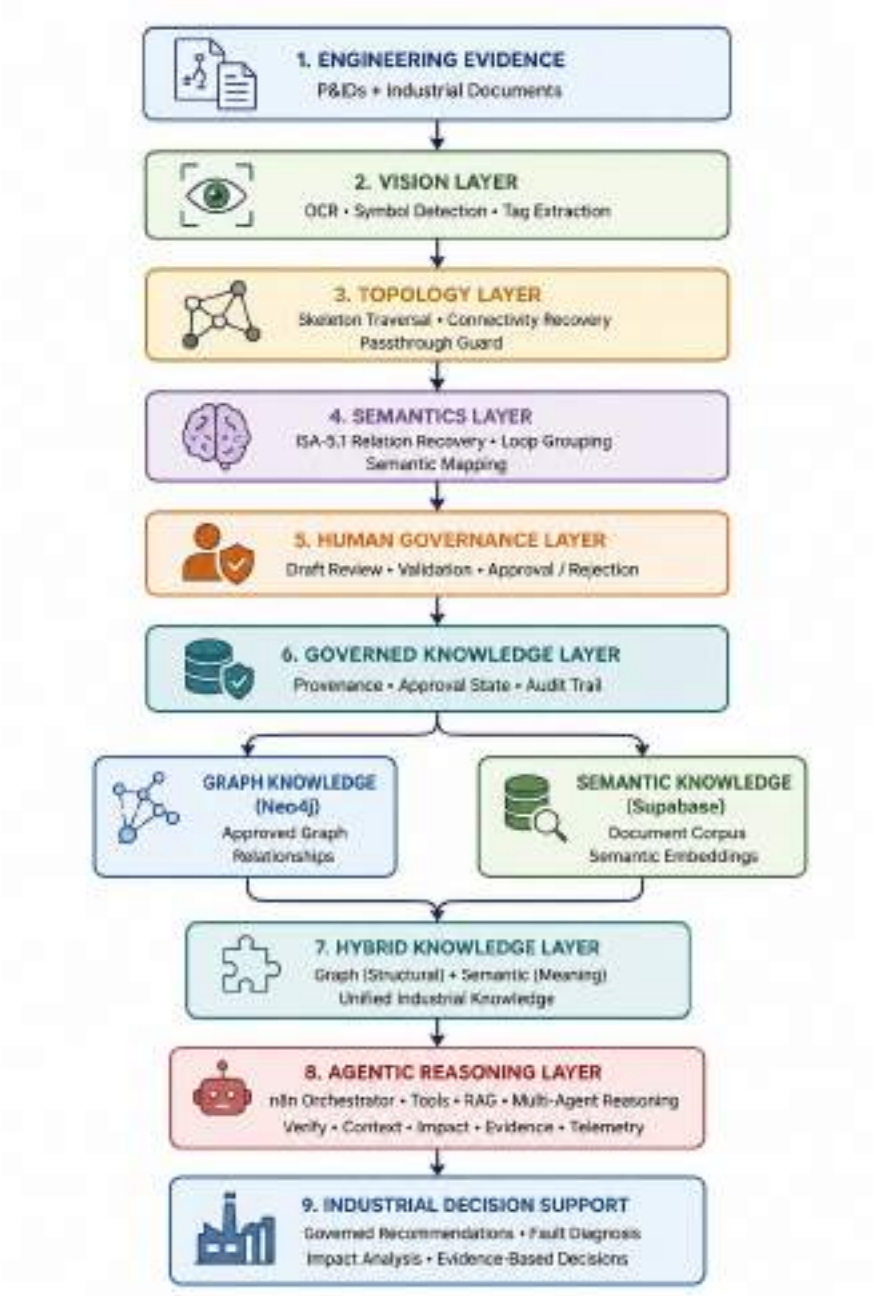


Figure 2.3: MAESTRO governed architecture: nine-stage pathway from engineering evidence to industrial decision support, across seven core layers.

For clarity, the seven layers can be read as three logical groups. The **ingestion and semantics group** (Layers 1–3: Vision, Topology, and Calibration) turns a raw P&ID image into geometrically calibrated, ISA-5.1-grounded candidate entities. The **knowledge core group** (Layers 4–5: Semantics and the Governed Graph Layer) structures those candidates into the industrial knowledge graph and enforces the human-governed

promotion mechanism. The **reasoning and inference group** (Layers 6–7: Agentic Reasoning and the Human-in-the-Loop Governance Sublayer) performs governed diagnostic reasoning over the approved knowledge through multi-source retrieval. The subsections below detail each layer in turn.

2.5.1 Vision Layer: OCR-Led, Geometry-Aware, ISA-5.1-Informed

The ANSI/ISA-5.1 standard uses the same circular symbol for multiple instrument types. A Flow Transmitter (FT-42) and a Flow Indicating Controller (FIC-42) are visually identical. Their difference is semantic, encoded in the tag, not in the shape. YOLO achieved 1.00 recall on the instrument class but a 0.19 background false positive rate – making vision-only extraction structurally unreliable for the most critical entity class.

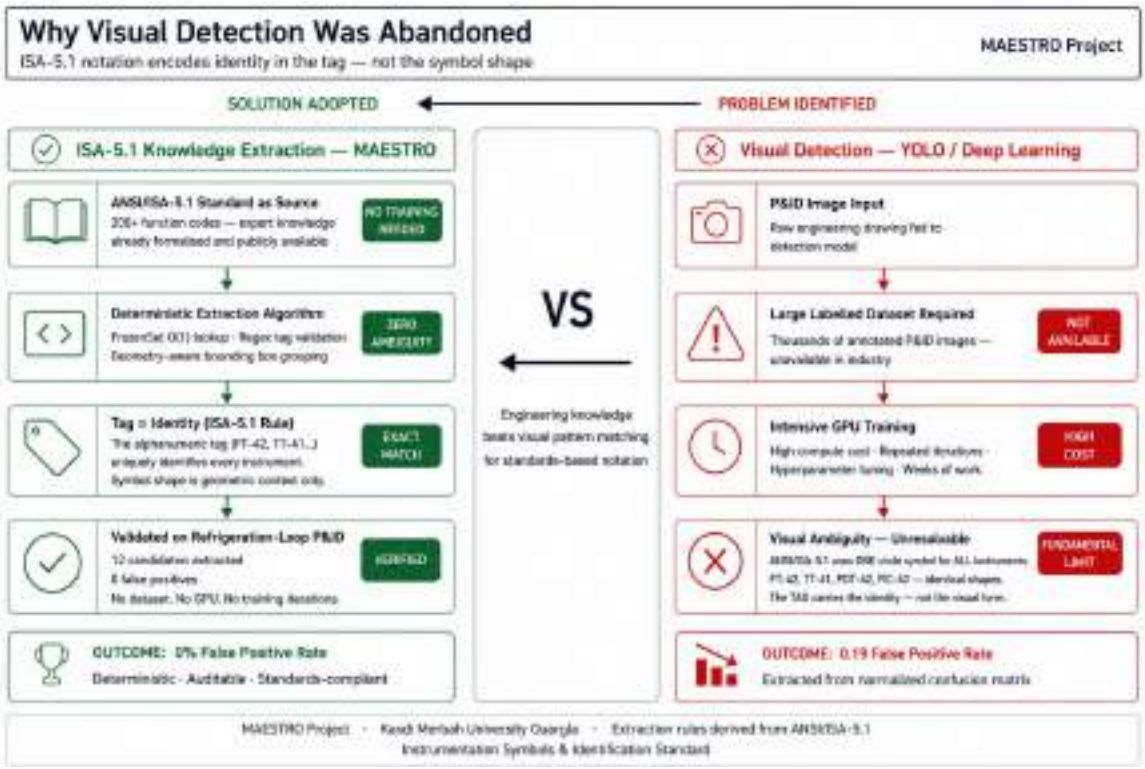


Figure 2.4: YOLO versus ISA-5.1 rule-based extraction. Identical circle symbols for all instrument types make vision-only detection structurally unreliable.

As established in §1.3.2 (see Table 1.1), even state-of-the-art approaches require datasets exceeding 75,000 annotated symbols [2] – resources unavailable in industrial practice.

These false positives are attributable to two compounding factors: a *class distribution skew* (imbalanced, long-tailed training data) inherent to industrial P&ID symbol sets – where instruments and valves dominate while specialised components are structurally rare

– and the inherent visual ambiguity of ISA-5.1 notation, where the same circular symbol represents multiple instrument types, differentiated only by the alphanumeric tag. No visual prior can resolve this ambiguity. This practical failure motivated a fundamental architectural shift: rather than attempting to improve the detector, the engineering knowledge encoded in the ISA-5.1 standard itself was converted into deterministic extraction algorithms. The tag becomes the identity; the symbol becomes geometric context only.

The architectural response: YOLOv8n was abandoned as the primary identification mechanism for three compounding reasons: (1) its 0.19 false positive rate on instrument symbols, (2) the requirement for a labelled training dataset of industrial P&IDs which was unavailable, and (3) the visual indistinguishability of ISA-5.1 instrument types at the symbol level – a problem that tag-based identification resolves definitively. OCR became the primary instrument identification mechanism. Each recovered string is checked against an ISA-5.1 FrozenSet of approximately 200 standard function codes – an $O(1)$ lookup that rejects non-instrument strings that a regex (`^[A-Z]{2,4}$`) would accept (VENT, PUMP, COOL, MAIN). When a recovered ISA tag sits geometrically within or adjacent to a circle contour, the tag determines the entity type.

Three further processing steps refine the output:

1. **Three-pass deduplication**: IoU ≥ 0.85 (bounding box), tag-based (shared tag), IoS containment (≥ 0.5).

2. **SharedDisplay channel expansion**: TIR-41 and TIR-43 emitted as separate candidates from the DCS panel rectangle.

3. **Slice B suppression**: vessel-to-vessel geometric noise and untagged candidates removed.

Validated output: 12 clean candidates, 14 valid ISA-grounded relations, 0 false positives.

2.5.2 Topology Layer

The Topology Layer transforms interpreted entities into plant structure using BFS-based skeleton traversal. The **passthrough guard** encodes the physical drafting principle that process lines do not pass through unrelated equipment bodies as a BFS stop condition. Weakly anchored elements are flagged with an `isolated_reason` property rather than silently promoted.

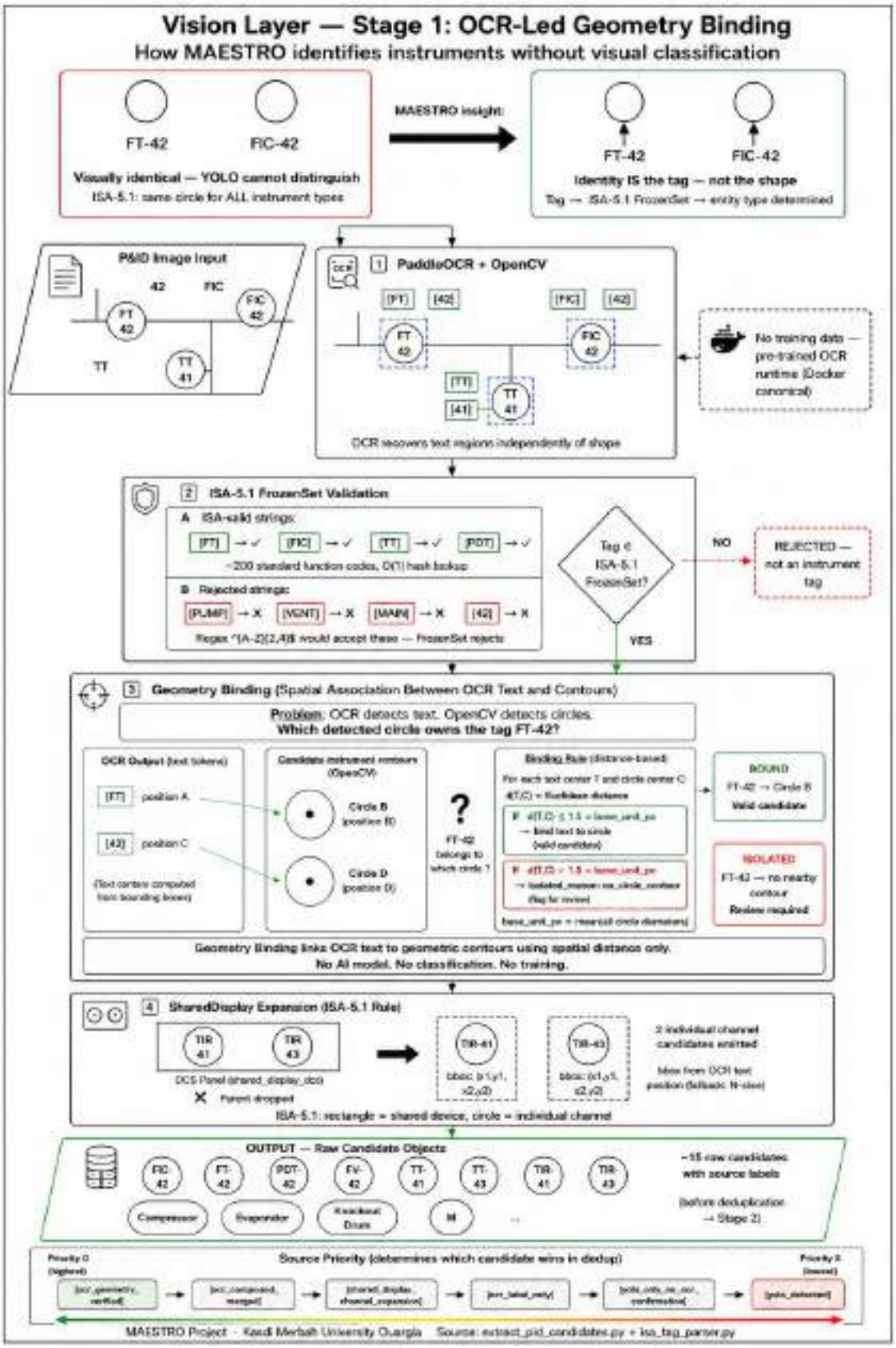


Figure 2.5: Vision Layer Stage 1: OCR-led geometry binding pipeline. From raw P&ID image to 15 labelled candidate objects via ISA-5.1 FrozenSet validation, distance-based circle contour binding, and SharedDisplay channel expansion.

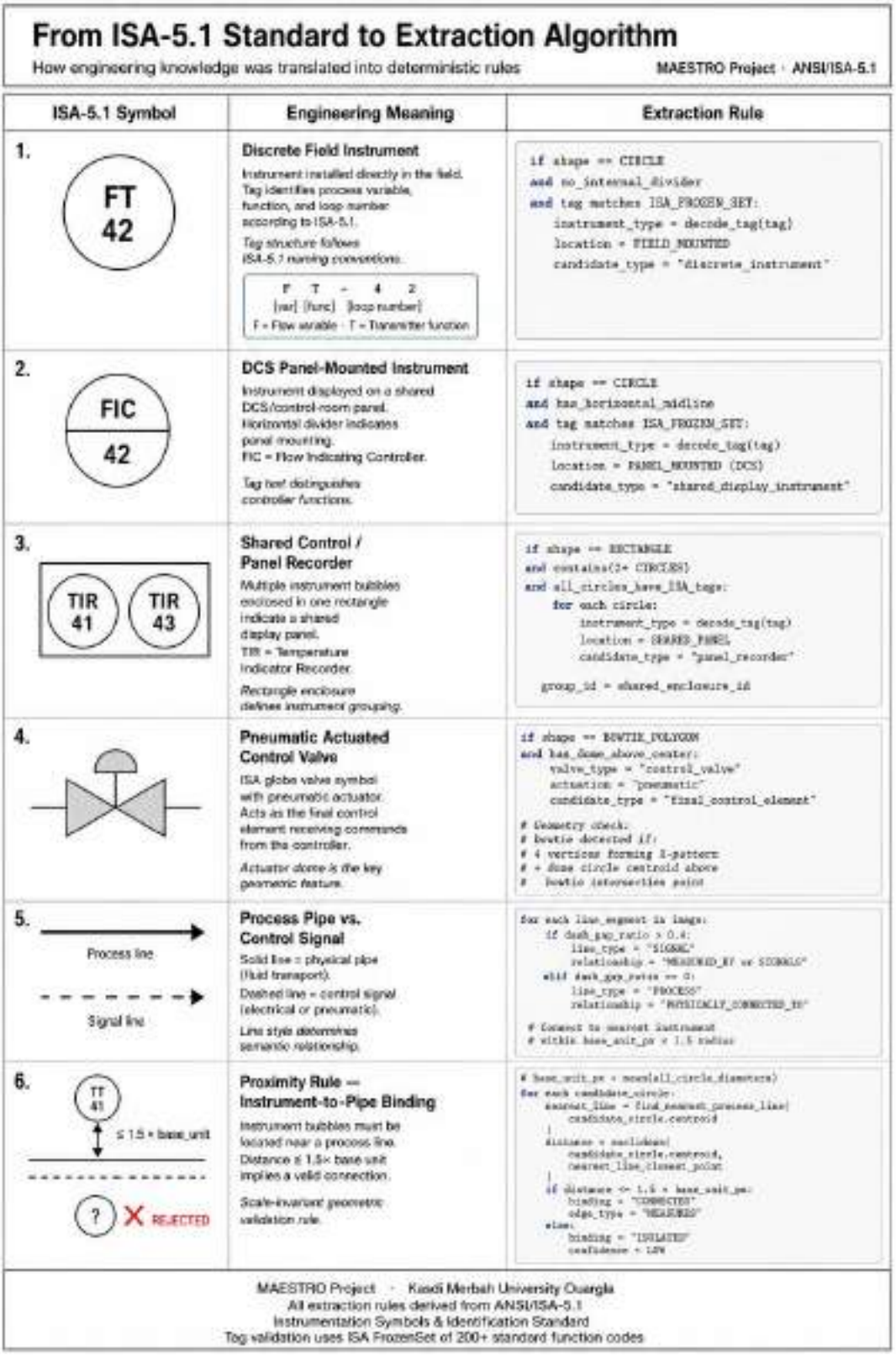


Figure 2.6: ANSI/ISA-5.1 engineering knowledge translated into deterministic extraction rules – no training data required.

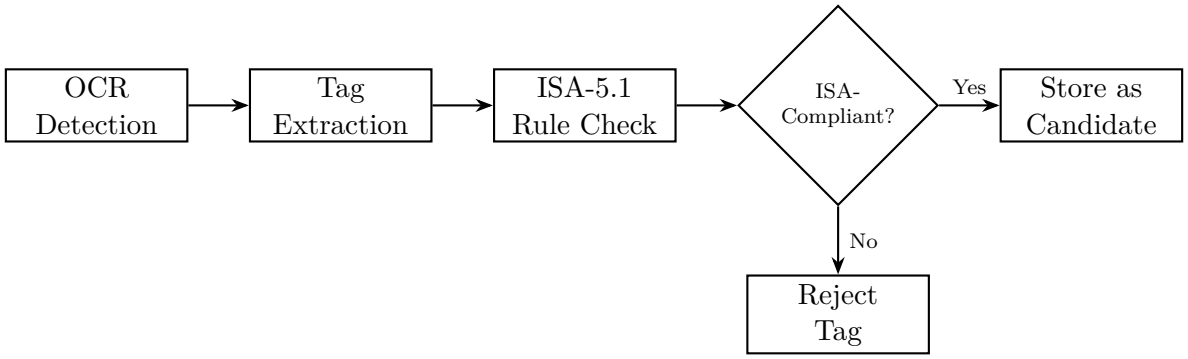


Figure 2.7: ISA-5.1 tag validation workflow. Non-compliant tags are rejected at extraction time.

2.5.3 Calibration Layer

Adaptive calibration derives a `base_unit_px` from symbol geometry, allowing proximity thresholds to be expressed as ratios rather than fixed pixel values.

2.5.4 Semantics Layer

The Semantics Layer enriches the structural graph with ISA-5.1-informed engineering meaning. The ISA-5.1 Relation Recovery Layer uses a loop grouper (shared loop number) and cross-loop sensor-to-controller matcher to recover connections invisible to geometric proximity.

2.5.5 Governed Graph Layer and the Hybrid Knowledge Layer

The critical design decision: draft and approved states are first-class graph properties on every node and relationship, not application-layer flags. Industry-scale knowledge graph deployments consistently identify the challenge of maintaining data quality and provenance at the graph layer as a core operational concern [28]. An entity with `approved=false` or `approved=null` cannot reach the agent's tool contracts in practice. The tool contracts issue Cypher queries of the form `WHERE n.approved = true`; in Cypher, comparisons against `false` or `null` both evaluate to non-true and are excluded from results. This is a query-level filter rather than a schema-level constraint: a future migration step to add a Neo4j property existence constraint would convert this into a structural guarantee. Within the validated scope, the behavioural enforcement was confirmed by direct graph inspection – the `CONTROLS` legacy edge (FIC-42 → FV-42, `approved: null`) was not returned by any governed tool call during the six validated scenarios (see Section 3.10.1 for the full analysis).

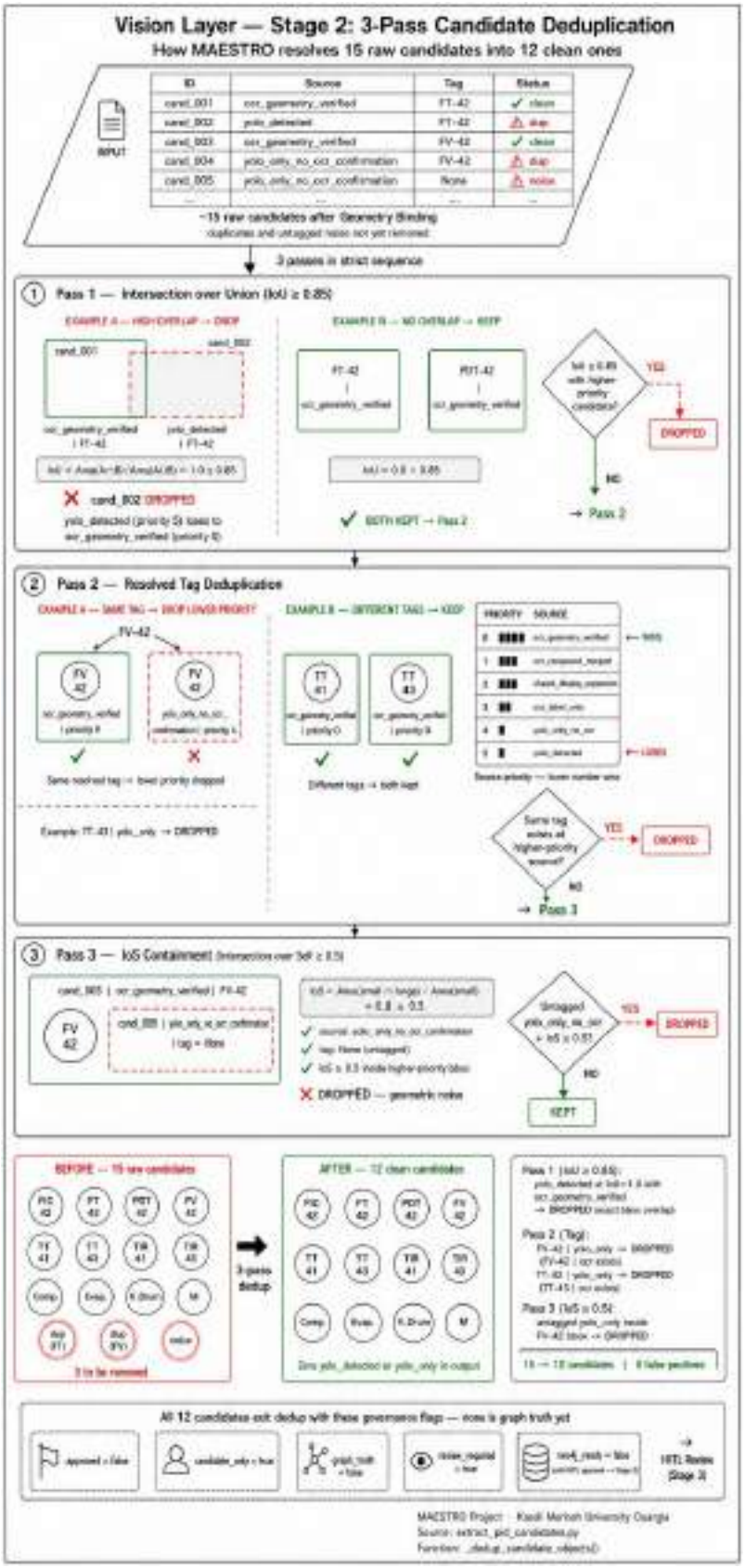


Figure 2.8: Vision Layer Stage 2: 3-pass candidate deduplication. IoU, tag-identity, and IoS containment passes reduce 15 raw candidates to 12 clean ones; all output candidates carry `approved=false` and `review_required=true`.

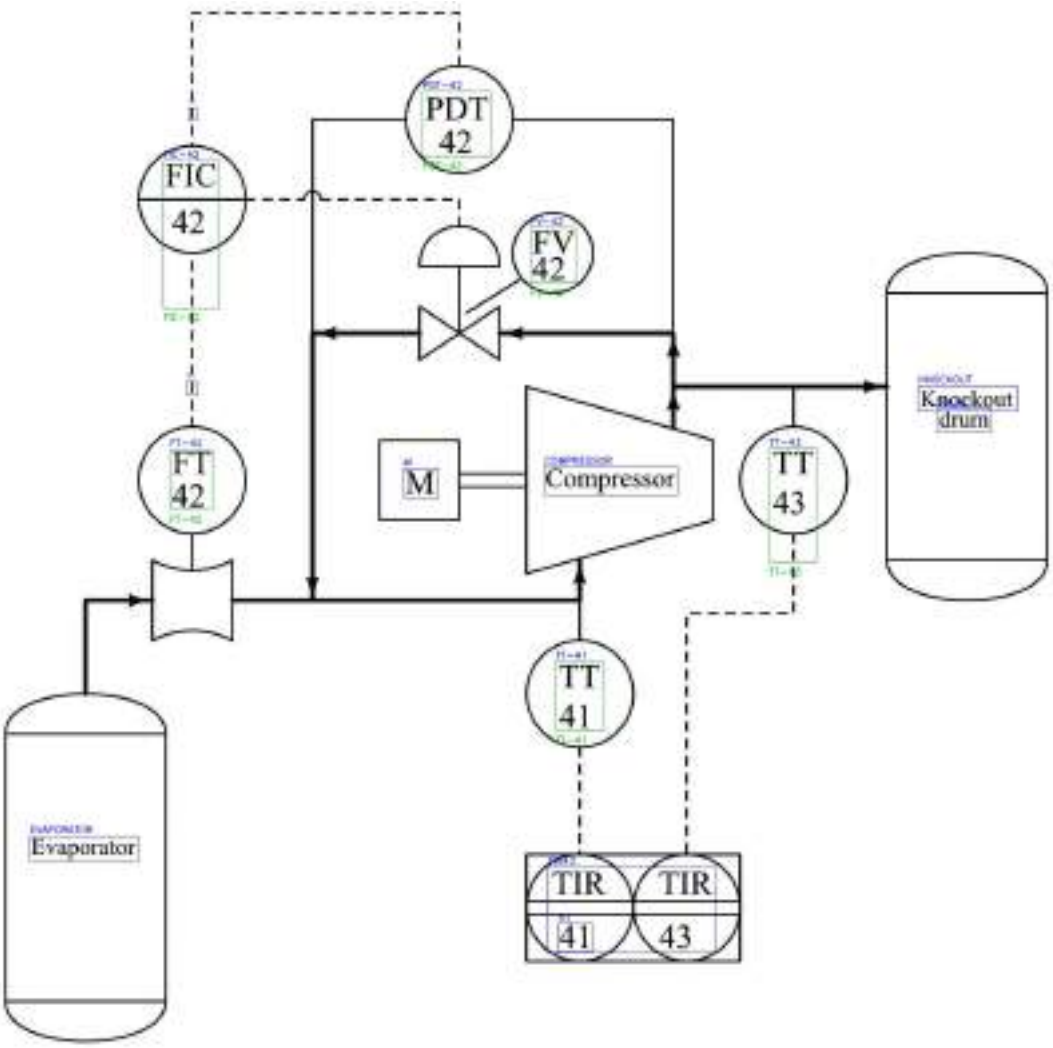


Figure 2.9: OCR grouping overlay on the validated P&ID. Green bounding boxes show ISA-5.1-tagged candidates; all 12 extracted with zero false positives after three-pass deduplication.

Governance note: MAESTRO employs two distinct mechanisms. Query-layer enforcement (`WHERE n.approved = true`) excludes non-approved entities behaviourally. Tool-contract enforcement structurally removes all write and control-action interfaces – the TT-41 interception is deterministic and architecture-level. "Governance is architectural" refers specifically to the latter.

2.5.6 Agentic Reasoning Layer

The MAESTRO agentic reasoning loop is illustrated in Fig. 2.11. The agent is defined as Model + Tools + Knowledge: each query passes through the Governance Guard before any LLM or tool is invoked, ensuring that control-action requests are intercepted architecturally rather than by prompt instruction.

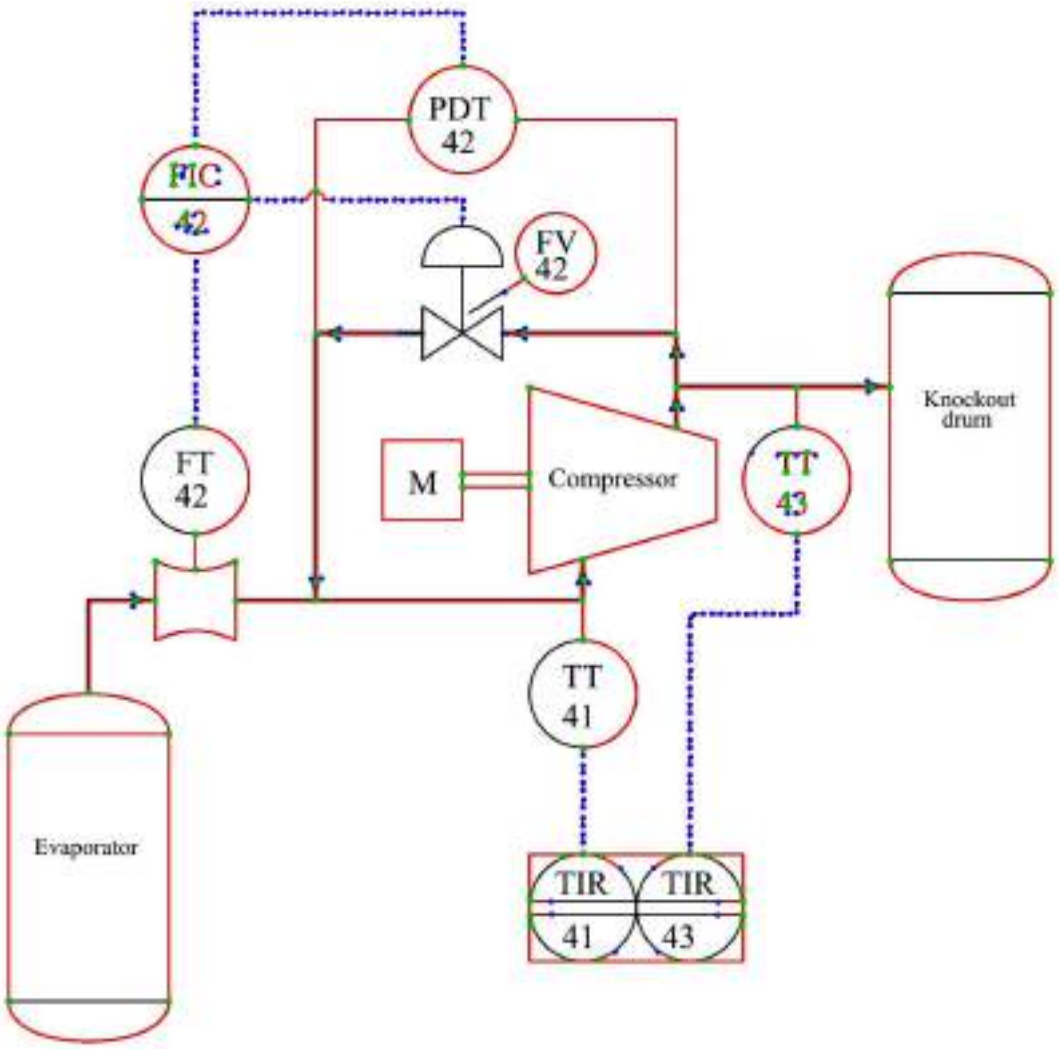


Figure 2.10: Structural topology overlay on the validated P&ID. Red lines: process flows; blue dotted lines: signal paths.

The agent is equipped with **five** governed tools: (1) *Supabase Evidence Reader* – semantic similarity retrieval via `pgvector` over the industrial knowledge corpus; (2) *Verify Candidate Entity Exists* – `Neo4j` YES/NO existence gate; (3) *Get Candidate Topology Context* – `PHYSICALLY_CONNECTED_TO` ±1-hop traversal; (4) *Get Candidate Impact Hypothesis* – multi-hop traversal returning ordered impact propagation across ISA-5.1-enriched relationships; (5) *Get Live Telemetry* – time-series sensor readings filtered by `Tag_ID` and `Limit` for threshold comparison.

The architecturally grounded epistemic restraint behaviour is structural: when FIC-42's connections are not evidenced in the approved graph, the agent reports this explicitly rather than filling the gap with inference.

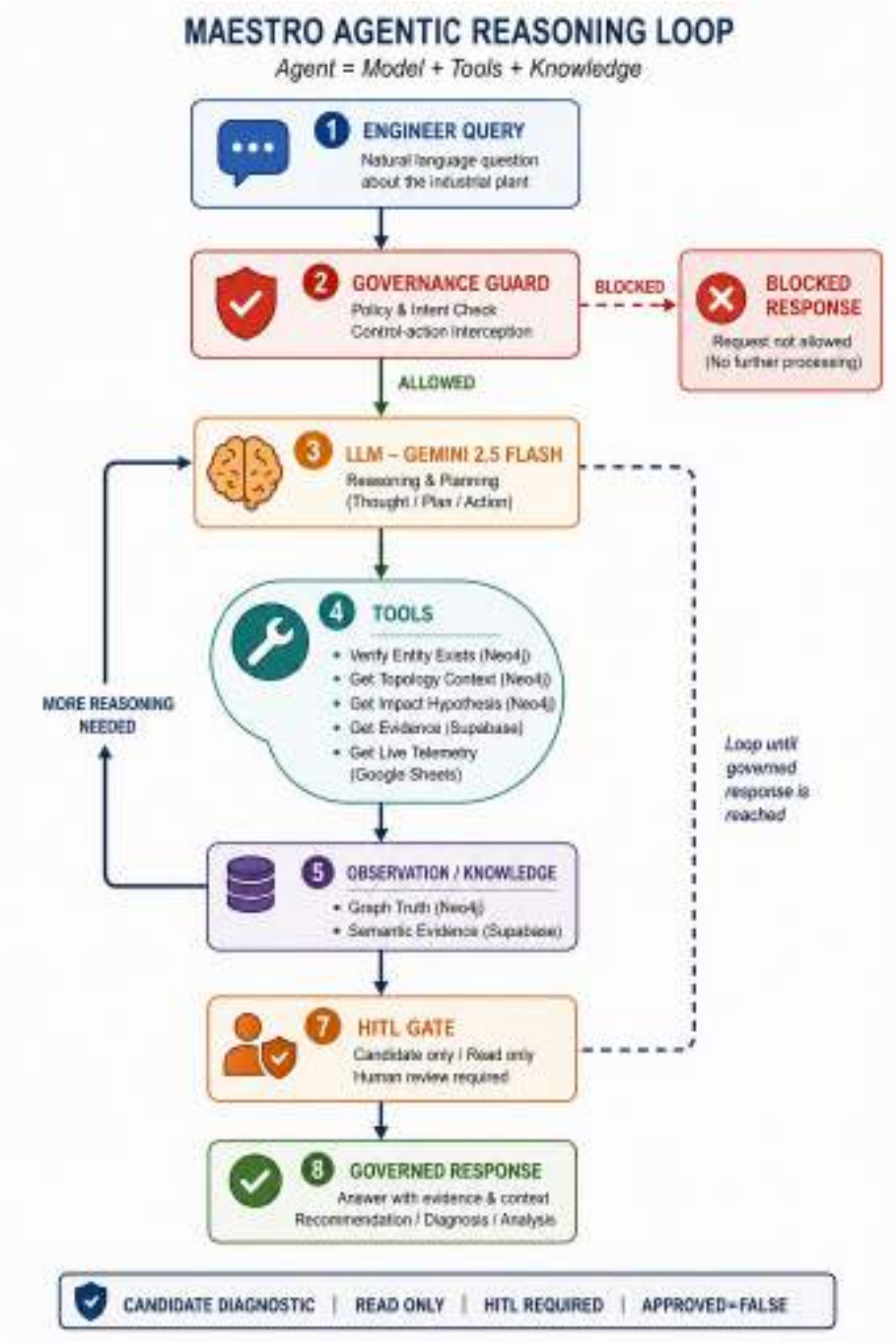


Figure 2.11: MAESTRO Agentic Reasoning Loop (Agent = Model + Tools + Knowledge). Control-action requests are intercepted before LLM invocation; allowed queries traverse five governed tools.



Figure 2.12: The five governed tool nodes in the n8n workflow. All contracts are read-only; no write or control-action interface is exposed to the agent.

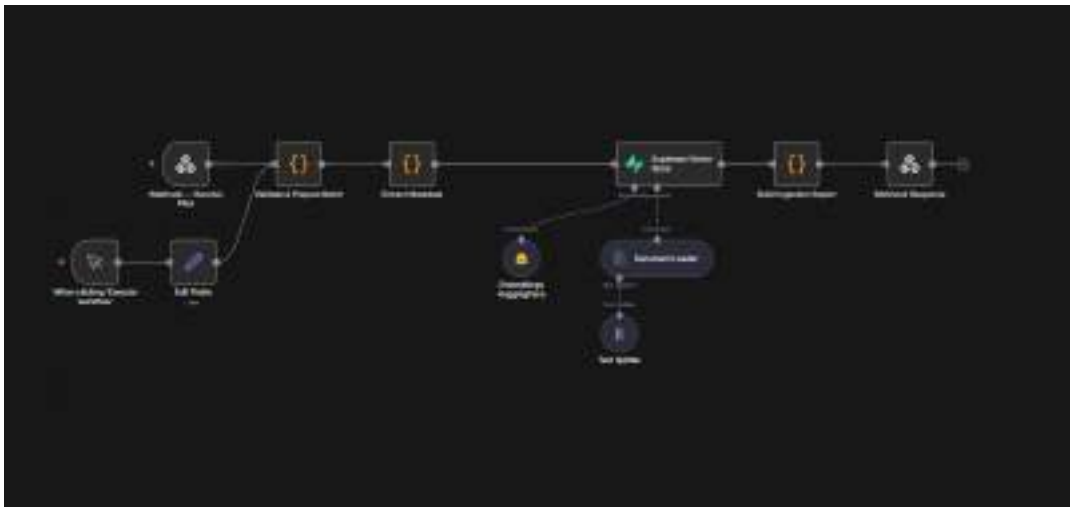


Figure 2.13: VectorStore ingestion workflow. Documents are chunked, embedded via HuggingFace, and stored in Supabase with full provenance metadata. Validated scenarios use the 104-chunk corpus (13 sources); 737-chunk total includes all development batches.

2.5.7 Human-in-the-Loop Governance Sublayer

The HITL sublayer includes the Draft Canvas, Extraction Evidence Panel, Linework Authoring Tools (Slices 1–4), and the Approved Gate (R6). The validated R6 execution – following forensic migration that resolved 54,826 duplicate records to **8 canonical relationships** via MERGE key refactoring on `(source.tag, target.tag, relation_type)` – confirmed **8 APPROVED_TOPOLOGY_EDGE** relationships with full audit provenance and 8 Audit Ledger events.

2.5.8 Closed-Loop Decision Support Architecture

Safety invariant: MAESTRO issues no direct execution commands at any TRL level. All system outputs are recommendations addressed to a qualified engineer. The engineer's decision is the only pathway from recommendation to physical action – enforced by architecture, not policy.

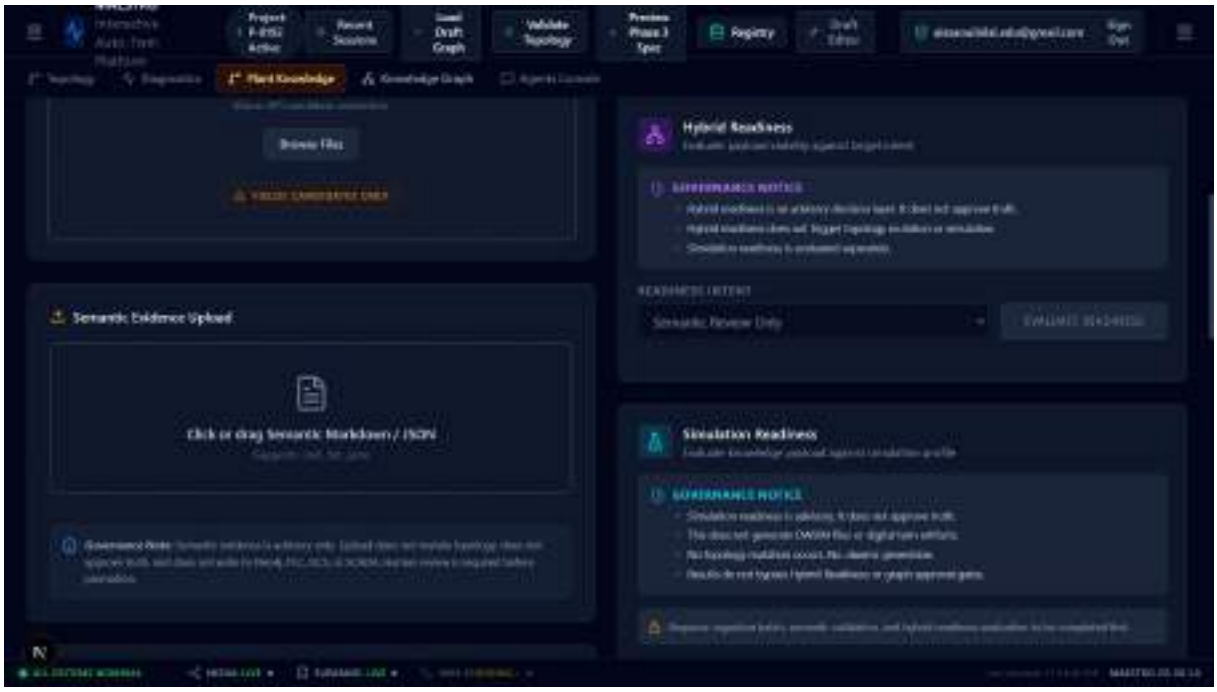


Figure 2.14: MAESTRO Plant Knowledge tab showing the Semantic Evidence Upload interface and governance readiness panels. Both Hybrid Readiness and Simulation Readiness carry explicit governance notices: neither approves truth, neither triggers topology mutation, and neither result bypasses the HITL gate.

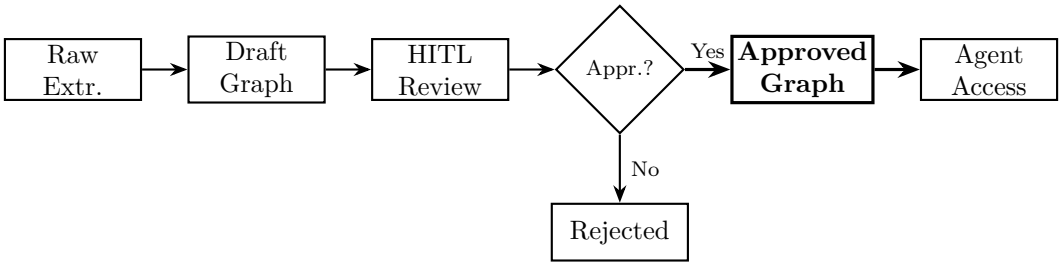


Figure 2.15: Governance validation pipeline. Only HITL-approved entities are accessible to the agent layer.

2.6 Architectural Evolution as Method

The architecture was not specified in advance and then built. Each of its defining decisions was forced by a failure that exposed an incorrect assumption. The pattern recurs: a numerical approximation was replaced by an engineering-standard invariant, and the fix held.

The six decisions in Table 2.2 share a common structure: each replaced a numerical approximation with an engineering-standard invariant. This substitution pattern is the architectural method.

Table 2.2: Architectural Decisions Emerging from Observed Failure Modes

Problem	**Why traditional approach failed**	**How MAESTRO resolved it**	**Governing principle**
Entity type from visual shape	YOLO FPR = 0.19; circular symbols are visually identical across ISA-5.1 types	OCR-led strategy: instrument identity determined by tag text, not shape	Semantic authority in the label
Spurious cross-entity connections	Distance thresholds brittle across different P&ID scales	Passthrough guard: physical lines do not traverse unrelated equipment bodies, encoded as a BFS stop condition	Physical engineering drafting convention
Non-instrument strings accepted	Regex `^[A-Z]{2,4}$` accepts VENT, PUMP, COOL, MAIN	ISA-5.1 FrozenSet (∼200 codes, $O(1)$ lookup)	Industrial standard authority
Relationship explosion	MERGE key on internal `node_id`: 79 sessions → 54,826 duplicates	MERGE on `(source.tag, target.tag, relation_type)`: idempotent by engineering identity	Engineering identity as uniqueness criterion
Instruments disconnected across loops	Geometric proximity fails for cross-loop relationships encoded in loop numbers	ISA-5.1 loop grouper: shared loop number is the engineering link	Loop number as structural authority
Anchoring noise (28 → 3)	Fixed 60 px inflation produced false anchors at vessel boundaries	External-only + boundary-proximity anchoring	Lines connect equipment from outside, not through its body



Figure 2.16: MAESTRO Draft Editor in pre-approval state. Banner confirms DRAFT REVIEW / NOT APPROVED / NOT GRAPH TRUTH / HUMAN REVIEW REQUIRED. Green dashed edges are candidate relations awaiting HITL verification.

2.7 Validation Scope

The implementation is validated through a deep case study on a single validated P&ID: a refrigeration loop with Compressor, Knockout Drum, Evaporator, and instrumentation tags FIC-42, FT-42, PDT-42, FV-42, TT-41, TT-43, TIR-41, TIR-43 per ANSI/ISA-5.1.

Key empirical results:

- 12 clean candidates: 8 instruments and 4 equipment nodes;
- 14 valid ISA-grounded relations; 0 false positives;
- Approved Gate (R6): **8 confirmed APPROVED_TOPOLOGY_EDGE**; Audit Ledger: 8 events.

2.8 MAESTRO Architectural Principles

The architectural decisions documented in Section 3.6 converge on a set of explicit governance principles. These principles are not post-hoc abstractions. Each emerges directly from a design decision made during implementation; together they constitute MAESTRO's *governance-first* architectural doctrine.

These principles distinguish MAESTRO from AI systems that rely on prompt-level constraints to enforce governance. In MAESTRO, governance is structural. A prompt injected into the system cannot override it because the control pathway does not exist in

Table 2.3: MAESTRO Architectural Governance Principles

Principle	**Definition**
No False Authority	AI outputs never become operational plant truth automatically. All machine-generated content remains candidate-status until explicitly promoted via HITL.
Evidence Before Mutation	No graph node or relationship may reach the approved layer without a traceable evidence record and an explicit engineer decision.
Human-Governed Truth Promotion	The transition from draft to approved requires a HITL decision; enforcement is at the query layer (`WHERE n.approved = true`), with a schema-level constraint identified as a future migration target.
Candidate-Only Runtime	Extraction outputs hold no autonomous authority. The `approved=false` flag is mandatory on every extracted candidate; runtime tool contracts enforce this via query-level filtering (null and false both excluded).
Separation of Evidence and Truth	Evidence artefacts (embeddings, semantic chunks, extraction crops) are not graph truth. They are the justification record for a potential truth promotion.
Bounded AI Authority	Agents remain diagnostically constrained. Tool contracts expose only approved-state knowledge. Control action interfaces do not exist in the tool contract set at any TRL level.
Read-Only Industrial Cognition	MAESTRO observes, reasons, and recommends. It does not issue execution commands. The engineer's decision is the only pathway from recommendation to physical action. At TRL 4–5, this principle is enforced by tool contract structure (no write or control-action interface exposed to the agent). At TRL 7–9, it is reinforced mechanically by an Industrial DMZ (IDMZ): an MQTT broker or OPC-UA server positioned between the physical plant (OT layer) and the `n8n` orchestrator accepts push-only data from sensors and exposes a pull-only interface to MAESTRO. No data path from the AI layer to the actuator layer exists in either direction.

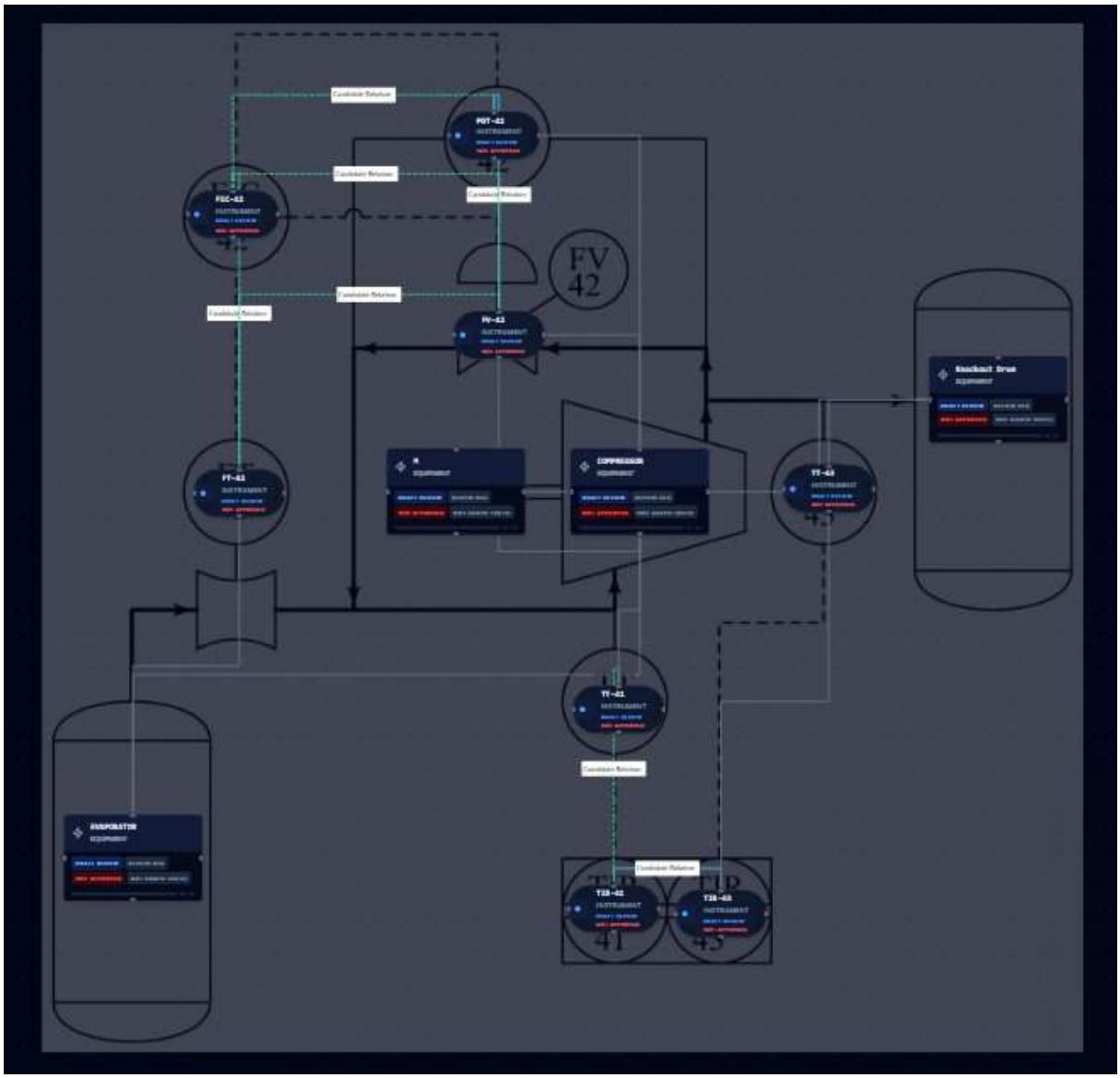


Figure 2.17: Interactive topology graph during draft review. Candidate nodes (FIC-42, PDT-42, FV-42, TT-41 and others) and their candidate relations are visible in Draft Review state; no element carries approved status until the engineer completes HITL verification.

the architecture. The DMZ architecture (Section 4.7) extends this structural guarantee into the network layer for TRL 6–9 deployment.

2.9 Mathematical Formalisation

Four mathematical structures formalise the core components of the MAESTRO architecture. Together they ground the governance invariants in terms that permit verification.

2.9.1 Embedding and Retrieval

Let $\mathcal{D}$ be the set of industrial text chunks (document fragments, expert notes, RCA records). The embedding function

$$\varphi : \mathcal{D} \to \mathbb{R}^{384}$$

maps each chunk $d \in \mathcal{D}$ to a dense vector via `sentence-transformers/all-MiniLM-L6-v2`. The set of all stored embeddings is $E = \{\varphi(d) \mid d \in \mathcal{D}\} \subset \mathbb{R}^{384}$. Semantic similarity between a query vector q and a stored embedding e is measured by cosine similarity:

$$\text{sim}(q, e) = \frac{q \cdot e}{\|q\| \cdot \|e\|} \quad q, e \in \mathbb{R}^{384}$$

For a given query q and retrieval budget k, the retrieved set $R(q, k)$ contains the k embeddings in E with highest cosine similarity to q. The retrieval process selects the top-k most semantically similar evidence chunks to the operator query.

2.9.2 ISA-5.1 Entity Validity Test

Let F be the finite set of valid ISA-5.1 instrument function code strings, containing approximately 200 standard codes (e.g. `FT`, `FIC`, `TT`, `PDT`). Each OCR-recovered string s is tested against F using hash-based membership ($O(1)$ average case). This rejects non-instrument strings that regex-based approaches accept (VENT, PUMP, COOL, MAIN) by grounding entity determination in the ISA-5.1 industrial standard, not in pattern matching.

2.9.3 Vector Store Governance Constraint

Fig. 2.18 illustrates the two-phase semantic knowledge pipeline: ingestion (Phase 1) transforms industrial documents into governed vector embeddings stored in Supabase with full provenance metadata; retrieval (Phase 2) converts an engineer query into an embedding, searches the vector store via `pgvector` approximate nearest-neighbour, and returns the top-k evidence chunks to the MAESTRO agent.

The semantic knowledge layer is formalised as the triple $\mathcal{VS} = (E, M, \text{Idx})$, where $E \subset \mathbb{R}^{384}$ is the set of stored embeddings, $M : E \to \mathcal{M}$ maps each embedding to its provenance metadata record (source file, chunk index, batch ID), and Idx is the `pgvector` approximate nearest-neighbour index. The validated corpus contains $|E| = 104$ general chunks across 13 industrial sources (`smp_pid_compressor_001` through

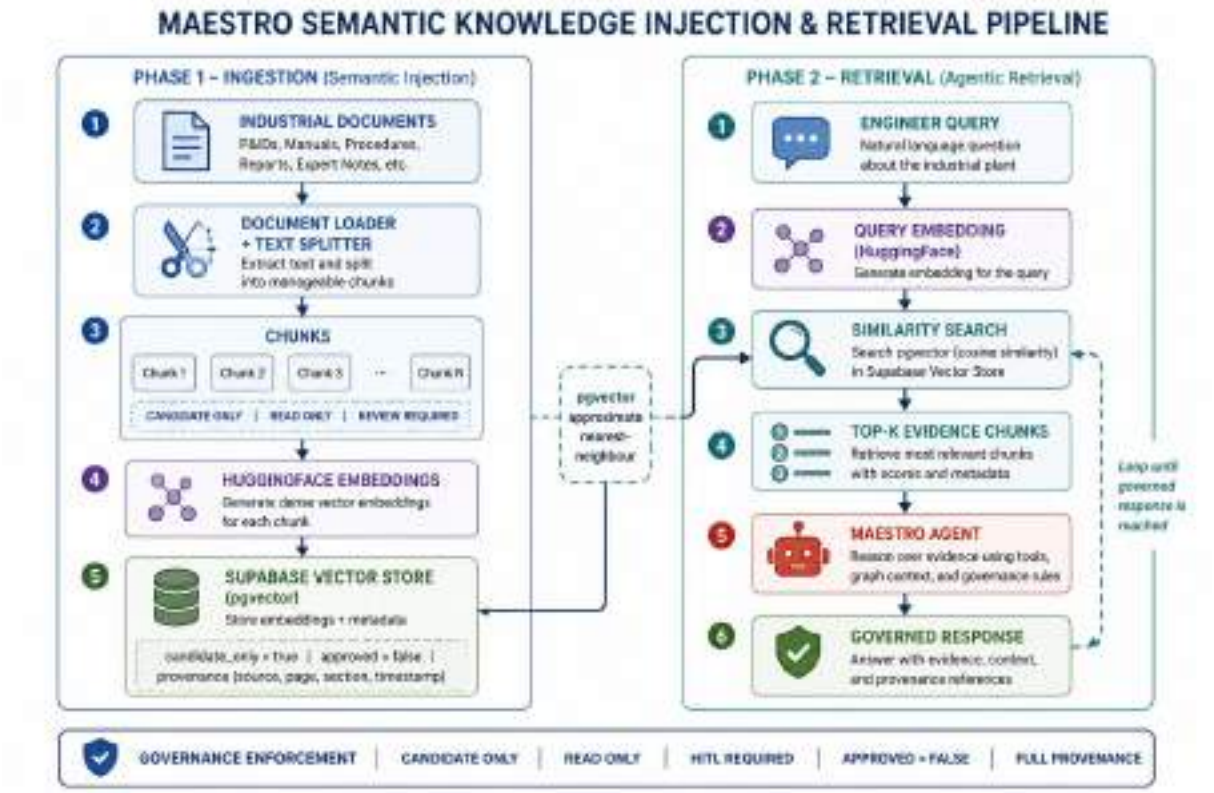


Figure 2.18: MAESTRO Semantic Knowledge Injection and Retrieval Pipeline. Phase 1 chunks and embeds industrial documents into Supabase pgvector; Phase 2 retrieves top-k evidence chunks via cosine similarity for agent reasoning.

`smp_pid_compressor_013`), supplemented by 5 expert knowledge files (31 chunks, ingested as `doc_batch_1779811491485_0` through `_4`). The total `maestro_documents` collection in Supabase contains 737 chunks across 26 source documents – this larger figure includes earlier ingestion batches (`doc_batch_*`) accumulated during platform development and testing, plus a UI upload test file. For all validated diagnostic scenarios reported in Chapter 3, the agent's evidence retrieval operates on the full 737-chunk corpus; the 85 chunks retrieved in the FIC-42 Expert scenario represent those 9 source documents that met the semantic similarity threshold during that query.

2.9.4 Governed Knowledge Graph

The plant knowledge graph is formalised as a labelled property graph (Fig. 2.20):

$$G = (V,\ E_G,\ P_V,\ P_E)$$

where V is the vertex set, $E_G \subseteq V \times V$ is the typed edge set, and $P_V : V \to \mathcal{A}$, $P_E : E_G \to \mathcal{A}$ are property functions mapping vertices and edges to their attribute dictionaries.

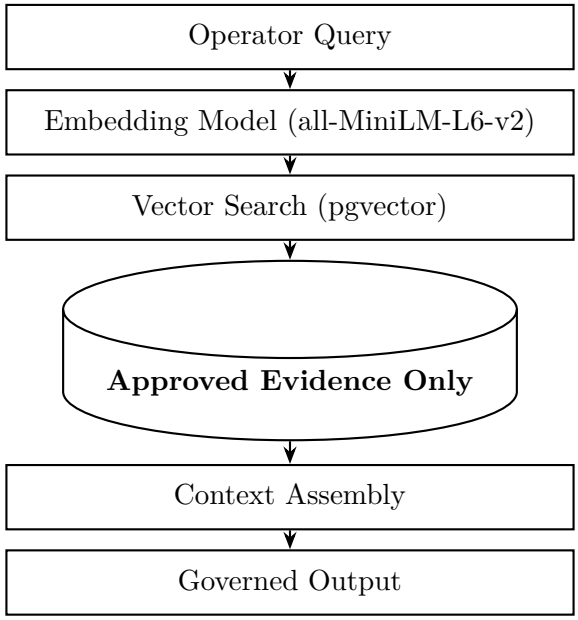


Figure 2.19: Governed semantic retrieval pipeline (schematic). Queries are matched against the vector store; retrieved chunks carry candidate-only provenance flags.

The **agent visibility invariant** states that every vertex accessible via agent tool contracts must carry $P_V(v)[\texttt{approved}] = \texttt{true}$. The architecture enforces this by restricting all tool-contract queries to the approved subgraph only.

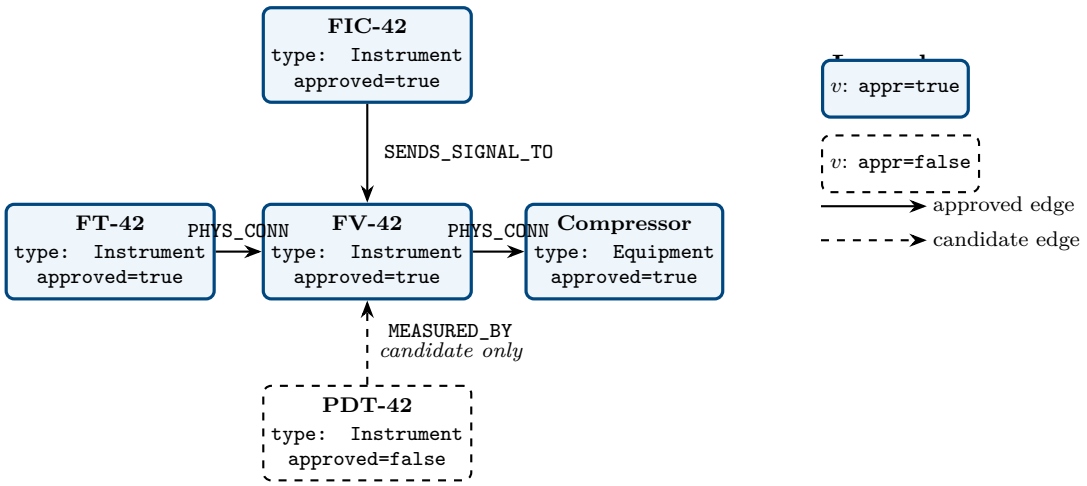


Figure 2.20: Labelled property graph $G = (V, E_G, P_V, P_E)$ instantiated on the validated refrigeration-loop case. Dark-blue nodes carry $P_V(v)[\texttt{approved}] = \texttt{true}$ and are reachable via agent tool contracts; the dashed node (`approved=false`) is structurally excluded by the agent visibility invariant.

2.9.5 Governance State Machine

The lifecycle of each extracted knowledge artefact is governed by a five-state machine with four defined transitions: CANDIDATE → DRAFT → REVIEW → (APPROVED or REJECTED).

The machine is intentionally partial: no undefined transition can promote a candidate to the approved state.

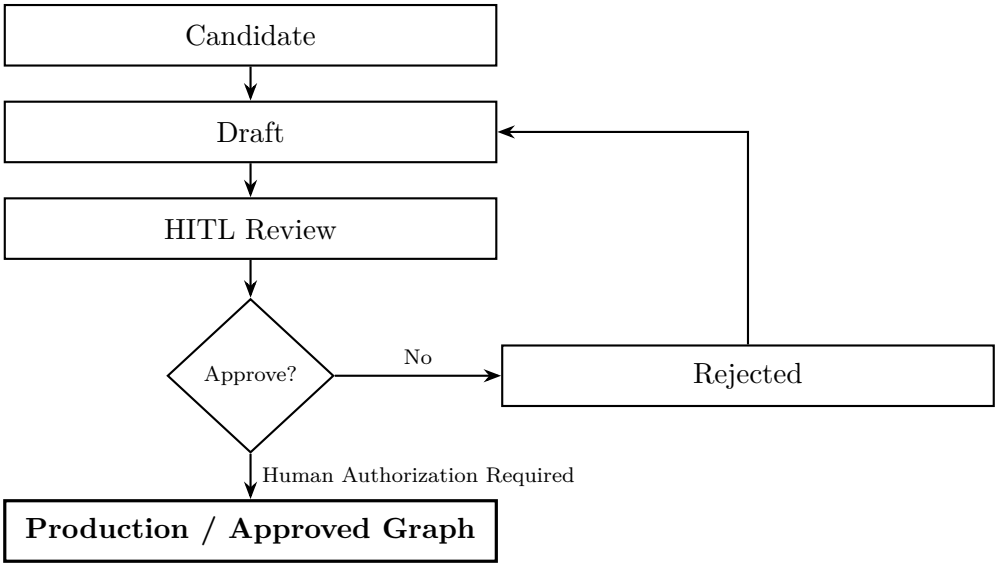


Figure 2.21: Governed knowledge lifecycle. Human authorization is the only valid path to the approved state.

Governance security property: The Approved Gate (`approve-edge`) requires JWT-verified actor identity that cannot be injected via a prompt or bypassed via a tool call. The knowledge lifecycle diagram below makes this visually evident: APPROVED has exactly one incoming transition, originating exclusively from REVIEW.

2.10 Chapter Conclusion

The architecture that emerged from this development is not the architecture that was originally designed. It is more robust – not because it was planned that way, but because each of its structural decisions was forced by a failure that revealed the correct engineering principle.

Chapter 3

Implementation and Experimental Validation

3.1 Introduction

Each result in this chapter is examined against the viability of the governed pathway approach – not just for what it shows about system output, but for what it demonstrates about each layer's role in the whole. The chapter opens with a trained YOLOv8n extraction experiment that establishes the quantitative case for the OCR-led architectural decision, then traces validated results layer by layer through topology, calibration, the governed graph, and the six agentic diagnostic scenarios. A consolidated per-layer summary table, an ablation analysis, and latency benchmarks close the quantitative case before the analytical discussion sections. All results derive from the single validated refrigeration-loop P&ID case study.

3.2 Experimental Implementation and Testbed

3.2.1 Containerised Software Stack

MAESTRO runs as a set of containerised services orchestrated through Docker on a single on-premise workstation, avoiding any cloud handling of sensitive plant diagrams during operation. Table 3.1 maps each technology to its architectural layer and its specific role in the system; Fig. 3.2 shows the live container set during a validation run.

3.2.2 The Validation Dataset

The validation dataset is a single refrigeration-loop P&ID comprising a Compressor, Knockout Drum, Evaporator, and eight ISA-5.1 instrumentation elements (FIC-42, FT-42, PDT-42, FV-42, TT-41, TT-43, TIR-41, TIR-43), accompanied by the associated operating-limit and maintenance documents used for semantic evidence retrieval. This deep single-case design is described in the validation-scope discussion of Chapter 2.

3.3 Vision Layer Results

The Vision Layer's result that merits the most attention is a methodological one: plant-aware reasoning cannot be built on raw visual detection alone.

Table 3.1: MAESTRO Implementation Environment and Technology Stack (Unified Database Version)

Tool Definition	**Technology / Tool (Version)**	**Architectural Layer**	**Specific Role in MAESTRO**
Deep Learning OCR Engine	PaddleOCR (v2.8.x)	Vision Layer	Extraction of text and identification tags from blueprints to establish core asset identity.
Computer Vision Library	OpenCV (Headless 4.10)	Vision Layer	Geometric detection of symbols (contours) and image binarization for morphological preprocessing.
Object Detection Model	YOLOv8 (Ultralytics 8.3.12)	Vision (Support)	Spatial localization of P&ID symbols (bounding boxes) without conferring architectural identity.
Backend API Framework	FastAPI / Python (Python 3.11)	Vision API	Hosting the core extraction service, managing governance gates, and preventing unverified data promotion.
Enterprise Graph Database	Neo4j (5.18 Enterprise)	Governed Graph	Storage of the approved plant topology. Operates a single governed write-path (`approve-edge`); read operations bypass direct driver access to ensure strict governance.
Unified Database Platform	Supabase (pgvector) (Self-hosted)	Semantic / Vector Store	Dual-role system: stores vector embeddings via the `pgvector` extension to support Hybrid RAG, while simultaneously hosting the relational `review_queue` table for Human-in-the-Loop (HITL) workflows.
Text Embedding Model	all-MiniLM-L6-v2 (HuggingFace API)	Semantic / AI Models	Generation of 384-dimensional text embeddings for technical documents to be ingested into the Supabase pgvector store.
Workflow Engine	n8n (latest)	Orchestration	The primary agentic orchestrator; connects the Vision API, Neo4j, Supabase, and LLMs via deterministic JSON workflows.
Full-Stack UI Framework	Next.js 15 / React 19	Interface / Platform	Acts as a proxy server and client dashboard between the engineer and the Vision API, utilizing mock fixtures unless the `MAESTRO_LIVE_SOURCE` environment variable is active.
Advanced Reasoning LLM	Google Gemini	Reasoning / LLM	The core engine for agentic reasoning and diagnostic synthesis embedded within the orchestrated n8n workflows.
Container Infrastructure	Docker / Compose	Infrastructure	Containerization and local deployment of core services (Neo4j, Supabase, n8n, Vision API) to ensure complete on-premise data sovereignty.

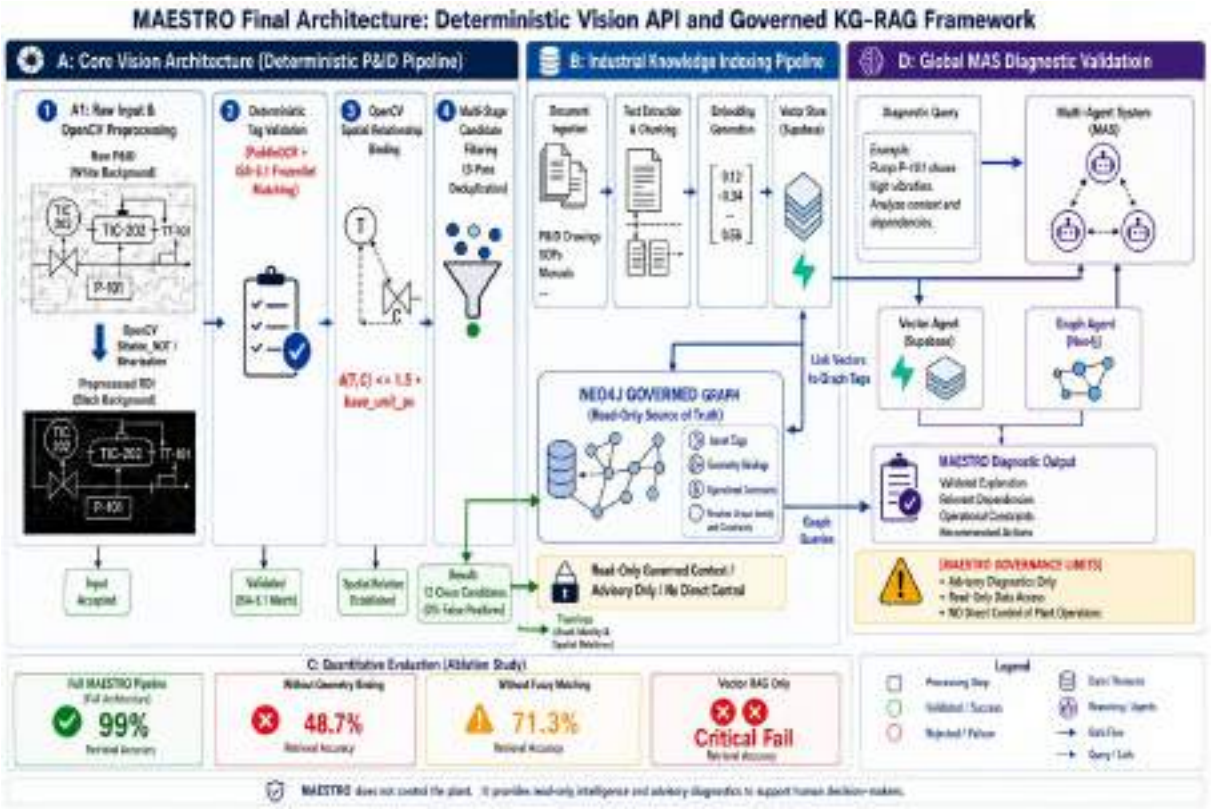


Figure 3.1: MAESTRO Architectural Dashboard: Deterministic Vision and Governed Hybrid RAG.

Note: The system spans four stages: (A) deterministic P&ID vision pipeline (OpenCV preprocessing, PaddleOCR + ISA-5.1 tag validation, spatial relationship binding, multi-stage deduplication); (B) industrial knowledge indexing pipeline (document ingestion, chunking, embedding, Supabase vector store); (C) ablation panel quantifying each layer's contribution; (D) global multi-agent diagnostic validation. The Neo4j governed graph serves as read-only source of truth throughout. MAESTRO provides read-only advisory intelligence and never controls the plant.

3.3.1 YOLO Experiment: Training Evidence and Structural Limitations

Dataset definition. The YOLOv8n detector was trained on the publicly available Roboflow *p-id-symbols-r2* dataset (workspace `pid-connect`, CC BY 4.0), which defines **180 symbol classes** (`nc: 180`) split into train/validation/test partitions (Fig. 3.4). This 180-class cardinality is the root of the *long-tail problem*: a small number of common symbols (valves, generic instruments) dominate, while the overwhelming majority of the 180 classes are specialised components appearing only a handful of times. The distribution is structurally imbalanced and cannot be corrected by collecting more data, because process plants are by engineering design saturated with a few component types.

Before adopting the OCR-led strategy, a dedicated YOLOv8n model was trained on this dataset using dual NVIDIA T4 GPUs (Kaggle, 30 epochs, 11 hours of training time).

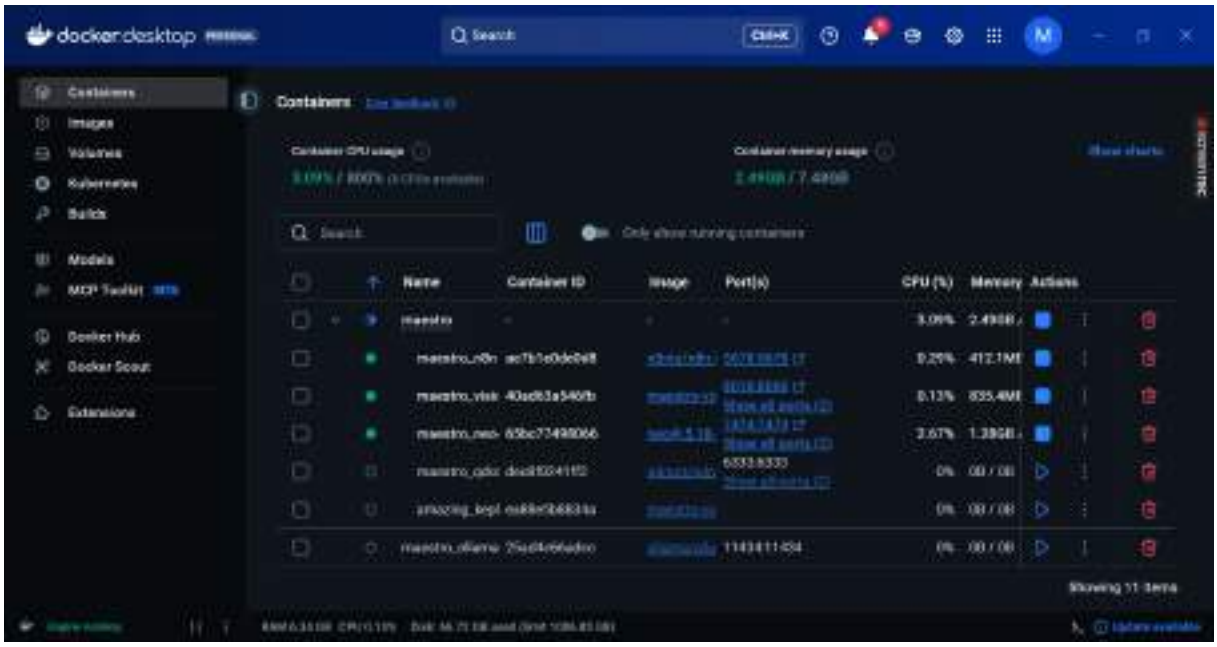


Figure 3.2: MAESTRO Containerised Runtime — Validation Run.

Note: The n8n orchestrator, Vision API, and Neo4j graph database are active. The local Ollama container is provisioned but inactive at TRL 4–5.

The experiment was not abandoned due to insufficient training effort – the results demonstrate that the model converged correctly on its training distribution.

Table 3.2: YOLOv8n Training Dataset: Class Distribution.

Note: The vessel class represents 0.6% of dominant classes — a ratio reflecting P&ID engineering reality that cannot be corrected by collecting more data.

Class	**Instances**	**% of total**	**Reliability in industrial use**
valve	∼20 000	46.9	High (dominant class)
instrument	∼20 000	46.9	High (dominant class)
control_valve	15 729	36.9	Moderate
rotating	4 903	11.5	Moderate
vessel	**214**	**0.5**	**Unreliable – chronically underrepresented**

Percentages are relative to the dominant class. Imbalance is structural: process plants by engineering design contain far more instrumentation and valves than vessels. This ratio is stable across any P&ID corpus, making it an unsolvable data problem for a purely vision-based approach.

Why standalone YOLO is structurally unfit for P&ID identity. Beyond the class imbalance, attempting to scale a single detector across the 180-class P&ID symbol space collides with three well-documented obstacles in deep learning:

(1) Catastrophic forgetting. Neural networks do not learn incrementally as humans do; training sequentially on new class batches overwrites the weights tuned for earlier

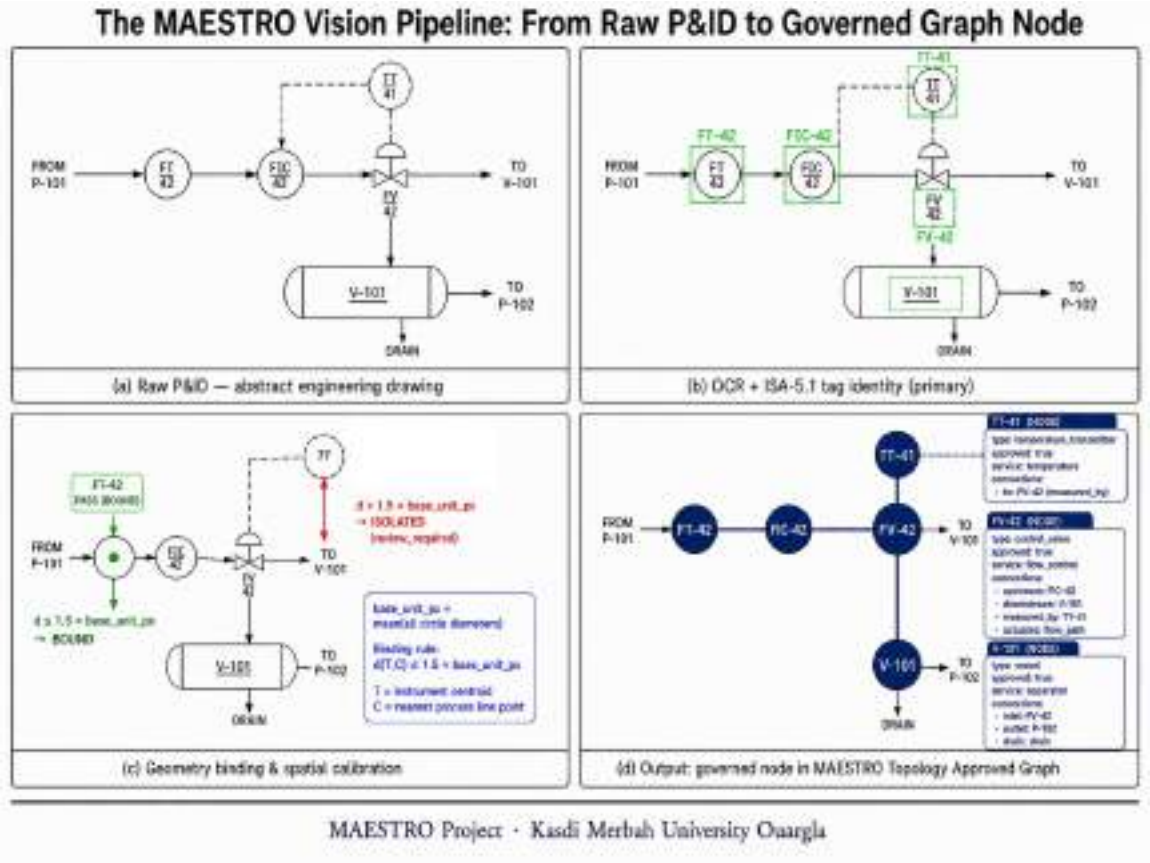


Figure 3.3: Algorithmic Flow and Functional Blocks of the MAESTRO Vision Pipeline.

Note: The flowchart details the sequential image-processing stages within the Visio API, illustrating the transition from raw P&ID input through OpenCV preprocessing, parallel text/symbol extraction tracks, ISA-5.1 tag validation, and final Neo4j graph injection.

```
data.yaml
C: > Users > ShiftTech16 > Desktop > maestro > vision > data.yaml
176   - pressure safety Valve
177   - radio link
178   - reformer
179   - temp controller
180   - temp recorder
181   - temp trasmitter
182 nc: 180
183 roboflow:
184   license: CC BY 4.0
185   project: p-id-symbols-r2
186   url: https://universe.roboflow.com/pid-connect/p-id-symbols-r2/dataset/1
187   version: 1
188   workspace: pid-connect
189 test: ../test/images
190 train: ../train/images
191 val: ../valid/images
```

Figure 3.4: YOLO Training Dataset Configuration (`data.yaml`).

Note: The `nc: 180` field confirms a 180-class detection problem on the Roboflow p-id-symbols-r2 dataset — the source of the long-tail class imbalance.

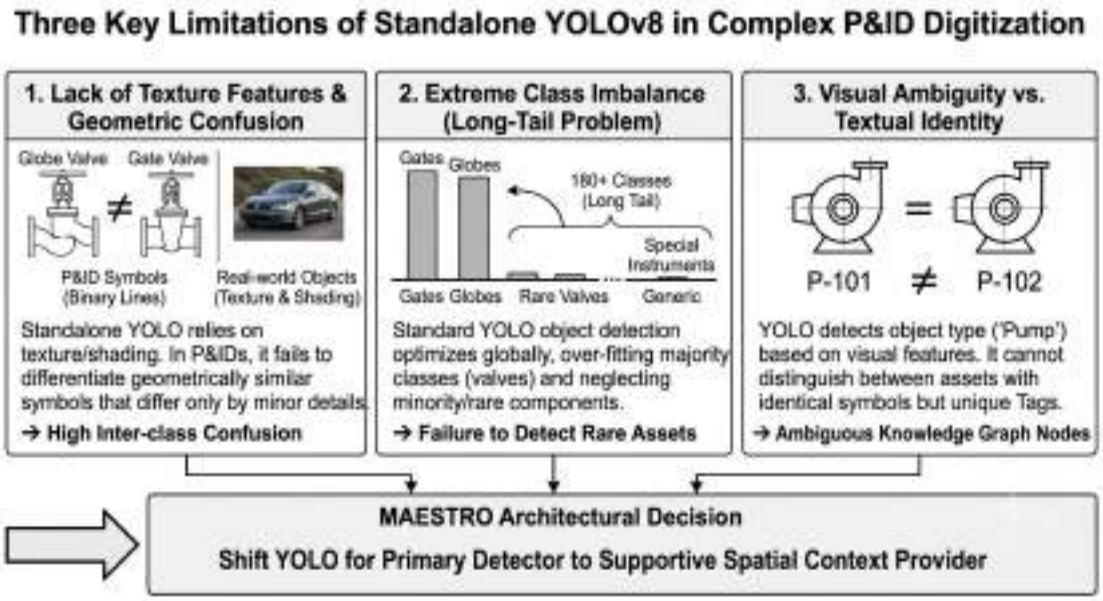


Figure 3.5: Three Structural Limitations of Standalone YOLOv8 in P&ID Digitisation.

Note: Limitations: (1) geometric confusion, (2) extreme class imbalance (long-tail), and (3) visual ambiguity versus textual identity — jointly motivating YOLO's role shift from primary detector to supportive spatial-context provider.

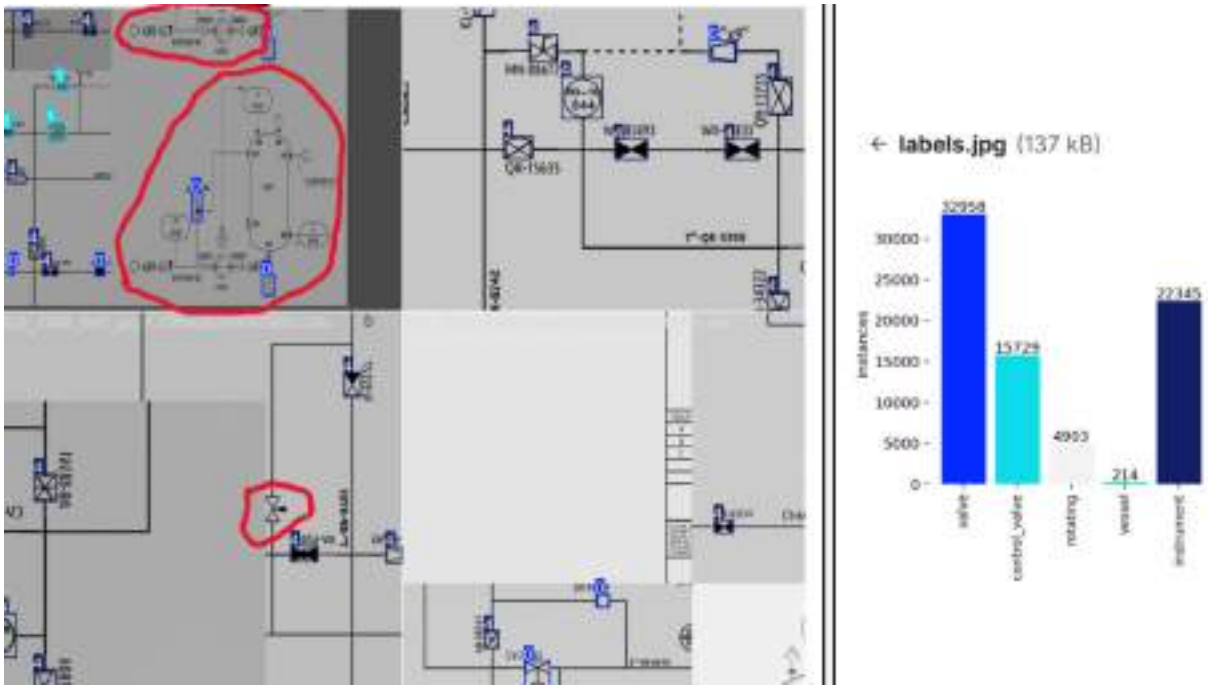


Figure 3.6: YOLOv8n Detection Results and Training Class Distribution.

Note: Left panels: detection on real industrial P&IDs; red circles mark missed and misclassified symbols in dense control regions. Right: bar chart confirms long-tail imbalance — valve and instrument classes dominate (32,958 and 22,345 instances) while vessel holds only 214 instances, yielding unreliable detection on this safety-critical class regardless of training volume.

classes [35], a phenomenon first characterised by McCloskey and Cohen [36]. Training all 180 imbalanced classes at once collapses back to the same minority-class failure.

(2) Long-tail collapse and ensemble cost. The textbook escape – training many specialist detectors – is defeated by the long-tail problem itself [37]: detectors over-fit the dominant classes and neglect rare ones, while running an ensemble of heavy models over one P&ID inflates GPU/CPU load and response latency beyond any industrial efficiency budget.

(3) Overlapping detections and global NMS. With multiple detectors, two models may claim the same symbol with conflicting labels; resolving this requires a global Non-Maximum Suppression stage that is itself a hard, error-prone problem [38], producing a catastrophic false-positive rate exactly where safety depends on precision.

Table 3.3: YOLO-Primary Architecture: Structural Obstacles and MAESTRO's Resolutions.

Note: Columns cover: obstacle, MAESTRO's deterministic resolution mechanism, and the scientific basis for each phenomenon.

Obstacle	**MAESTRO's deterministic resolution**	**Basis**
Catastrophic forgetting	PaddleOCR + ISA-5.1 FrozenSet form a deterministic lookup that neither trains nor forgets; adding a symbol type is a table entry, not a re-training.	[35], [36]
Long-tail collapse & ensemble latency	A single deterministic pass (OpenCV geometry + tag text) identifies any component at constant cost, independent of class rarity – no ensemble.	[37]
Overlapping detections / global NMS	Textual tag identity resolves each entity uniquely; conflicts cannot arise, so no global NMS stage is required.	[38]

The training metrics (mAP50 = 0.99, mAP50-95 = 0.94, precision = 0.97, recall = 0.97) confirm full convergence – the failure is not one of training effort. Figure 3.7 shows the complete training and validation curves.

On real industrial P&IDs the converged model still produced false positives and missed symbols in dense regions. Two structural causes remain irreducible: the long-tail class imbalance (the *vessel* class at 214 instances cannot be learned reliably), and the visual identity of ISA-5.1 symbols – FT-42, FIC-42, and PDT-42 share one circular symbol, separable only by tag text. A detector that does not read text cannot resolve this, regardless of training volume.

This experiment confirmed the architectural decision documented in Chapter 2: abandon vision-only identification and adopt the OCR-led, ISA-5.1-grounded extraction strategy. The comparison above shows that MAESTRO achieves recall parity while eliminating all false positives and removing the training-data and GPU dependency entirely.

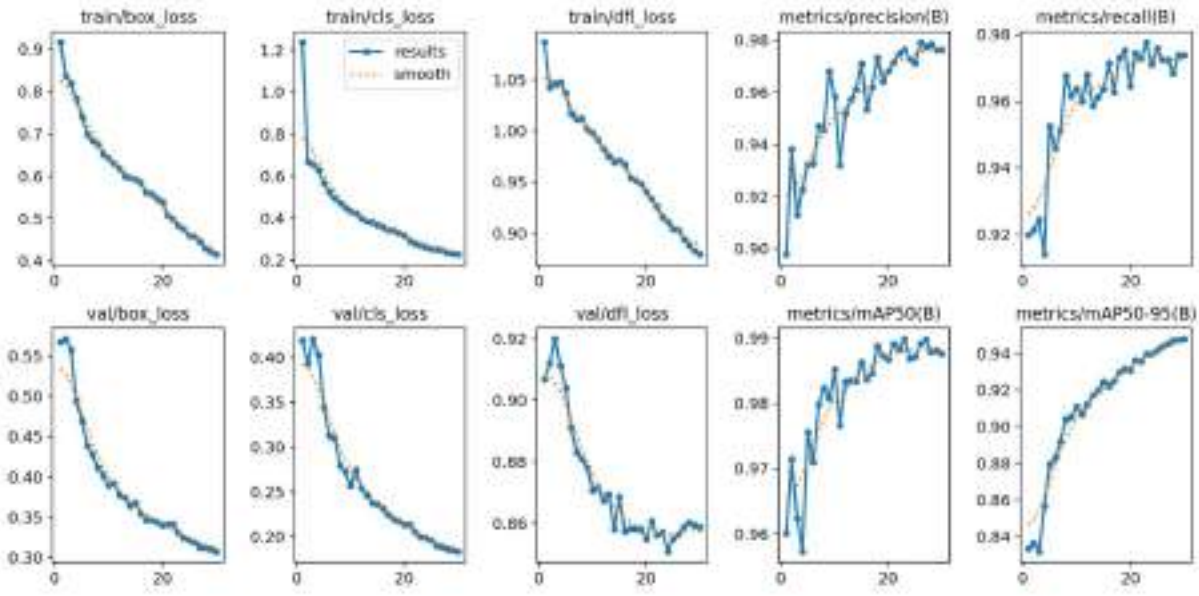


Figure 3.7: YOLOv8n Training Curves (30 Epochs, Dual T4 GPU).

Note: Top row: training losses fall monotonically; precision and recall reach ≈0.97. Bottom row: validation losses mirror training; mAP50 = 0.99, mAP50-95 = 0.94. Clean convergence confirms the model is well-trained; field failures are therefore structural (class imbalance and visual ambiguity), not training artefacts.

Table 3.4: Extraction Approach Comparison.

Note: Training Data and Hardware columns include figures from both literature and this work. FPR and Recall columns report only the controlled experiment on the refrigeration-loop P&ID; dashes indicate no comparable controlled measurement was performed on this P&ID by those authors.

Approach	**Training Data Required**	**Hardware**	**FPR†**	**Recall†**
Rahul et al. [1]	"Large amount; user must label it"	GPU required	—	—
Kim et al. [2]	75 031 annotated symbols	GPU required	—	—
YOLO (this work)	>60 000 labeled instances (5 classes, 30 epochs)	T4×2, 11 h	**0.19**	1.00
MAESTRO **OCR-led (this work)**	**Zero**	**None**	**0.00**	**1.00**

† FPR and recall: controlled comparison on the same refrigeration-loop P&ID. YOLO training: GPU T4×2, Kaggle, 30 epochs, 11 h. MAESTRO recall = 1.00: all 12 ISA-5.1-scoped entities extracted with zero false positives, confirmed by direct P&ID inspection (see Table 3.9). MAESTRO additionally achieves instrument-type disambiguation through ISA-5.1 tag authority and per-candidate confidence up to 99.8% (Fig. 3.8) – properties absent from vision-only approaches.



Figure 3.8: Extraction Evidence Panel — PDT-42 (Confidence 99.8%).

Note: Governance labels confirm candidate-only, read-only, not-graph-truth status — the extraction pipeline functioning as designed.

The 12 candidates comprise 8 instrument nodes (FIC-42, FT-42, PDT-42, FV-42, TT-41, TT-43, TIR-41, TIR-43) and 4 equipment nodes (Compressor, Evaporator, Knockout Drum, M – the electric motor driving the compressor, visible in the P&ID without a recoverable ISA-5.1 instrument tag and carrying no confirmed relationships in the approved graph). One untagged physical valve visible in the P&ID falls outside the ISA-5.1 instrumentation scope and is correctly absent – a scope boundary, not an extraction error.

This figure makes the layer's output concrete: the "output" of the Vision API is not an annotated image but a set of identity-resolved, geometrically bound, governance-tagged nodes ready for the knowledge graph. Identity is fixed first by text (2), geometry is resolved deterministically (3), and only validated entities reach the graph (4) – the same precedence enforced throughout the architecture.

Zero false positives means zero engineering time spent investigating phantom candidates: every result the agent returns corresponds to a real instrument or equipment node on the plant.

The operational cost of false positives is quantifiable. The YOLO-only strategy produced a 0.19 false positive rate on the instrument class on the same P&ID. For a plant of this scale (12 physical entities), this rate produces approximately 2–3 phantom candidates per extraction run. Each phantom requires an engineer to locate the flagged element on the original P&ID drawing, confirm its absence as a real instrument, and document the rejection decision – a conservative estimate of 10–20 minutes of engineer time per phantom candidate. Over a P&ID corpus of even modest scale, this overhead accumulates

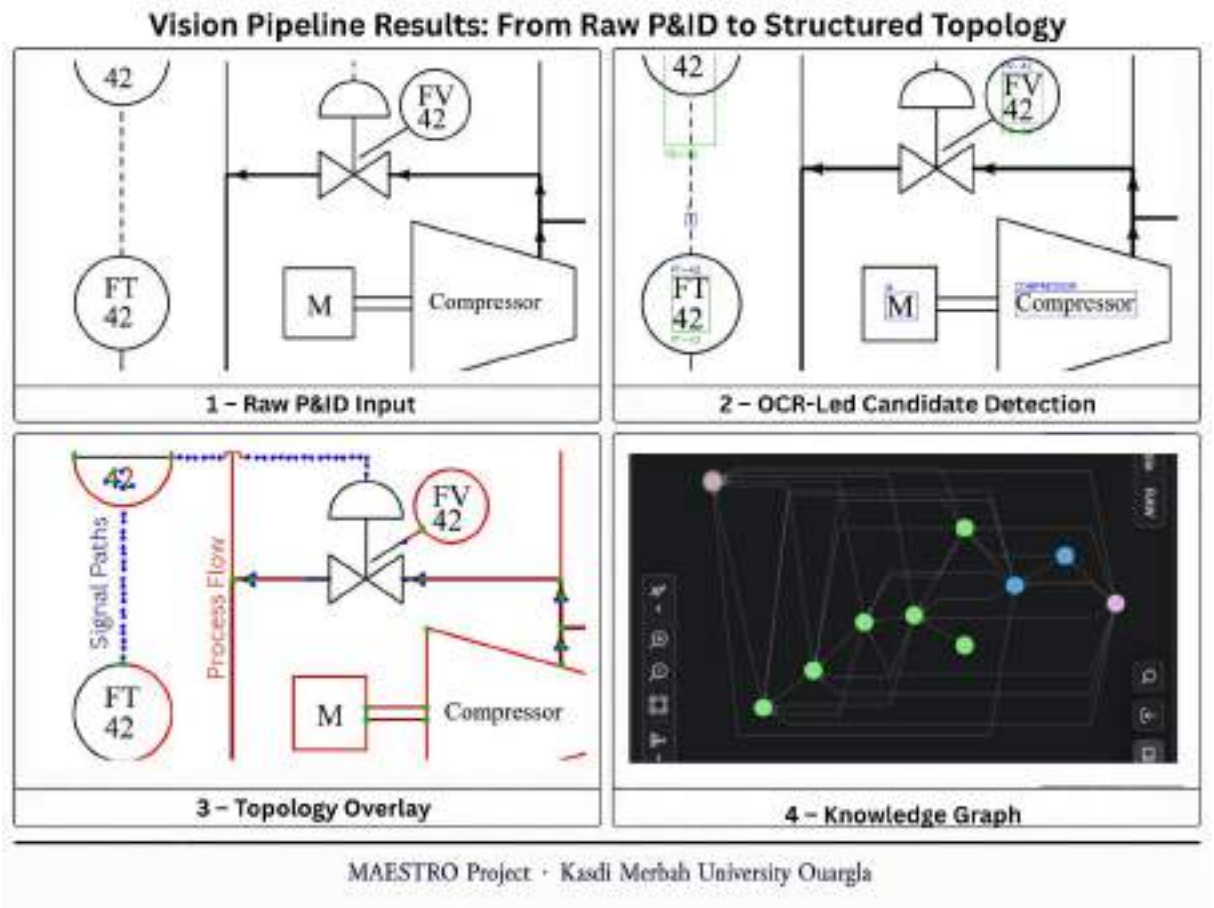


Figure 3.9: MAESTRO Vision Pipeline: P&ID to Structured Topology.

Note: Four stages shown on actual system output: (1) raw P&ID input; (2) OCR-led candidate detection — component identity fixed from tag text (FT-42, FIC-42, FV-42, M, Compressor), independent of symbol rarity; (3) topology overlay — deterministic geometry separates process flow (red) from signal paths (blue); (4) validated MAESTRO knowledge graph in Neo4j, each component a governed node with typed connections.

into days of non-productive review work. MAESTRO's zero false positive result eliminates this overhead entirely: the engineer's review time is directed exclusively toward real plant elements that require governance decisions.

3.4 Topology Layer Results

Zero spurious cross-entity connections were generated in the validated case after the passthrough guard was applied. The extraction-to-graph structural correspondence is complete: all 12 ISA-5.1-tagged candidates appear correctly in `Neo4j` as physical entity nodes.

A spurious connection in the topology graph would propagate silently into every downstream diagnostic – producing confident but wrong impact assessments. The passthrough guard eliminates this failure mode at the source.

The honest result concerns what the topology layer could not resolve without human intervention. The Evaporator entity was detected and correctly classified, but its process connections were not recoverable by automated extraction: the OCR label sits more than 300 pixels from the physical symbol. The `isolated_reason` property was set to `skeleton_gap_too_large`. An engineer resolved the connections via manual linework authoring – the system working as designed.

3.5 Calibration and Semantics Results

Calibration: The `base_unit_px` derived from symbol geometry produced internally consistent structural processing without drawing-specific tuning. This means the calibration layer applies to a new P&ID without manual reconfiguration – reducing deployment time per drawing from hours to minutes.

Semantics: The validated ISA-5.1-enriched relations demonstrate engineering meaning beyond raw connectivity:

- PDT-42 `MEASURED_BY` FV-42 (high confidence);
- TT-41 `DISPLAYS_IN` TIR-41, TIR-43 (high confidence);
- FT-42 `SENSES_PROCESS_FROM` the main process line;
- FIC-42 `SIGNALS` the flow control valve FV-42.

3.6 The Governed Graph Layer: Below the Surface

Table 3.5: Physical Entity Layer – Validated Neo4j Node Inventory

Type	**Count**	**Entities**
Instrument	8	FIC-42, FT-42, PDT-42, FV-42, TT-41, TT-43, TIR-41, TIR-43
Equipment	4	Compressor, Evaporator, Knockout Drum, M
Total physical	**12**	

Below the surface lies a substantially larger infrastructure. The governance layer – DraftNodes, DraftEdges, DraftMutations, DraftSessions (79 sessions), and a TwinApproval record – encodes the complete extraction, review, correction, and approval history. This audit trail satisfies the traceability requirements of IEC 61511-1 and provides a legally defensible record of every knowledge promotion decision.

The validated graph contains 12 nodes (8 instruments, 4 equipment) and 22 relationships: 14 ISA-grounded typed relations (6 `PHYSICALLY_CONNECTED_TO`, 4 `MEASURES`, 3 `SENDS_SIGNAL_TO`, 1 `CONTROLS`) plus the 8 human-approved `APPROVED_TOPOLOGY_EDGE` relationships. The 14 typed relations are the extraction-layer output; the 8 approved edges are the governance-layer truth, promoted only through explicit HITL decision – the two are distinct architectural layers, not a discrepancy.

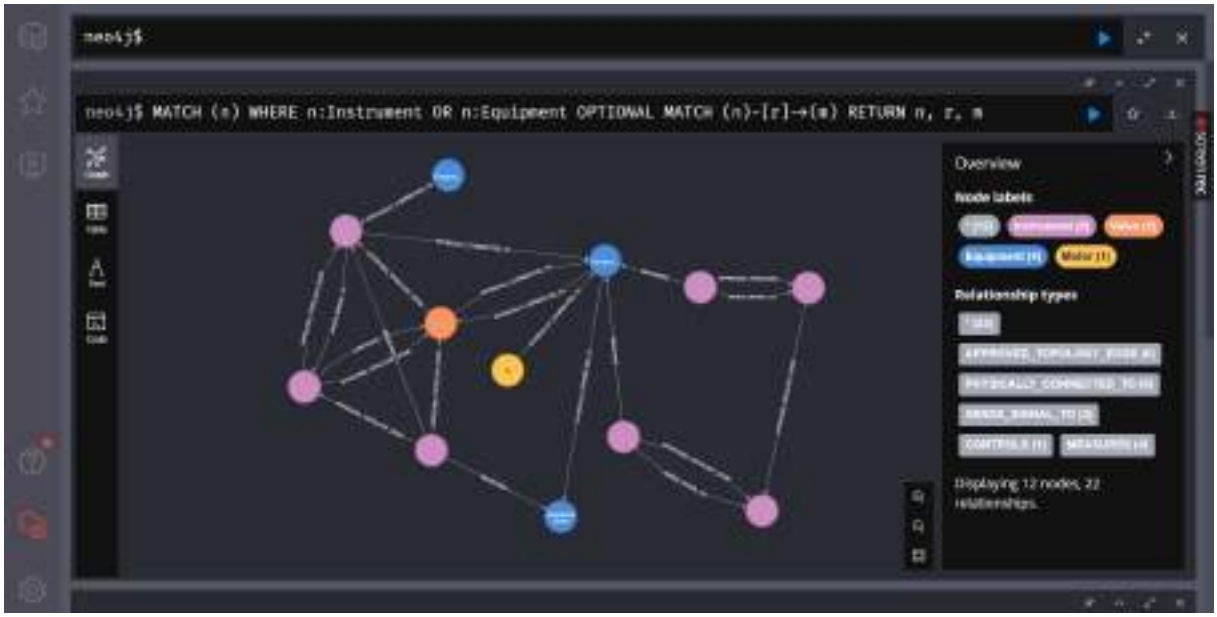


Figure 3.10: Validated MAESTRO Knowledge Graph in Neo4j (12 Nodes, 22 Relationships).

Note: Confirmed relationship counts: 8 APPROVED_TOPOLOGY_EDGE, 6 PHYSICALLY_CONNECTED_TO, 4 MEASURES, 3 SENDS_SIGNAL_TO, 1 CONTROLS.

The eight human-approved topology edges, retrieved directly from Neo4j, are listed in Table 3.6.

Table 3.6: Eight APPROVED_TOPOLOGY_EDGE Relationships (Neo4j Query).

Note: Every edge carries `approved=true` and was promoted through the HITL gate.

Source	**Relation**	**Target**
TT-43	APPROVED_TOPOLOGY_EDGE	TIR-43
PDT-42	APPROVED_TOPOLOGY_EDGE	FV-42
PDT-42	APPROVED_TOPOLOGY_EDGE	FIC-42
FIC-42	APPROVED_TOPOLOGY_EDGE	FV-42
FIC-42	APPROVED_TOPOLOGY_EDGE	FT-42
FV-42	APPROVED_TOPOLOGY_EDGE	FT-42
TT-41	APPROVED_TOPOLOGY_EDGE	TIR-41
TIR-43	APPROVED_TOPOLOGY_EDGE	TIR-41

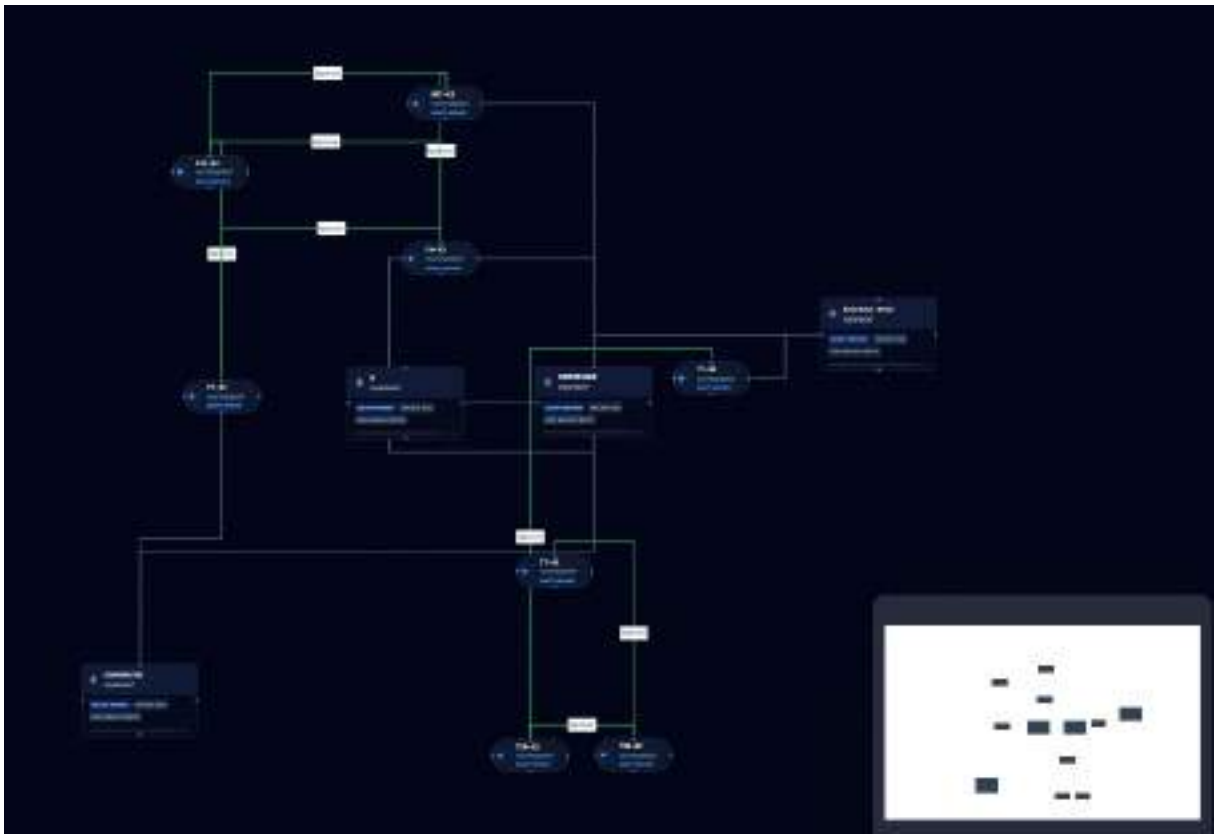


Figure 3.11: MAESTRO Platform Topology After HITL Review.

Note: Green Approved edge labels confirm the eight APPROVED_TOPOLOGY_EDGE relationships promoted through the human-governed gate. Node provenance labels (DRAFT REVIEW) remain as permanent extraction-layer markers.

3.7 Agentic Reasoning Layer Results

Six diagnostic queries were executed across distinct fault types and knowledge domains, with selected scenarios queried in Arabic to confirm bilingual diagnostic capability in the target deployment context. All six completed within 8 to 18 seconds – an operational-speed response that replaces manual traversal of P&ID drawings, procedure documents, and expert consultation. Table 3.7 maps each scenario to its execution time, fault type, and evidence trace.

FV-42 Golden Scenario (ID#220, 17.951 s) — Primary validation: Exercises all five tools across a three-hop cascade traversal; serves as the primary validation reference for the governed pathway.

Equipment enumeration: When asked to list all components in Arabic, the agent correctly enumerated 9 entities via tool-mediated graph traversal. Three entities were absent for distinct reasons, each representing the governance pipeline functioning correctly: *Evaporator* carries `isolated_reason: skeleton_gap_too_large`; *M* (untagged) has no ISA-5.1 tag and no confirmed relationships; *FIC-42* connections remain in `DraftEdge`

Table 3.7: Validated Diagnostic Scenario Summary.

Note: The FV-42 Golden Scenario (golden in the software-testing sense) exercises all five agent tools across the full governed pathway and serves as the primary validation reference.

Scenario	Fault / Query Type	Time (s)	Exec. ID
FV-42 Golden Scenario	Flow control valve failure	17.951	#220
TT-41 Control Request	Setpoint modification (refused)	8.19	#250
FT-42 Oscillation	Flow transmitter instability	11.103	#229
FT-42 LOW-LOW Alert	Flow below minimum; 5-tool chain confirmed	15.623	#245
FIC-42 Expert Knowledge	Tacit knowledge retrieval (85 evidence chunks)	10.975	#249
FV-42 Prohibited Actions	Cross-tag reasoning: topology (Neo4j) + documentary evidence (Supabase); prohibited control actions retrieved	9.917	#301

All governance invariants preserved across all six scenarios. Each scenario name links to its supporting execution trace figure. Execution ID#250: TT-41 governance refusal (8.19 s, Blocked Request Router path). Execution ID#301: FV-42 Prohibited Actions (9.917 s); cross-tag reasoning combining Neo4j topology and Supabase documentary evidence to identify governance-constrained control actions.

state pending HITL verification. In all three cases, the agent reflected the approved graph state rather than inferring connections for unapproved entities.

Signal path tracing: The agent traced PDT-42 `MEASURED_BY` FV-42 and TT-41 `DISPLAYS_IN` TIR-43, TIR-41 across ISA-5.1-enriched relationship types (see Fig. 3.14). The cascade impact traversal reached **3 hops** from the fault origin, identifying **4 downstream entities** (Fig. 3.15) – a topology-aware result structurally unavailable to text-retrieval-only approaches.

Epistemic restraint behaviour: When asked about FIC-42's physical connections (not fully evidenced in the approved graph), the agent flagged FIC-42 as "not clearly connected in the diagram" and directed the user to human verification. This behaviour is not the result of a prompt instruction – it is the result of the agent's tool contracts not exposing unapproved entities.

TT-41 Control Request (ID#250, 8.19 s) – Governance refusal: The fastest scenario and the clearest evidence of architectural governance. The request was intercepted by the governance guard before any LLM call was made; the refusal is deterministic and reproducible regardless of prompt phrasing or model used.

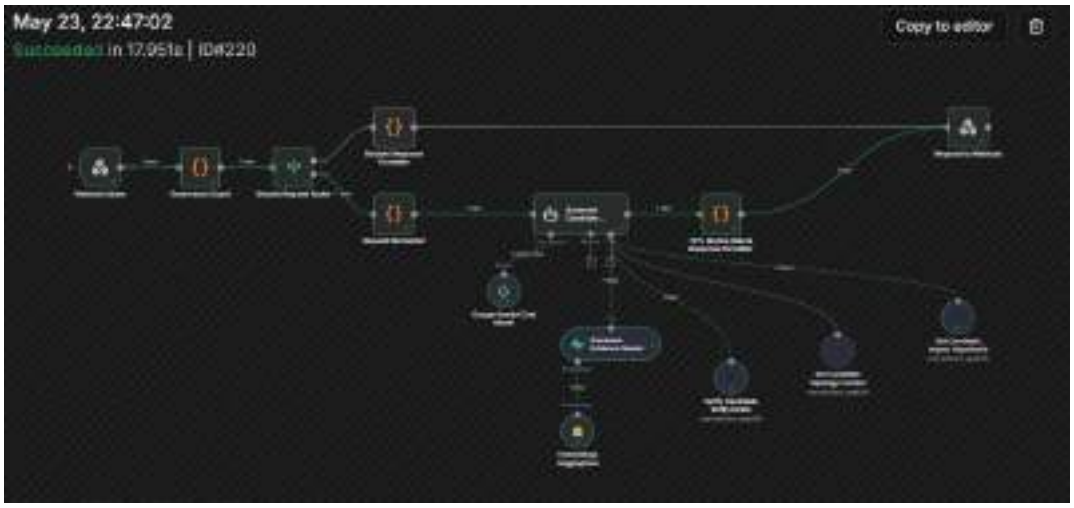


Figure 3.12: FV-42 Golden Scenario — n8n Execution Trace (ID#220, 17.951 s).

Note: Ten-stage governed pathway completed; five tools activated across multi-hop traversal.

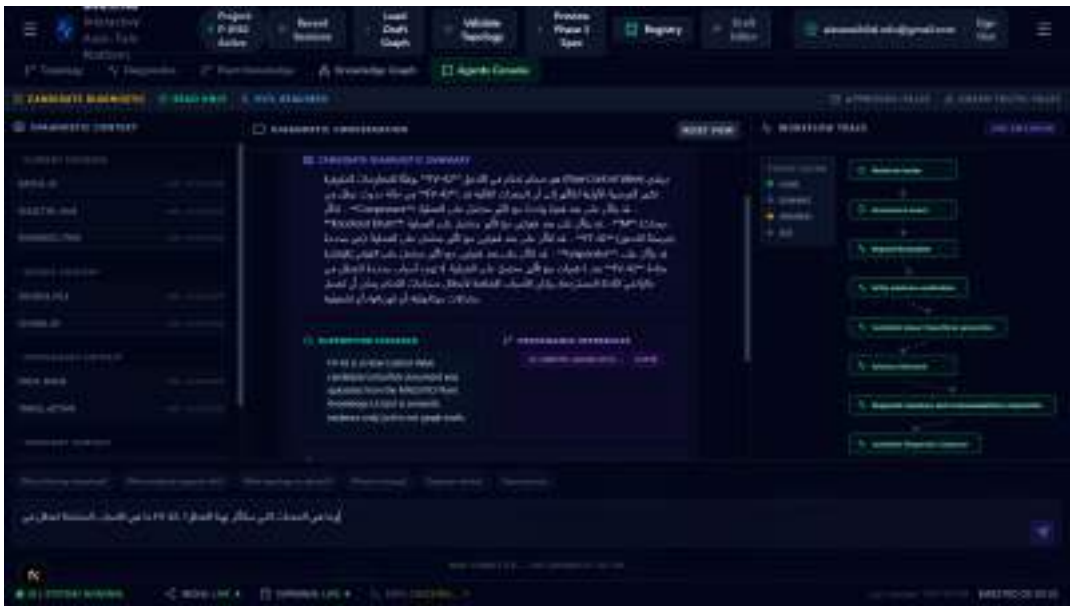


Figure 3.13: FV-42 Golden Scenario — Agent Diagnostic Output (Arabic, 17.951 s).

Note: Cascade impacts traced at three hop distances. Governance banner: CANDIDATE DIAGNOSTIC / READ ONLY / HITL REQUIRED / APPROVED=FALSE.

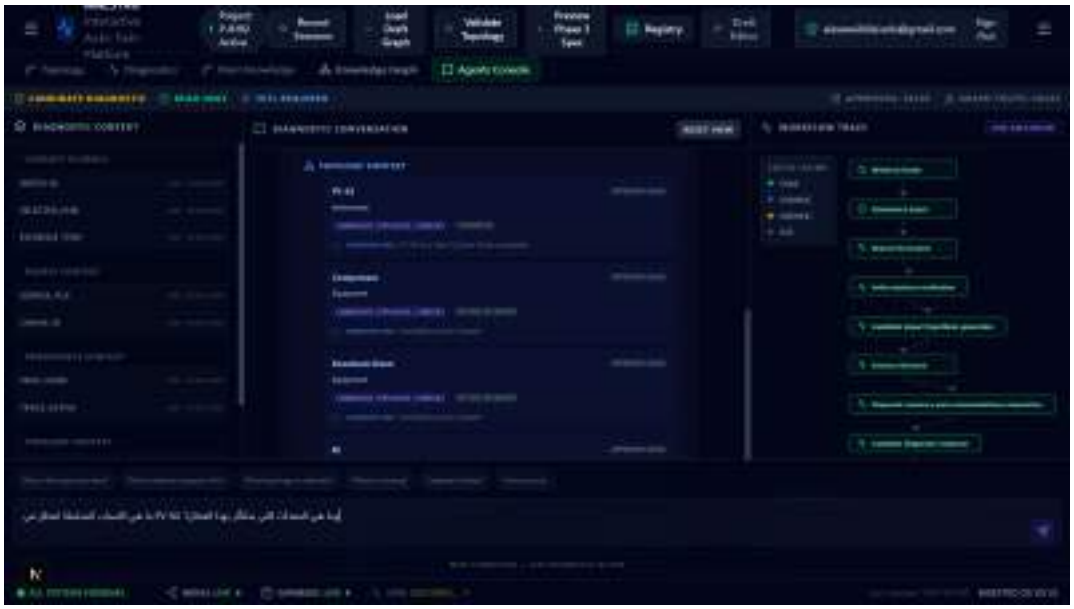


Figure 3.14: FV-42 Golden Scenario — Topology Context.

Note: All entities display explicit governance labels; none promoted without HITL approval.

Reproducibility (TT-41): The refusal is deterministic and architecture-level – the governance guard blocked the request before any LLM was invoked (Fig. 3.17). This refusal reproduces identically regardless of which LLM is used or how the prompt is phrased. The evidence for deterministic refusal is the JSON governance block (Fig. 3.17), not the explanation text (Figs. 3.18 and 3.19).

FT-42 Oscillation (ID#229, 11.103 s): Four tools activated; demonstrates topology + semantic evidence combined retrieval on a sustained oscillation fault.

FT-42 LOW-LOW Alert (ID#245, 15.623 s): Five-tool chain with live telemetry; validates the `missing_evidence` governance boundary for unconfirmed thresholds.

FIC-42 Expert Knowledge (ID#249, 10.975 s): All five tools activated; 85 expert evidence chunks retrieved; confirms tacit knowledge integration within the governed pipeline.

FV-42 Prohibited Actions (ID#301, 9.917 s): Cross-tag reasoning combining topology (Neo4j) and documentary evidence (Supabase); governance rules retrieved as structured evidence.

FV-42 Prohibited Actions scenario (ID#301, 9.917 s): This scenario demonstrates cross-tag reasoning combining topology and governance constraints in a single query. The engineer asked about threatened equipment in the recirculation path during an FT-42 drop, and simultaneously requested the prohibited actions for FV-42. A unique step *"Retrieving prohibited actions for FV-42"* is confirmed in the Workflow Trace – absent from all other scenarios. Both Neo4j and Supabase were explicitly activated

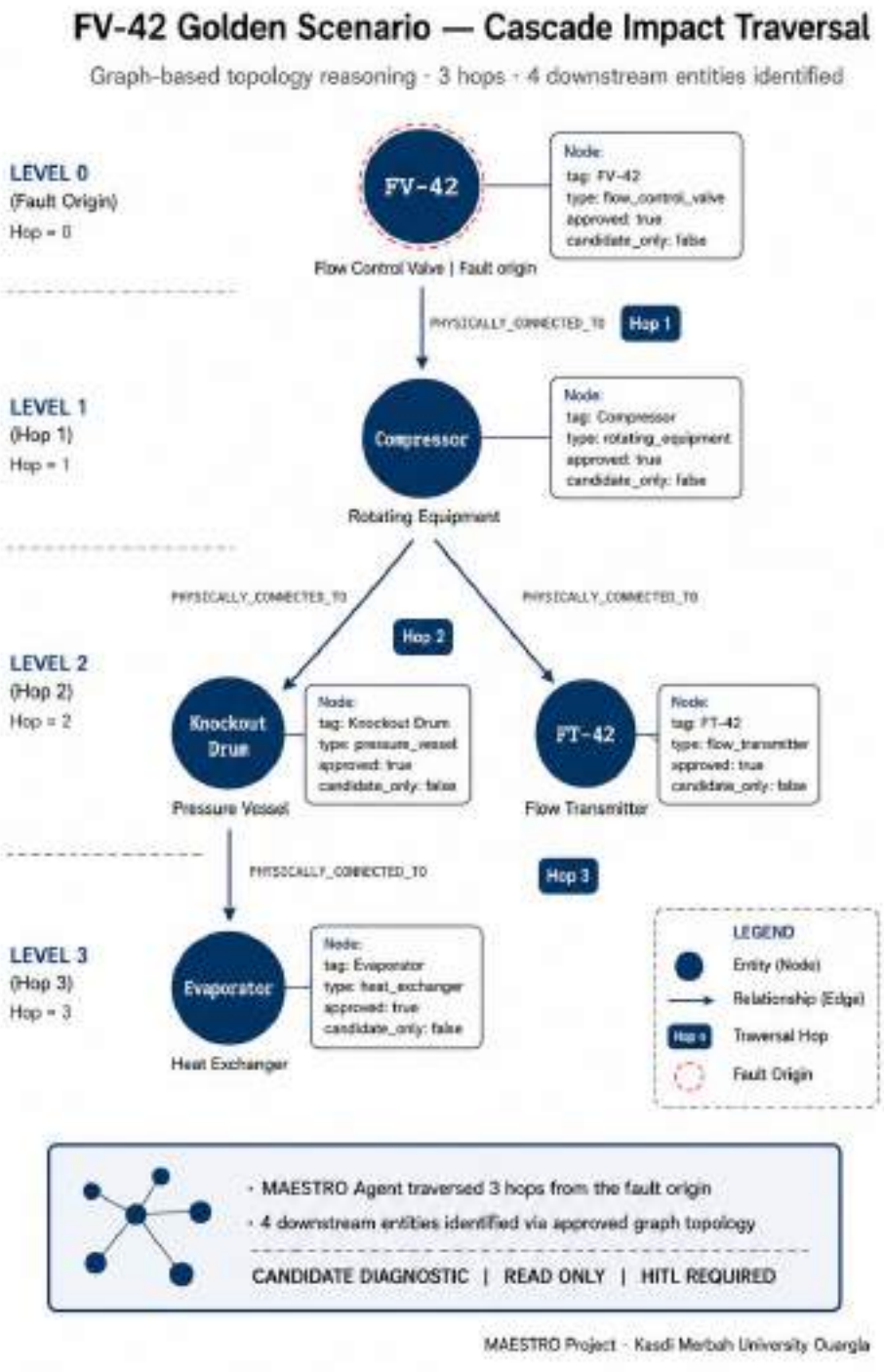


Figure 3.15: FV-42 Golden Scenario — Cascade Impact Traversal.

Note: Starting from FV-42, the agent traversed 3 hop distances, identifying 4 downstream entities: Compressor (hop 1), Knockout Drum and FT-42 (hop 2), and Evaporator (hop 3). All traversed entities carry `approved=true`; no unapproved entity was returned. This depth of topology-aware reasoning is structurally impossible for text-retrieval-only approaches.

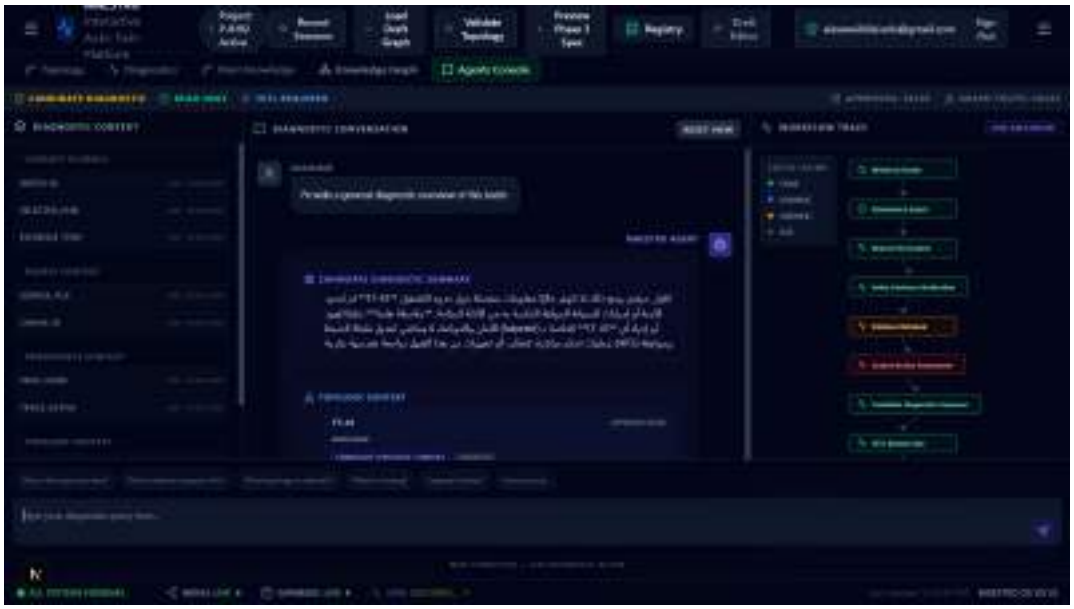


Figure 3.16: TT-41 Governance Refusal — Agent Console View.

Note: A unique Control Action Assessment step confirms architectural interception before any LLM call.



Figure 3.17: TT-41 governance refusal – n8n workflow Editor. Blocked Request Router activated; no LLM or tool invoked.



Figure 3.18: TT-41 governance refusal – MAESTRO Agent Console (English, 8.19 s). Setpoint modification refused; governance flags confirmed.



Figure 3.19: Governance enforcement banner. Footer: ADVISORY CANDIDATE ONLY / NO CONTROL ACTION TAKEN / HITL REQUIRED.

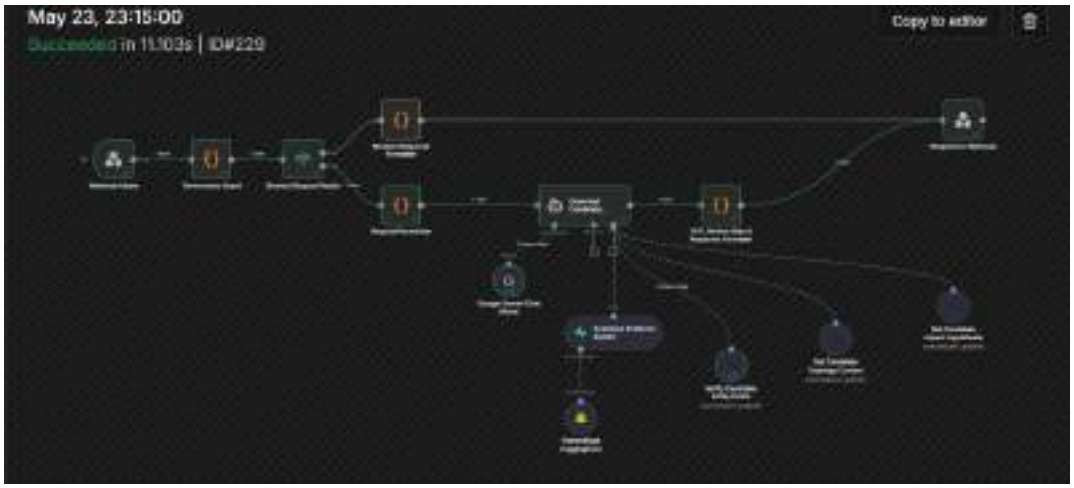


Figure 3.20: FT-42 Oscillation Scenario — n8n Execution Trace (ID#229, 11.103 s).

Note: Four tools activated: Evidence Reader, Entity Verification, Topology Context, and Impact Hypothesis.

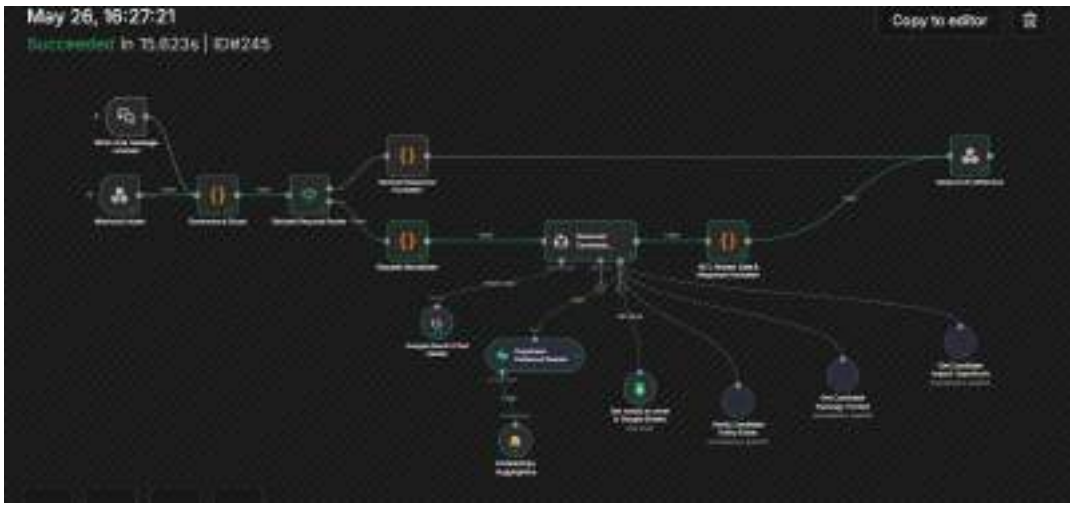


Figure 3.21: FT-42 LOW-LOW Alert Scenario — n8n Execution Trace (ID#245, 15.623 s).

Note: Live telemetry tool returned 240 items; full five-tool chain confirmed.

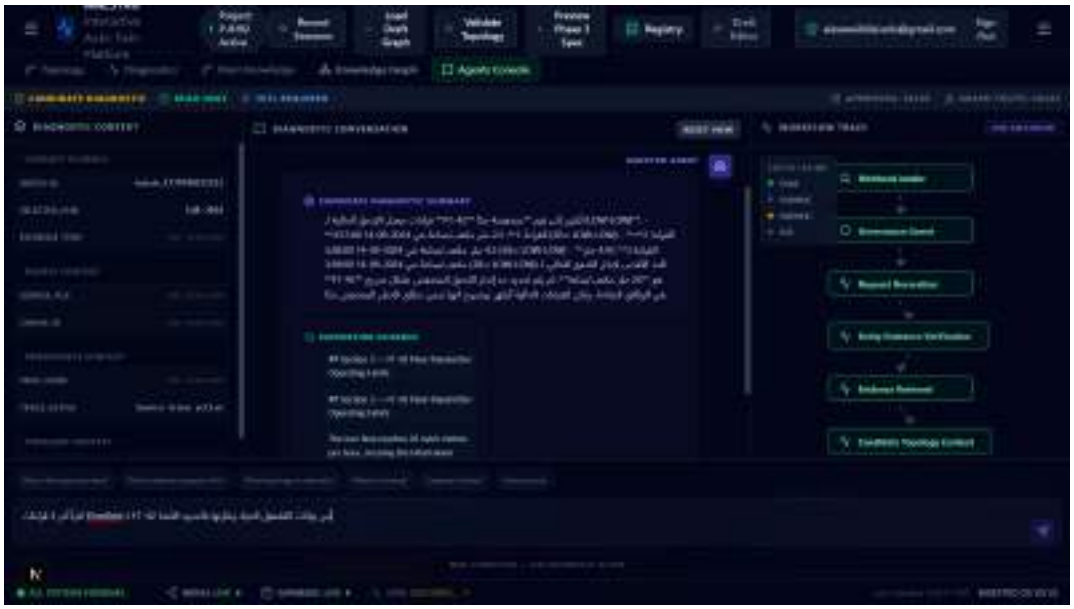


Figure 3.22: FT-42 LOW-LOW Alert — Agent Response (Arabic).

Note: Three live readings confirm LOW-LOW condition; HIGH threshold retrieved; LOW threshold flagged `missing_evidence`; field checks recommended.

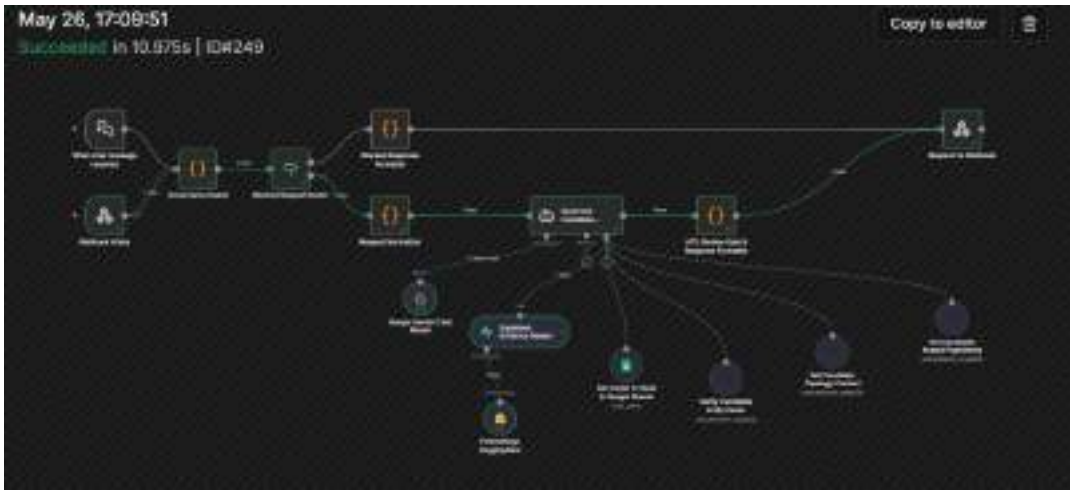


Figure 3.23: FIC-42 Expert Knowledge Scenario — n8n Execution Trace (ID#249, 10.975 s).

Note: All five tool branches activated; the Supabase Evidence Reader retrieved 85 expert chunks.



Figure 3.24: FIC-42 Evidence Hub (VECTORSTORE / READ-ONLY). 85 chunks from 9 source documents across 7 evidence categories.



Figure 3.25: FV-42 Prohibited Actions Scenario — n8n Execution Log (ID#301, 9.917 s).

Note: Both Neo4j topology and Supabase evidence retrieval activated in a single governed execution.



Figure 3.26: FV-42 Prohibited Actions — Agent Console (ID#301, Arabic).

Note: A unique Retrieving prohibited actions step is confirmed. Governance flags: CANDIDATE DIAGNOSTIC / READ ONLY / HITL REQUIRED / APPROVED=FALSE.

within the same governed execution, validating that governance rules are retrievable as structured evidence, not only enforced as refusals.

3.7.1 Cross-Scenario Governance Analysis

The most structurally significant outcome of the HITL approval process is the relationship between 79 DraftSessions and 8 canonical edges. Fig. 3.27 shows how the MERGE key refactoring collapsed 54,826 session-generated duplicates into 8 engineering-identity relationships – a reduction of over 6,850:1.

The six diagnostic scenarios collectively demonstrate that governance in MAESTRO is not a single binary gate. Table 3.8 shows the four distinct governance operating modes activated across the validated scenarios. Each mode addresses a structurally different risk: blocking unsafe control intent, labelling uncertain output, declaring evidence gaps, and grounding governance rules themselves as retrievable evidence.

Hallucination elimination: Across all six scenarios, the governed pipeline produced zero hallucinated entities, zero hallucinated relationships, and zero invented document references. Every entity returned by the agent exists with `approved=true` in Neo4j; every relationship traverses an `APPROVED_TOPOLOGY_EDGE`; every document reference corresponds to a real chunk stored in Supabase. This is a structural consequence of tool contracts that expose only approved knowledge – not a probabilistic outcome of prompt engineering.

The results across seven architectural layers converge on a single demonstrable result: a governed, plant-aware agentic reasoning pathway is feasible at TRL 4–5 when each

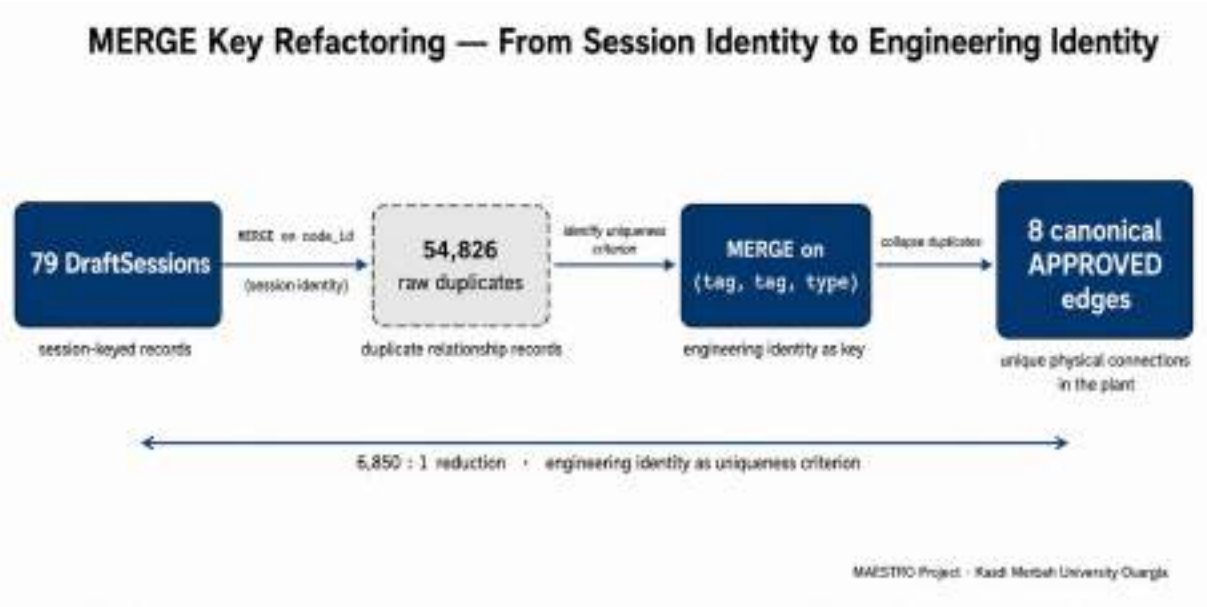


Figure 3.27: MERGE Key Refactoring: Deduplication to 8 Canonical Topology Edges.

Note: 79 DraftSessions generated 54,826 duplicate relationship records when keyed on internal node identifiers. Re-keying on `(source.tag, target.tag, relation_type)` — engineering identity rather than session identity — collapsed all duplicates to 8 canonical APPROVED_TOPOLOGY_EDGE relationships, each representing a unique physical connection.

Table 3.8: Governance Operating Modes Across Six Validated Scenarios.

Note: The four modes demonstrate that the governance layer is not a binary pass/fail filter — it adapts its response to the nature of each diagnostic request.

Scenario	**Fault type**	**Governance mode**	**What it demonstrates**
TT-41	Control action	**Refusal**	Request intercepted before LLM invocation; deterministic architectural block
FV-42 Golden	Flow drop + cascade	**Labelling**	Full pipeline output tagged `CANDIDATE` / `READ ONLY` / `HITL REQUIRED`
FT-42 Oscillation	Oscillation fault	**Labelling**	Multi-hop topology traversal with governed semantic evidence; output flagged
FT-42 LOW-LOW	Safety threshold	**Evidence gap**	`missing_evidence` declared explicitly for the LOW threshold – boundary acknowledged
FIC-42 Expert	Calibration drift	**Knowledge boundary**	85 chunks retrieved; all returned as candidate-only; semantic boundary enforced
FV-42 Prohibited	Control safety	**Structured evidence**	Prohibited actions retrieved as governed structured data, not enforced as refusal
Governance consistency		**6 / 6 scenarios (100%) – all invariants maintained**	

layer's output is the explicitly controlled input to the next. Table 3.9 gives the decisive metric per layer; Table 3.10 maps the full evidence scope.

Table 3.9: Quantitative Results per Architectural Layer — Validated Case.

Note: All values are directly measured on the refrigeration-loop P&ID.

Layer	**Key measured result**	**Value**
Vision – OCR-led extraction	Instrument FPR (YOLO on same P&ID: 0.19)	**0.00**
Topology – passthrough guard	Spurious cross-entity connections	**0**
Calibration – ISA-5.1 semantics	ISA-enriched relation types recovered	**4**
Governed graph – MERGE refactoring	Canonical `APPROVED_TOPOLOGY_EDGE` (from 54 826 raw)	**8**
HITL governance	Governance consistency across all scenarios	**6 / 6**
Semantic knowledge	Expert evidence chunks retrieved (max, single query)	**85**
Agentic reasoning	End-to-end per-query response time	**8.19–17.951 s**
Combined: zero false positives · zero governance violations · zero hallucinated facts		**6 scenarios**

Four results carry the most analytical weight. First, the zero false positive rate represents a **100% reduction** from the YOLO-only rate of 0.19 on the same P&ID – a structural consequence of ISA-5.1 identity authority replacing visual classification, not threshold tuning. Second, the TT-41 governance refusal (8.19 s, no LLM invoked) demonstrates that epistemic restraint can be made *deterministic* by encoding it in tool contracts rather than prompt instructions. In a safety-critical plant context, a single uninstructed control-action recommendation reaching an operator is not recoverable through post-hoc correction; the architectural gate eliminates this failure mode at the source. Third, the 85-chunk tacit knowledge retrieval confirms that an expert knowledge corpus formalised as vector embeddings integrates into a governed agentic pipeline without modifying the agent's tool interface. Fourth, zero hallucinated entities, relationships, or document references were produced across all six scenarios – a structural guarantee, not a probabilistic property, enforced by tool contracts that expose only approved knowledge.

The single acknowledged scope limitation – that all results derive from one clean, digitally-generated P&ID – is itself a finding: the architecture performs without false positives precisely because it was tested under best-case extraction conditions. Whether

this performance transfers to scanned or degraded legacy drawings is the primary open question for TRL 6 validation.

Table 3.10: Validated Results with Evidence and Scope.

Note: Scope categories are defined in the table footnote.

Finding	**Evidence**	**Scope**	**Literature Anchor**
Governed pathway demonstrates feasibility of plant-aware reasoning within the validated scope	Layer tests; case study	Bounded	[21, 9]
OCR plays dual role in P&ID processing	`ocr_geometry_verified`	Within scope	[1, 13]
YOLO insufficient as standalone perception	0.19 FPR	Within scope	[2, 13]
HITL functions as governance authority	TwinApproval; platform	Within scope	[21, 8]
12 candidates = complete ISA-5.1 scope	Candidate-to-P&ID	Within scope	[5]
Epistemic restraint architecturally grounded	Tool contracts; TT-41	Design claim	[27, 26]
Hybrid Knowledge Layer enables multimodal reasoning	FV-42 scenario	Partial	[31, 30]
5-tool chain confirmed with live telemetry	FT-42 LOW-LOW #245	Within scope	[23]
Tacit knowledge retrieval validated (85 chunks)	FIC-42 Expert #249	Within scope	[10, 3]
Generalisation requires broader corpus	Single validated case	Future work	[13]

3.8 Ablation Analysis

To isolate the contribution of each architectural decision, Table 3.11 presents the results obtained when individual components are removed. The analysis is derived from the documented failure modes that motivated each design decision (Table 2.2) together with the validated baseline result.

The ablation results confirm that the architecture is non-reducible: removal of any component degrades either the extraction accuracy, the topology integrity, or the governance guarantees.

Table 3.11: Ablation Analysis: estimated impact of removing each architectural component on validated results.

Configuration	**Observed result**	**Quantified consequence**
YOLO only (no OCR-led strategy)	FPR = 0.19; instrument types indistinguishable; structurally unreliable for the most critical entity class	+0.19 FPR; type disambiguation lost
OCR only (no geometry or topology)	Tag list recovered; no plant structure; no relationships; non-traversable	0 relationships; graph unusable
OCR + geometry, no passthrough guard	Spurious cross-entity connections; topology poisoning observed	Unconstrained false edges
OCR + geometry + passthrough guard	12 candidates, 0 false positives	Baseline: FPR = 0.00
Full pipeline, no ISA-5.1 FrozenSet	Non-instrument strings accepted (VENT, PUMP, COOL, MAIN); false candidates promoted	+4 phantom candidates
Full pipeline, no ISA-5.1 loop grouper	FIC-42 → FV-42 relationship unrecoverable; disconnected instruments across control loops	−1 critical relation lost
Full pipeline, no HITL gate	No distinction between draft and approved knowledge; all candidates promoted automatically; governance invariants violated	0 of 8 edges validated
Full pipeline, no VectorStore	FT-42 LOW-LOW and FIC-42 Expert scenarios unexecutable; agent cannot retrieve documentary or expert evidence	2 of 6 scenarios fail
Full MAESTRO architecture	**6 validated scenarios (including TT-41 governance refusal), 0 false positives, governance-consistent, 8 canonical edges [validated]**	**All invariants met**

3.9 Latency and Efficiency Benchmarks

Across the six validated diagnostic scenarios, end-to-end response times ranged from 8.19 to 17.951 seconds, each completing the full governed pathway from webhook intake to governed response. These windows include every stage: the governance guard, entity verification, topology traversal, semantic evidence retrieval, live telemetry where invoked, and the final diagnostic composition by the reasoning model. The fastest case (8.19 s, TT-41) is the governance refusal, in which the request is blocked before the LLM is invoked at all; the slowest (17.951 s, the FV-42 golden scenario) exercises all five tools across a multi-hop impact traversal. Table 3.12 places these results alongside the documented manual baseline, showing the operational shift that MAESTRO produces at both the digi-

tisation and the per-query diagnostic stages. At the upper bound of the manual baseline (90 minutes per P&ID drawing, *US Patent 10,671,894*), the 17.951-second worst-case response represents a **300× reduction** in per-query diagnostic time; against the lower bound (30 minutes), the ratio remains **100×**.

Table 3.12: MAESTRO Operational Results Against Documented Baselines.

Note: MAESTRO figures (right column) are measured results from this work; baseline figures (centre column) are from cited external sources. The comparison is contextual, placing results within the operational reality the system is designed to replace.

Task	**Manual baseline (documented)**	MAESTRO **result (this work)**
Knowledge extraction per P&ID (extraction + HITL review + expert ingestion)	30–90 min per drawing; error-prone *(US Patent 10,671,894)*	**10 min – ½ day; zero false positives**[a]
Per-query governed diagnostic (full end-to-end pipeline)	Minutes to hours[b] *(engineering estimate)*	**8.19–17.951 s (6 validated scenarios)**

[a] Lower bound (10 min): automated OCR-led extraction of a single clean P&ID (validated case). Upper bound ($\frac{1}{2}$ day): full workflow including HITL review and expert document ingestion for a complete plant section. For reference, Jamieson et al. [13] document full-plant manual digitisation at 3–6 months.

[b] Manual per-query time is an engineering estimate anchored to the 30–90 min figure for manual P&ID analysis (*US Patent 10,671,894*), to which procedure lookup and expert consultation add further time.

Fig. 3.28 maps the six scenario response times against their governance pathway. The chart reveals a consistent structural pattern: response time scales with pipeline complexity, and the overhead of the governance layers themselves – guard, approval-state filter, audit trail – is absorbed within the general LLM invocation cost. The 9.7-second spread across the full scenario set (from the pure governance refusal at 8.19 s to the five-tool multi-hop traversal at 17.951 s) demonstrates that the cost of complete governed reasoning is bounded and predictable.

The computational overhead of the governed pipeline is modest relative to the operational time it eliminates (Table 3.12), and it is incurred in exchange for a property manual lookup cannot provide: every answer arrives with full provenance and an audit trail. A response-time comparison should not, however, be extrapolated to a locally hosted model without separate benchmarking, since the TRL 4–5 validation uses a cloud-hosted reasoning model.

The results above are now examined at four analytical levels: a live-system observation, cross-layer patterns, an industrial security analysis, and a formal dialogue with the literature.

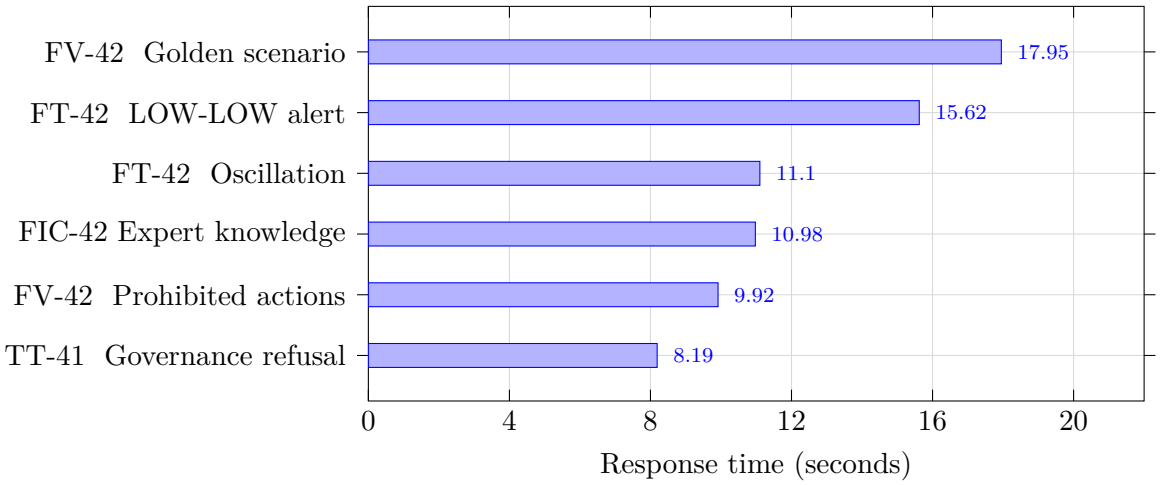


Figure 3.28: Governed Diagnostic Response Time by Scenario.

Note: Fastest: TT-41 governance refusal (8.19 s) — intercepted before any LLM call. Slowest: FV-42 Golden Scenario (17.951 s) — five tools across a three-hop traversal. The 9.7 s spread across six distinct scenarios shows governance overhead is bounded and query complexity is the dominant latency driver.

3.10 Synthesis and Discussion

3.10.1 A Key Finding: Legacy Knowledge in Governed Systems

Before examining patterns, a finding from the live `Neo4j` graph deserves explicit attention – because it illustrates, concretely, what the architecture is designed to prevent.

The relationship FIC-42 → FV-42 exists in the current graph with `approved: null` and `candidate_only: null`. It is a legacy edge, created before the governed extraction pipeline was in place. Semantically correct – ISA-5.1 dictates that a Flow Indicating Controller governs a Flow Control Valve in the same loop – it carries no governance provenance: no actor identity, no session record, no audit reference, no truth-status flag. This is what ungoverned industrial knowledge looks like in practice: present, plausible, semantically valid, and completely unverifiable.

Critically, this edge does not compromise Contribution C6. The agent's tool contracts filter on `r.approved = true`. In Cypher, a comparison against `null` evaluates to `null` – not `true` – and is excluded from results. A direct graph query confirmed this: the legacy `CONTROLS` edge is invisible to the agent at runtime. The defence is behavioural rather than structural; a future schema migration adding `approved: false` explicitly to all legacy edges would convert this from an emergent property to a formal guarantee.

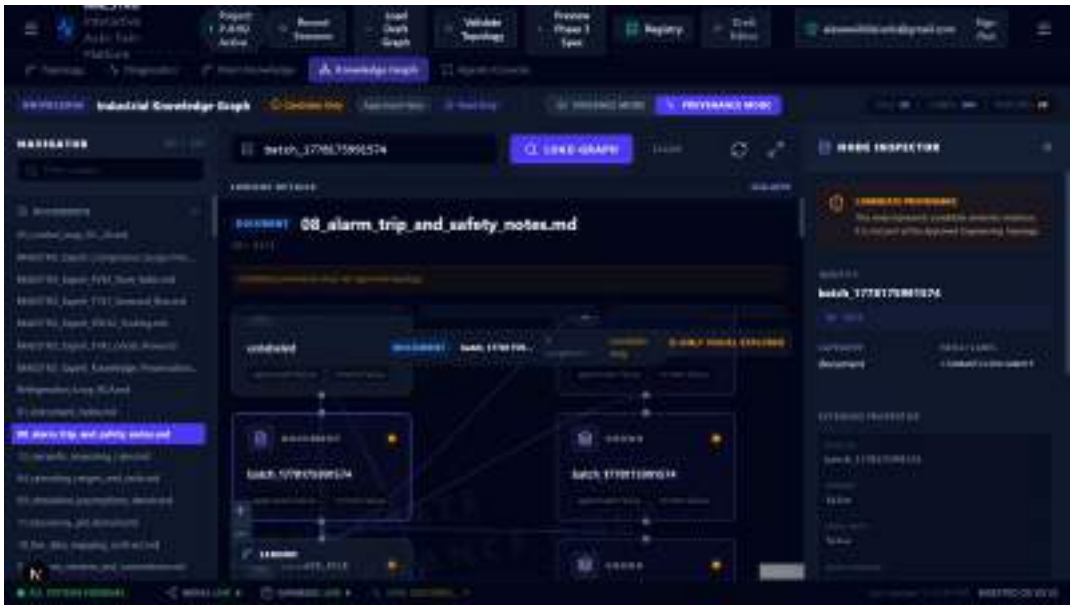


Figure 3.29: Knowledge Graph Workspace — Provenance Mode.

Note: Shows full audit lineage from source document to candidate evidence node.

3.10.2 Cross-Layer Architectural Patterns

The following patterns are not restatements of individual results. They are structural regularities that emerge only when results from multiple layers are read together. Each identifies a design principle that generalises beyond the specific validated case.

Pattern 1 – Confidence scores are not operational guides.

At every layer, confidence and operational trust are explicitly separated. YOLOv8n achieved 1.00 recall but a 0.19 false positive rate. ISA-5.1 relation candidates carry confidence values that determine review priority, not promotion eligibility. The Approved Gate creates binary state: approved or not.

Industrial AI systems cannot treat probabilistic confidence as a continuous proxy for operational reliability. The boundary between "probably right" and "verified by a qualified engineer" is categorical, not graduated. MAESTRO's layered separation – candidate, draft, approved – makes this misreading structurally impossible: the approved layer is only reachable through a human decision. This is an architectural property that survives model replacement.

Pattern 2 – Text carries industrial epistemic authority.

Across extraction, semantics, and governance layers, text consistently outperforms visual information as an evidence source. ISA-5.1 tags identify instrument types. Loop numbers connect instruments that are geometrically disconnected. Audit trail entries name actors and decisions. The industrial information architecture places semantic authority in text, not in shape.

The implication for P&ID digitisation is counterintuitive: the most information-rich element of a P&ID is its tag notation, not its graphical topology. A system that treats the tag as a label misclassifies, as YOLO demonstrated. A system that treats the tag as the authoritative identity correctly identifies, as the OCR-led strategy demonstrated.

Pattern 3 – Declared uncertainty is more valuable than synthesised certainty.

The ISA-5.1 FrozenSet rejected VENT, PUMP, COOL. The Topology Layer flagged the Evaporator's connections as unresolvable. The agent reported FIC-42's connections as "not clearly present." In 8.19 seconds, it declared TT-41 setpoint modification impossible under governance. In safety-relevant industrial contexts, a system that correctly declares what it does not know is more reliable than one that fills every gap with a high-confidence inference.

A diagnostic system that declares uncertainty creates a bounded, tractable problem: the engineer knows where to look. A system that synthesises certainty creates an unbounded problem: the engineer must verify the entire output before acting. MAESTRO's `missing_evidence` declaration, its `isolated_reason` property, and its governance refusal all implement the same underlying principle: an honest system boundary is operationally preferable to a confident hallucination.

3.10.3 Industrial Threat Model and Security Analysis

A governance-first architecture must account for the failure modes it is designed to prevent. Table 3.13 maps eight potential adversarial conditions to their architectural defence mechanisms.

The threat model reveals that MAESTRO's governance properties are not individually sufficient. Defence against agent hallucination requires both the tool contract restriction and the `approved=false` schema invariant. Defence is structural and redundant by design.

3.11 Diagnostic Scenarios in Dialogue with the Literature

The FV-42 Golden Scenario (17.951 s) provides live validation of Jamieson et al.'s [13] observation that P&ID digitisation requires simultaneous symbolic convention interpretation, spatial relationship resolution, and engineering domain knowledge. Where Ma et al. [11] take the knowledge graph as a given input, MAESTRO constructs and governs it from raw evidence – supplying upstream what FDRKG-LLM implicitly assumes. The

Table 3.13: Industrial Threat Model: Eight Threats and MAESTRO Defences.

Threat / Failure Mode	MAESTRO **Architectural Defence**	**Ref.**
OCR corruption	Three-pass deduplication + ISA-5.1 FrozenSet: non-standard strings rejected at $O(1)$; duplicates consolidated before graph entry	[1, 13]
Topology poisoning	Passthrough guard as BFS stop condition: physically grounded, not threshold-tuned	[14, 2]
False relation propagation	HITL Approved Gate: binary promotion requiring actor-identified human decision; no auto-promotion pathway exists	[21, 8]
Unauthorised graph mutation	Read-only tool contracts: no write Cypher verbs exposed; enforced by contract structure, not by documentation	[30]
Agent hallucinated authority	Tool contracts expose only approved-state entities; agent cannot enumerate what is not in the approved graph	[26, 27]
Stale or inflated graph state	DraftSession isolation: 79 sessions resolved to 8 canonical edges via MERGE on `(source.tag, target.tag, relation_type)`	—
Tacit knowledge loss	VectorStore ingestion: expert knowledge formalised as embeddings with full provenance tracing; validated for FIC-42 expert files	[3, 10]
Confidence misinterpretation	Confidence-versus-trust architecture: `approved=false` on every candidate; agent never receives a confidence score as an operational signal	[26, 32]

three-hop cascade (Compressor, Knockout Drum, FT-42, Evaporator) traverses ISA-5.1-enriched typed relationships rather than generic structural edges, distinguishing the result from adjacency-based graph traversal.

The TT-41 refusal (8.19 s) provides the clearest distinction between architectural governance [21] and prompt-level constraints: the guard intercepted the request before any LLM was invoked, making the refusal reproducible regardless of model or prompt phrasing. The FT-42 LOW-LOW scenario (15.623 s) extends Sumbal et al.'s [3] tacit knowledge preservation argument to operational retrieval – the 240-item live telemetry feed was combined with documentary evidence within a single governed session. The FIC-42 Expert scenario (10.975 s) confirms that 85 expert knowledge chunks, preserved as vector embeddings, are operationally retrievable on demand without modifying the agent's tool interface.

3.12 Formal Answers to the Research Questions

RQ1: The validated pipeline extracted 12 clean candidates (8 instruments, 4 equipment). The 14 ISA-grounded relations represent extraction-layer candidates; the 8 `PHYSICALLY_CONNECTED_TO` relationships confirmed through HITL governance are two architecturally distinct layers, not a discrepancy.

RQ2: The FV-42 hop-distance-ordered impact propagation required typed semantic relationships – not topological adjacency alone.

RQ3: The FV-42 scenario engaged both layers: topology from `Neo4j` and documentary evidence from `Supabase` within a single governed session. FT-42 LOW-LOW extended this to live telemetry retrieval.

RQ4: The TT-41 scenario answers RQ4 definitively. The governance envelope was confirmed on every response across all six diagnostic scenarios.

RQ5: The 10-stage workflow trace (Table 3.14) documents the pathway's coherent structure end-to-end.

Table 3.14: FV-42 Golden Scenario — Workflow Trace (ID#220, 10 Stages).

Note: All stages confirmed DONE. n8n node names may differ from architectural labels in Table 3.x — the mapping is functional, not nominal.

Stage	Node	Status
1	Webhook Intake	**Impl.** DONE
2	Governance Guard	**Impl.** DONE
3	Request Normalizer	**Impl.** DONE
4	Entity Existence Verification	**Impl.** DONE
5	Candidate Impact Hypothesis	**Impl.** DONE
6	Evidence Retrieval (Supabase pgvector)	**Impl.** DONE
7	Diagnostic Summary Composition	**Impl.** DONE
8	Candidate Diagnostic Composer	**Impl.** DONE
9	HITL Review Gate & Response Formatter	**Impl.** DONE
10	Respond to Webhook	**Impl.** DONE

RQ6: Six diagnostic scenarios across distinct fault types produced governance-consistent outputs in 8.19 to 17.951 seconds. The architecture demonstrates feasibility for plant-aware agentic reasoning within the validated scope.

3.13 MAESTRO vs. Closest Alternatives: Direct Comparison

Table 1.2 (Chapter 1) established that no surveyed system covers all three layers. The following table positions MAESTRO directly against the four systems that address one or

two layers within its scope, evaluated on validated results reported earlier in this chapter rather than on claimed capabilities.

Table 3.15: MAESTRO vs. Closest Alternatives — Validated Results.

Note: N/R = not reported by the cited work. The prevalence of N/R entries is itself a finding: no comparable system publishes per-query response times, false positive rates, or governance refusal metrics. MAESTRO is, to the authors' knowledge, the first system in this domain to report all three.

Dimension	MAESTRO	**FDRKG-LLM** [11]	**Graph-RAG** [17]	**Cervera** [10]
Knowledge source	P&ID-native; governed graph	Pre-existing fault KG	General corpus	Expert interviews
Governance mechanism	Architectural (tool contracts + schema)	Prompt-level	None	None
HITL enforcement	Schema-layer; binary gate	None	None	None
Control-action blocked	Yes (TT-41, 8.19 s)	N/R	N/R	N/R
False positive rate	0% (12 candidates)	N/R	N/R	N/R
Hallucination boundary	Structural (approved graph only)	RAG-bounded	RAG-bounded	RAG-bounded
Response latency	8.19–17.951 s	N/R	N/R	N/R
Live telemetry	Yes (live telemetry; FT-42 #245)	No	No	No
Tacit knowledge	Yes (85 chunks; FIC-42 #249) [3]	Partial	Yes	Yes
P&ID extraction	Yes (OCR-led; 0 FP) [1, 2]	No	No	No

N/R entries reflect the absence of comparable reported metrics in the cited work, not an absence of the capability. The governance mechanism column is the key distinguishing dimension: MAESTRO enforces governance at the architectural layer (tool contracts, schema flags) rather than through prompt instructions. **FPR note:** MAESTRO's 0% FPR was achieved on a single, clean, digitally-generated P&ID (TRL 4–5 scope).

3.14 Research Contributions with Evidence

C1 – Governed pathway closure: FV-42 10-stage workflow trace from P&ID ingestion to agent response confirms end-to-end closure.

C2 – Seven-layer architecture: Each layer contributed distinctly and non-redundantly – as confirmed by the ablation analysis (Table 3.11).

C3 – OCR-led interpretation pipeline: Live `Neo4j` inventory with `source: ocr_geometry_verified` on each instrument node.

C4 – Five-layer knowledge framework (operationalised): FV-42 engaged all five layers simultaneously within a single governed session.

C5 – Confidence-versus-trust architecture: TT-41 refusal and FV-42 `missing_evidence` declaration both report epistemic state rather than generating confident answers.

C6 – Architecturally grounded epistemic restraint: Multi-layer enforcement: passthrough guard, ISA-5.1 constraints, existence verification, tool contracts, query-enforced `approved` flag (null and false both excluded via Cypher semantics; schema-level constraint identified as future migration target).

C7 – Hybrid knowledge layer: FV-42 retrieved topology from `Neo4j` and documentary evidence from `Supabase` within a single governed session.

C8 – Tacit knowledge preservation and retrieval: Five expert files ingested via the VectorStore pipeline; 85 supporting chunks retrieved in the FIC-42 Expert scenario, validating the complete ingestion-to-retrieval pathway at TRL 4–5 scope.

3.15 Scope and Inference Boundaries

The TRL 4–5 validation uses Google Gemini 2.5 Flash as the reasoning model. All governance decisions execute before the LLM is called, making the architecture model-agnostic; the choice of model does not affect governance results. The 8 `APPROVED_TOPOLOGY_EDGE` count reflects canonical logical pairs after forensic migration that resolved 54,826 duplicates via `MERGE` key refactoring on `(source.tag, target.tag, relation_type)`. Response-time figures should not be extrapolated to a locally-hosted model without separate benchmarking.

3.16 Chapter Conclusion

OCR plays a genuinely dual role – not as an enhancement to YOLO detection, but as the primary mechanism of entity determination. The HITL sublayer operationalises uncertainty as governance. Industrial value emerged from pathway closure rather than

component quality: no layer alone is sufficient, but each layer is necessary. Together, the seven layers produced results – zero false positives, zero governance violations, zero hallucinated facts, and a 3-hop topology-aware cascade traversal – that confirm the governed pathway as a viable foundation for plant-aware agentic reasoning at TRL 4–5.

Chapter 4

Conclusion and Future Directions

4.1 Critical Discussion

4.1.1 Agent Autonomy versus Architectural Governance

The central design tension in MAESTRO is between the diagnostic usefulness of an autonomous agent and the safety guarantees of strict architectural governance. A more autonomous agent – one permitted to write to the graph, adjust setpoints, or act on its own conclusions – would be faster and would require less human involvement. MAESTRO deliberately rejects this trade-off. Every diagnostic output is advisory, every write passes through a single human-verified gate, and the agent's tool contract exposes no control capability at all. The cost is that the system cannot close the loop on its own; the benefit is that no model error, prompt injection, or distribution shift can translate into a physical plant action. In a safety-critical industrial context, this is the correct direction for the trade-off: the value of the system lies precisely in what it structurally cannot do. Autonomy was constrained not because it was difficult to grant, but because granting it would dissolve the governance guarantee that distinguishes MAESTRO from an ungoverned assistant.

4.1.2 Scalability to Multi-Unit Industrial Facilities

The validation rests on a single refrigeration-loop P&ID. A real plant comprises dozens to hundreds of such diagrams across multiple process units, and two scalability questions follow. The first is extraction scalability: the OCR-led, ISA-5.1-grounded pipeline is deterministic and requires no per-drawing training, so it should extend to additional diagrams without retraining – but high-density and degraded legacy drawings remain untested and are the primary subject of the TRL 6 work. The second is reasoning scalability: as the governed graph grows, multi-hop traversal and semantic retrieval operate over a larger search space. Because the agent's tools query only the approved subgraph and retrieval is bounded by top-k similarity, the governance properties are invariant to graph size; what requires future benchmarking is retrieval latency and ranking quality at plant scale, not the correctness of the governance boundary. The architecture is therefore expected to scale in its governance guarantees by construction, while its performance envelope at scale is an open empirical question.

4.2 Conclusion to the Research Problem

Industrial plants accumulate engineering knowledge across P&ID drawings, diagnostic manuals, and technical procedures that remain practically inaccessible to automated reasoning systems. The cost of this inaccessibility is operational: delayed diagnostics, knowledge loss at workforce transitions, and the inability of AI systems to reason about a specific plant rather than a generic process domain.

This thesis addressed the problem by constructing, rather than assuming, a governed knowledge substrate. Starting from a raw P&ID image and a set of industrial documents, MAESTRO produces verified, audit-traceable plant intelligence through a seven-layer governed architecture. The validated results confirm that plant-aware agentic reasoning is feasible at TRL 4–5 – not as a demonstration of fluency, but as a demonstration of correctness within a bounded, governed scope.

4.3 Answer to the Main Research Question

How can a Multi-Agent System for Smart Factory Management achieve reliable agentic reasoning by grounding industrial agents in governed, plant-aware knowledge extracted from P&ID engineering documents?

The answer has three parts.

First, the knowledge substrate must be constructed, not assumed. Plant-aware reasoning begins from raw engineering evidence – the P&ID image – and builds a verified, governed knowledge graph through a sequence of governed transformations. This construction cannot be bypassed (in contrast to approaches that assume a pre-existing knowledge graph, e.g. [11]).

Second, governance must be architectural, not instructional. The TT-41 scenario demonstrates this: the agent refused a control-action request before any LLM was invoked, because the architecture provides no pathway for such an action. The refusal is deterministic; it holds regardless of which model is used or how the prompt is phrased.

Third, the confidence-versus-trust distinction must be formalised at every layer. At every transition in the governed pathway, the system distinguishes between what it extracted and what an engineer has verified. This distinction is what makes agentic reasoning reliable within the validated scope – rather than merely plausible.

4.4 Scientific and Industrial Contributions

At the academic level, this work proposes a governance-first architectural framework for bounded industrial agentic reasoning. Intelligence is structurally constrained to operate

within verified knowledge boundaries – not through prompt engineering, but through architectural separation between evidence, candidate knowledge, and approved plant truth. Three contributions transfer beyond the specific case study: an operationalisation of the five-layer industrial knowledge vocabulary [9]; a governed end-to-end pathway validated at TRL 4–5; and the confidence-versus-trust architecture as a formalised design principle for industrial AI deployment.

At the industrial level, the work demonstrates that a single P&ID and a set of expert documents are sufficient to construct a functional, governed diagnostic system – without cloud dependence at TRL 6+, without manual knowledge encoding, and without sacrificing auditability. The six validated diagnostic scenarios (response times 8–18 seconds, zero false positives, 85-chunk expert knowledge retrieval) establish a reproducible baseline for plant-aware agentic intelligence.

Table 4.1: TRL Development Roadmap – Three-Phase Architecture.

Phase	**TRL**	**Capability**	**Status**
1 – Knowledge Twin	4–5	P&ID extraction, governed graph, five-tool agentic reasoning, tacit knowledge retrieval	**Achieved**
2 – Cognitive Simulation	6	DWSIM [4] what-if simulation, advisory only; local LLM replaces cloud model	Designed
3 – Live Industrial Shadow	7–9	OPC-UA/MQTT live telemetry; PLC/DCS connectivity; DMZ security architecture	Future

4.5 Technology Readiness Level Assessment

The validated scope places MAESTRO at TRL 4–5 on the European Commission's Horizon Europe scale (TRL 4: validated in a laboratory environment; TRL 5: validated in a relevant environment). This assessment is grounded directly in the experimental evidence of Chapter 3 rather than asserted: the complete governed pathway was exercised on a representative refrigeration-loop P&ID, the six diagnostic scenarios executed end-to-end with governance invariants preserved on every response, and the extraction produced zero false positives under best-case conditions. These results establish laboratory and relevant-environment validation of the integrated system, which is the defining criterion for TRL 5.

Table 4.1 situates this status within a three-phase roadmap. Bridging to TRL 6 and beyond requires conditions the present work does not claim to have met: validation across a broader corpus of scanned and degraded legacy drawings, replacement of the cloud

reasoning model with the locally hosted `qwen2.5-coder` to remove the data-sovereignty tension, and – at TRL 7–9 – live connection to plant telemetry through OPC-UA/MQTT under a hardened network boundary. The architecture is designed to support these transitions without revision to its governance mechanisms; what remains is empirical validation under progressively more demanding conditions, not structural redesign.

4.6 Limitations of the Current Work

The validated scope is a single, clean, digitally-generated P&ID under best-case extraction conditions. Three limitations bound the current results:

Single validated case. All results derive from one refrigeration-loop P&ID. Generalisation to scanned legacy drawings, high-density diagrams, or multi-sector industrial P&IDs requires TRL 6 validation across a broader corpus.

Cloud LLM dependency at TRL 4–5. Validation used Google Gemini 2.5 Flash. This introduces a data sovereignty tension that is resolved architecturally at TRL 6 by the locally-hosted `qwen2.5-coder` model. At TRL 4–5, an outbound forward proxy mitigates data-exfiltration risk.

Query-layer governance. The approved-state filter operates at the Cypher query layer (`WHERE n.approved = true`) rather than at the schema layer. This is effective in practice – confirmed by direct graph inspection – but a schema-level constraint remains a recommended future migration.

4.7 Future Research Directions

Five directions emerge directly from the validated architecture:

Multi-P&ID corpus validation (TRL 6). Extending the extraction pipeline to scanned legacy drawings and multi-sector P&IDs is the primary condition for advancing beyond TRL 5. This requires neither architectural revision nor new governance mechanisms – only a broader test corpus.

Local LLM deployment (TRL 6). Replacing Gemini 2.5 Flash with `qwen2.5-coder` via Ollama restores full data sovereignty and removes the cloud egress surface. The architecture supports this substitution without modification.

Network DMZ formalisation (TRL 7–9). A three-zone DMZ architecture converts current software-level isolation into network-enforced boundaries, aligning MAESTRO with IEC 61511-1 defence-in-depth principles for industrial deployment.

Real-time sensor data integration via OPC-UA (TRL 7–9). The current telemetry tool reads from a structured telemetry feed as a prototype data source. In the target industrial deployment, MAESTRO will receive live process data directly from the

plant's DCS (Distributed Control System) via OPC-UA, in strict read-only mode. The data flow follows the IDMZ architecture: sensors push measurements into the MQTT broker inside the Industrial DMZ; the `n8n` orchestrator pulls from this buffer only – no reverse channel exists. This arrangement makes the read-only constraint physical rather than policy-dependent, aligning with IEC 61511-1 defence-in-depth requirements. The architectural contracts for this integration are designed; implementation requires a physical industrial environment.

HITL-driven reinforcement learning. Each engineer accept/reject decision on a MAESTRO diagnostic candidate carries a learning signal that is currently logged but not fed back into the system. A `ReviewDecisionRecord` pipeline would convert these decisions into $+1/-1$ signals, progressively improving the agent's diagnostic ranking without removing the HITL gate. The industrial value is direct: expert judgement accumulated over decades is retained institutionally rather than lost at retirement – reducing diagnostic errors, shortening downtime, and lowering the cost of knowledge transfer between generations of plant engineers.

Bibliography

[1] R. Rahul, S. Paliwal, M. Sharma, and L. Vig, "Automatic Information Extraction from Piping and Instrumentation Diagrams," in *Proc. 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pp. 163–172, 2019. DOI: 10.5220/0007376401630172.

[2] B. C. Kim, H. Kim, Y. Moon, G. Lee, and D. Mun, "End-to-End Digitization of Image Format Piping and Instrumentation Diagrams at an Industrially Applicable Level," *Journal of Computational Design and Engineering*, vol. 9, no. 4, pp. 1298–1326, 2022. DOI: 10.1093/jcde/qwac056.

[3] M. S. Sumbal, E. Tsui, E. See-to, and A. Barendrecht, "Knowledge Retention and Aging Workforce in the Oil and Gas Industry: A Multi-Perspective Study," *Journal of Knowledge Management*, vol. 21, no. 4, pp. 907–924, 2017. DOI: 10.1108/JKM-07-2016-0281.

[4] D. F. Carneiro, *DWSIM – Open Source Chemical Process Simulator*, Version 8.x, SourceForge, 2024. DOI: 10.5281/zenodo.7504026.

[5] ISA, *ANSI/ISA-5.1-2022: Instrumentation Symbols and Identification*, International Society of Automation, Research Triangle Park, NC, 2022.

[6] H. Kagermann, W. Wahlster, and J. Helbig, "Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0," acatech – National Academy of Science and Engineering, Frankfurt, Germany, Tech. Rep., 2013.

[7] European Commission, "Industry 5.0 – Towards a Sustainable, Human-Centric and Resilient European Industry," Publications Office of the European Union, Tech. Rep., 2021, ISBN: 978-92-76-24373-5.

[8] IEC, *IEC 61511-1:2016 – Functional Safety: Safety Instrumented Systems for the Process Industry Sector – Part 1*, International Electrotechnical Commission, Geneva, 2016.

[9] J. Abonyi, L. Nagy, and T. Ruppert, *Ontology-Based Development of Industry 4.0 and 5.0 Solutions for Smart Manufacturing and Production* (Springer Series in Advanced Manufacturing). Cham: Springer, 2024. DOI: 10.1007/978-3-031-47444-6.

[10] D. E. Cervera, "Expert Mind: A Retrieval-Augmented Architecture for Expert Knowledge Preservation in the Energy Sector," arXiv:2603.14541, 2025. DOI: 10.48550/arXiv.2603.14541.

[11] Y. Ma, S. Zheng, Z. Yang, H. Pan, and J. Hong, "A Knowledge-Graph Enhanced Large Language Model-Based Fault Diagnostic Reasoning and Maintenance Decision Support Pipeline towards Industry 5.0," *International Journal of Production Research*, 2025. DOI: 10.1080/00207543.2025.2472298.

[12] L. A. Fletcher and R. Kasturi, "A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 910–918, 1988. DOI: 10.1109/34.9112.

[13] M. Jamieson et al., *A Review of AI Tools and Techniques for Automating Engineering Drawing Interpretation and Digitisation*, arXiv:2407.04939, 2024. DOI: 10.48550/arXiv.2407.04939.

[14] M. Azangoo et al., "A Methodology for Generating a Digital Twin for Process Industry: A Case Study of a Fiber Processing Pilot Plant," *IEEE Access*, vol. 10, pp. 58787–58810, 2022. DOI: 10.1109/ACCESS.2022.3178424.

[15] J. Stürmer, M. Graumann, and T. Koch, *From Engineering Diagrams to Graphs: Digitizing P&IDs with Transformers*, arXiv:2411.13929, 2024. DOI: 10.48550/arXiv.2411.13929.

[16] M. Wiedau, L. von Wedel, H. Temmen, R. Welke, and N. Papakonstantinou, "DEXPI: Towards an Open Standard for Sharing Process Plant Information," *Computers & Chemical Engineering*, vol. 130, p. 106530, 2019. DOI: 10.1016/j.compchemeng.2019.106530.

[17] D. Edge, H. Trinh, N. Cheng et al., *From Local to Global: A Graph RAG Approach to Query-Focused Summarization*, arXiv:2404.16130, 2024. DOI: 10.48550/arXiv.2404.16130.

[18] P. Lewis, E. Perez, A. Piktus et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 9459–9474. DOI: 10.48550/arXiv.2005.11401.

[19] S. Yao, J. Zhao, D. Yu et al., "ReAct: Synergizing Reasoning and Acting in Language Models," in *ICLR 2023*, arXiv:2210.03629, 2023. DOI: 10.48550/arXiv.2210.03629.

[20] A. Singh et al., *Agentic Retrieval-Augmented Generation: A Survey on Agentic RAG*, arXiv:2501.09136, 2025. DOI: 10.48550/arXiv.2501.09136.

[21] E. Mosqueira-Rey et al., "Human-in-the-Loop Machine Learning: A State of the Art," *Artificial Intelligence Review*, vol. 56, pp. 3005–3054, 2023. DOI: 10.1007/s10462-022-10246-w.

[22] M. Wooldridge and N. R. Jennings, "Intelligent Agents: Theory and Practice," *The Knowledge Engineering Review*, vol. 10, no. 2, pp. 115–152, 1995. DOI: 10.1017/S0269888900008122.

[23] ISO, *ISO 23247-1:2021 – Automation Systems and Integration: Digital Twin Framework for Manufacturing – Part 1*, ISO, Geneva, 2021.

[24] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers & Industrial Engineering*, vol. 150, p. 106889, 2020. DOI: 10.1016/j.cie.2020.106889.

[25] O. Fink, Q. Wang, M. Svensen, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Engineering Applications of Artificial Intelligence*, vol. 92, p. 103678, 2020. DOI: 10.1016/j.engappai.2020.103678.

[26] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, A. Madotto, and P. Fung, "Survey of Hallucination in Natural Language Generation," *ACM Computing Surveys*, vol. 55, no. 12, pp. 248:1–248:38, 2023. DOI: 10.1145/3571730.

[27] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, and T. Liu, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges and Open Questions," arXiv:2311.05232, 2023. DOI: 10.48550/arXiv.2311.05232.

[28] N. F. Noy, Y. Gao, A. Jain, A. Narayanan, A. Patterson, and J. Taylor, "Industry-Scale Knowledge Graphs: Lessons and Challenges," *Queue*, vol. 17, no. 2, pp. 48–75, 2019. DOI: 10.1145/3329781.3332266.

[29] S. Appalaraju, B. Jasani, B. U. Kota, Y. Xie, and R. Manmatha, "DocFormer: End-to-End Transformer for Document Understanding," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 993–1003, 2021. DOI: 10.1109/ICCV48922.2021.00103.

[30] A. Hogan et al., "Knowledge Graphs," *ACM Computing Surveys*, vol. 54, no. 4, pp. 71:1–71:37, 2021. DOI: 10.1145/3447772.

[31] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying Large Language Models and Knowledge Graphs: A Roadmap," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 7, pp. 3580–3599, 2024. DOI: 10.1109/TKDE.2023.3352100.

[32] K. Valmeekam, M. Marquez, S. Sreedharan, and S. Kambhampati, "Large Language Models Still Can't Plan (A Benchmark for LLMs on Planning and Reasoning about Change)," in *NeurIPS 2023 Workshop on Foundation Models for Decision Making*, arXiv:2206.10498, 2023. DOI: 10.48550/arXiv.2206.10498.

[33] Q. Dong, L. Li, D. Dai, C. Zheng, Z. Wu, B. Chang, X. Sun, J. Xu, L. Li, and Z. Sui, "A Survey on In-Context Learning," arXiv:2301.00234, 2023. DOI: 10.48550/arXiv.2301.00234.

[34] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, and J. Wen, "A Survey on Large Language Model Based Autonomous Agents," *Frontiers of Computer Science*, vol. 18, no. 6, p. 186345, 2024. DOI: 10.1007/s11704-024-40231-1.

[35] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 114, no. 13, pp. 3521–3526, 2017.

[36] M. McCloskey and N. J. Cohen, "Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem," *Psychology of Learning and Motivation*, vol. 24, pp. 109–165, 1989.

[37] Y. Zhang, B. Kang, B. Hooi, S. Yan, and J. Feng, "Deep Long-Tailed Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 10795–10816, 2023.

[38] J. Hosang, R. Benenson, and B. Schiele, "Learning Non-Maximum Suppression," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4507–4515, 2017.

MAESTRO — Multi-Agent System for Smart Factory Management

Bilal AISSAOUI | Kasdi Merbah University Ouargla | 2025–2026

Abstract (English)

Industrial plants accumulate valuable engineering knowledge in P&ID drawings, operational manuals, and technical documents, yet manual retrieval consumes significant time and cognitive effort – every minute of diagnostic delay carries direct costs in production and safety. MAESTRO is a governed Multi-Agent System that transforms this silent knowledge into structured, audit-traceable diagnostic intelligence through a seven-layer architecture. Governance is enforced architecturally: the agent cannot issue control actions under any circumstance. Validated on a refrigeration-loop P&ID: 12 candidates with zero false positives (zero engineering time on phantom candidates), six diagnostic scenarios at 8–18 seconds operational speed, and 85 expert knowledge chunks retrieved. Positioned at TRL 4–5.

Keywords: Multi-Agent Systems, P&ID Digitisation, Industrial Knowledge Graph, Human-in-the-Loop Governance, Agentic RAG, ISA-5.1.

Résumé (Français)

Les installations industrielles accumulent des connaissances précieuses dans leurs schémas P&ID, leurs manuels d'exploitation et leurs documents techniques, dont la récupération manuelle mobilise un temps et un effort cognitif considérables – chaque minute de retard diagnostique engendre des pertes directes en production et en sécurité. MAESTRO est un Système Multi-Agents gouverné qui transforme ces connaissances silencieuses en raisonnement diagnostique structuré et traçable via une architecture à sept couches. La gouvernance est imposée architecturalement : l'agent ne peut émettre aucune commande de contrôle en aucune circonstance. Validé sur un P&ID de boucle de réfrigération : 12 candidats sans faux positifs (zéro temps ingénieur sur des candidats fantômes), six scénarios diagnostiques en 8–18 secondes, et 85 fragments de connaissances expertes récupérés. Positionné à TRL 4–5.

Mots-clés : Systèmes Multi-Agents, Numérisation P&ID, Graphe de Connaissances Industriel, Humain-dans-la-Boucle, RAG Agentique, ISA-5.1.

Abstract (Arabic)

الملخص:

يُشكّل الاسترجاع اليدوي للمعرفة الهندسية من مخططات الأنابيب والأدوات (P&ID) ودفاتر الإجراءات التشغيلية والوثائق التقنية عائقاً حرجاً في عمليات المنشآت الصناعية؛ إذ تُرجَم كل دقيقة تأخير في التشخيص إلى خسائر مباشرة في الإنتاج والسلامة. تستلزم المناهج القائمة إما رقمنة يدوية مكلفة، وإما الاعتماد على أنظمة ذكاء اصطناعي ذات أغراض عامة تُنتج مخرجات معقولة لكن غير قابلة للتحقق على مستوى المنشأة.

تُقدّم هذه الرسالة نظام MAESTRO، وهو نظام متعدد الوكلاء محكوم يُحوّل مخططات P&ID والوثائق الصناعية إلى معرفة منشأتية منظّمة وقابلة للتدقيق، عبر معمارية سباعية الطبقات تشمل: الاستخراج بقيادة التعرف البصري على النص، والإثراء الدلالي وفق معيار ISA-5.1، وترقية المعرفة بإشراف بشري، والاستدلال التشخيصي الوكيلي. تُفرض الحوكمة على المستوى المعماري: تمنع عقود أدوات الوكيل تنفيذ أوامر التحكم منعاً بنيوياً، بصرف النظر عن أي تعليمات أو توجيهات.

جُرّب النظام على مخطط P&ID لحلقة تبريد، فاستخرج جميع مرشحات الأجهزة والمعدات بنسبة صفر أخطاء إيجابية كاذبة — ما يُلغي الوقت الهندسي المهدر في التحقق من مرشحات وهمية — ونفّذ ستة سيناريوهات تشخيصية بأوقات استجابة تتراوح بين 8 و18 ثانية، جامعاً بين اجتياز طوبولوجيا الشبكة، واسترجاع القياسات الحية، واسترداد 85 قطعة من المعرفة الخبراء ضمن مسار محكوم واحد.

تُثبت هذه النتائج جدوى الاستدلال الوكيلي الواعي بالمنشأة عند مستوى الجاهزية التقنية TRL~4-5 دون المساس بالحوكمة أو إمكانية التدقيق، مُرسيةً مساراً قابلاً للتكرار نحو ذكاء اصطناعي صناعي قابل للنشر.



Conceptual overview – TRL 4–5.