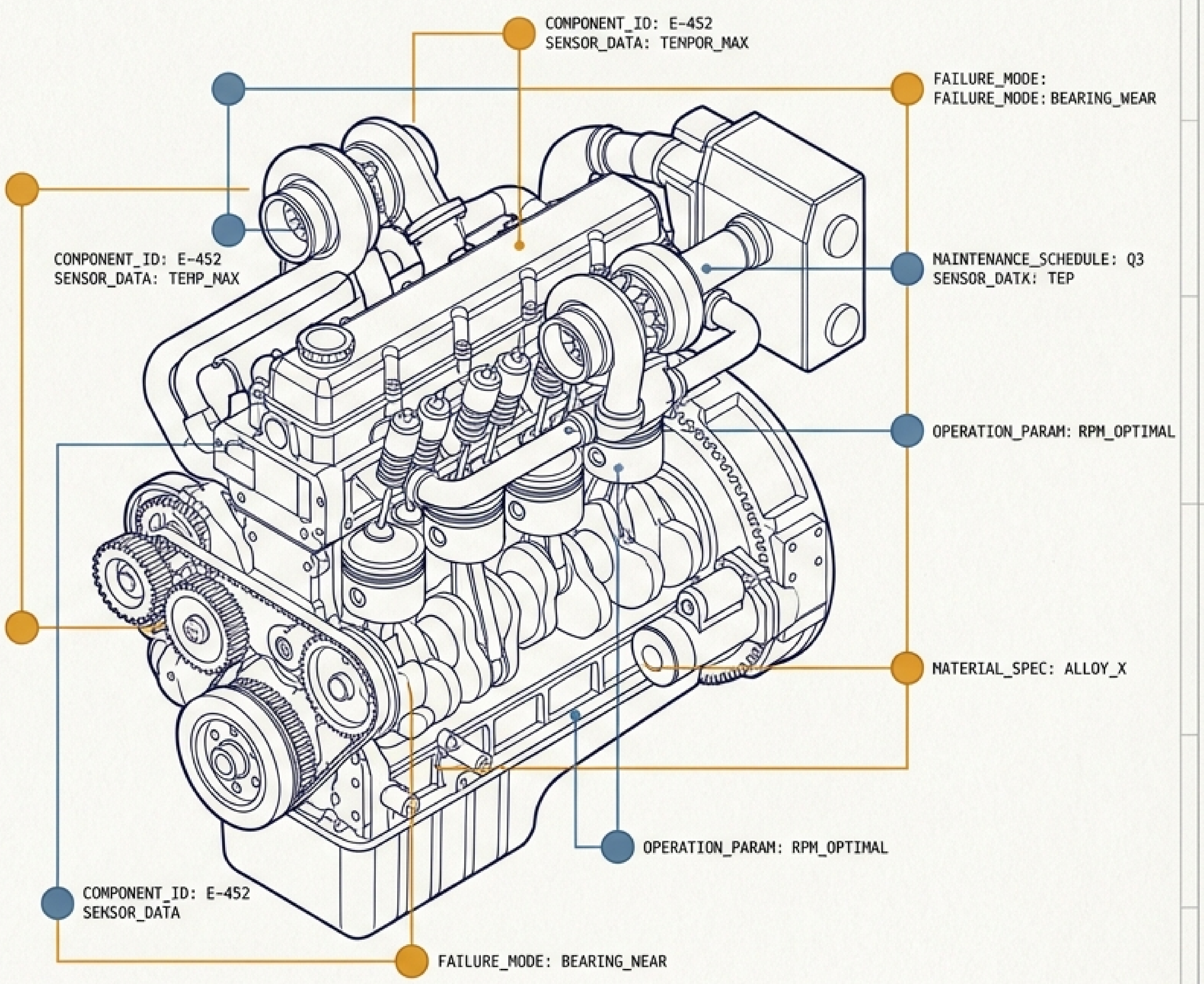
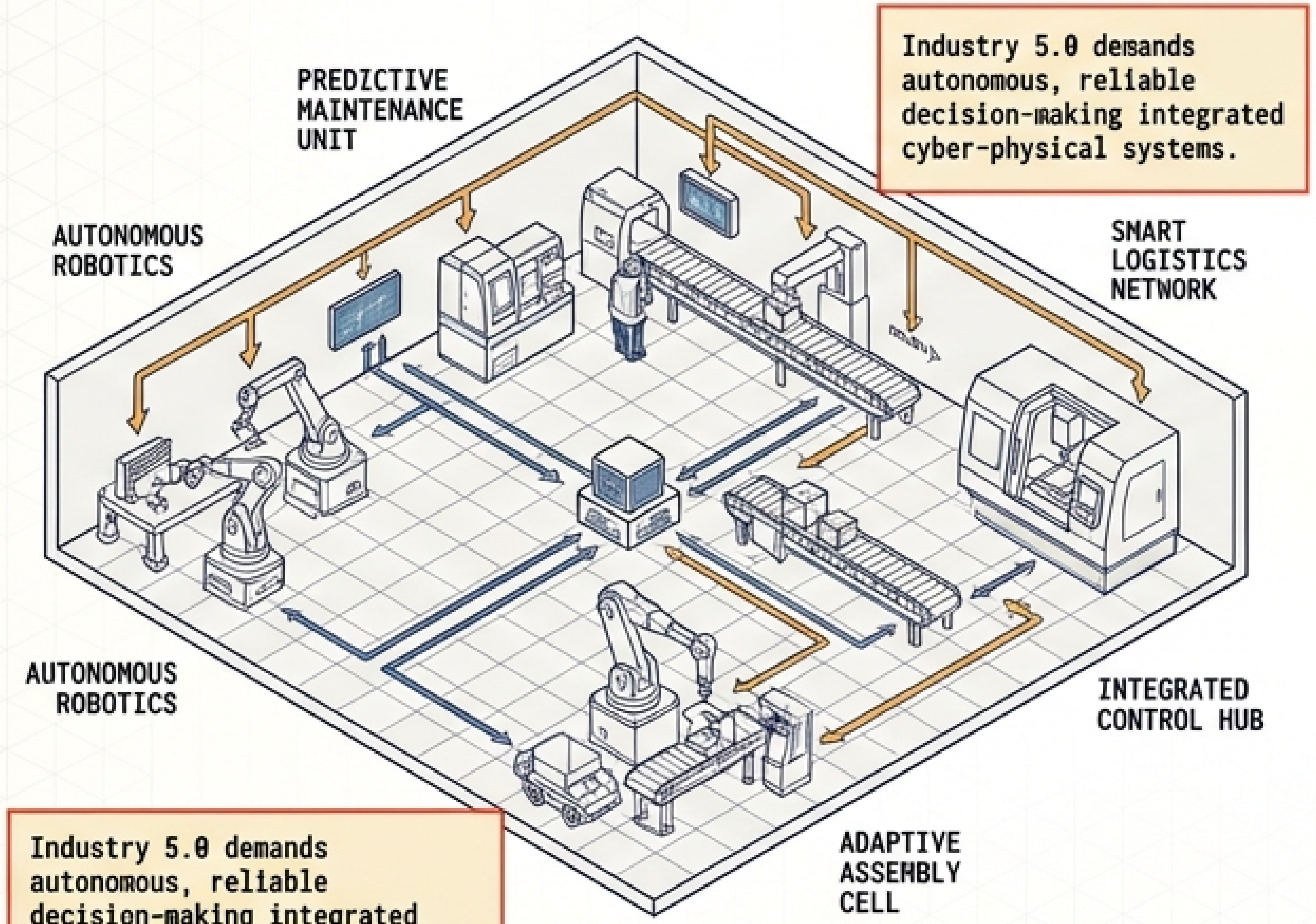
Explicit Knowledge, Explainable AI

The Knowledge-Graph-Informed Machine Learning (KGIML) Paradigm for Predictive Modelling in Industry 4.0 & 5.0



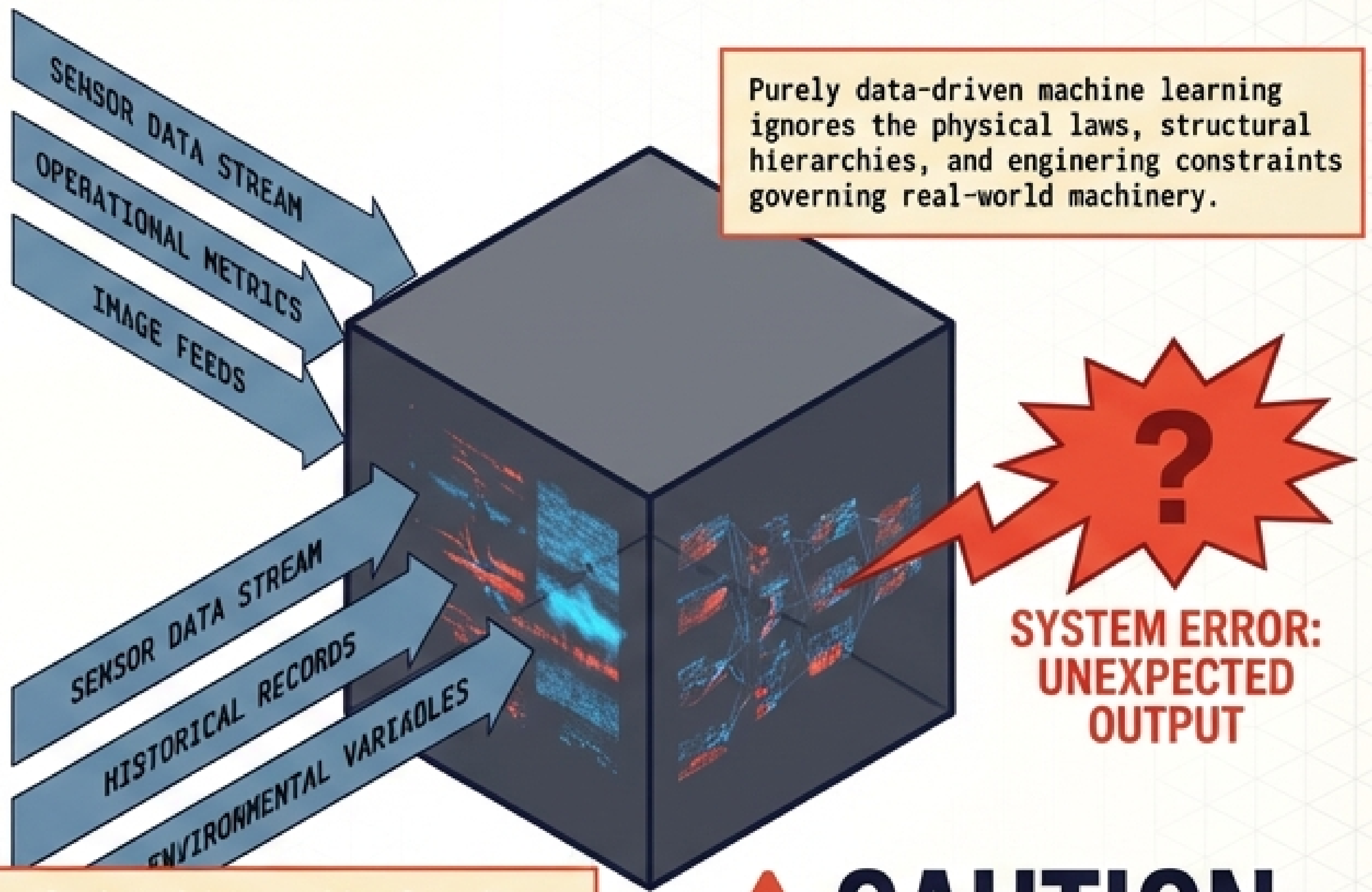
The AI Ceiling in Industry 4.0 & 5.0

The Demand: Cyber-Physical Autonomy



Industry 5.0 demands autonomous, reliable decision-making integrated across cyber-physical systems.

The Failure: The Opaque Black Box



Purely data-driven machine learning ignores the physical laws, structural hierarchies, and engineering constraints governing real-world machinery.

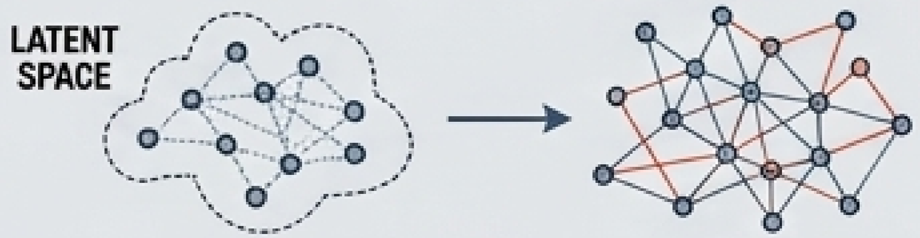
CAUTION: NO EXPLAINABILITY

Takeaway: Real-world industrial systems are inherently structured. Models that learn from flat data streams without explicit domain knowledge hit a performance and trust ceiling.

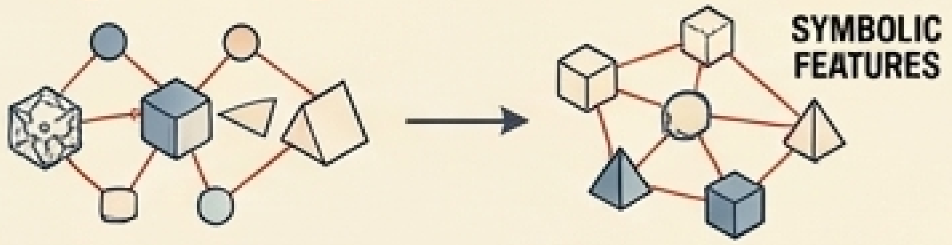
The Gap: Implicit Learning vs. Explicit Symbolic Enrichment

	Current Graph Neural Networks (GNNs)	Explicit KGIML
Mechanism	**Implicit Latent Spaces**	**Explicit Symbolic Features**
Transparency	**Opaque Black Box**	**White Box Explainable Provenance**
Representation	**Mathematically Embedded Parameters**	**Semantically Grounded Graph Structure**
Ontology	**Weakly Aligned**	**Strictly Governed**

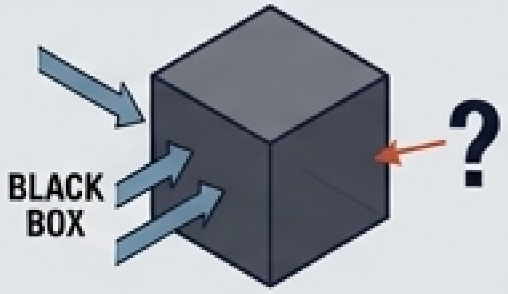
Mechanism — Implicit Latent Spaces



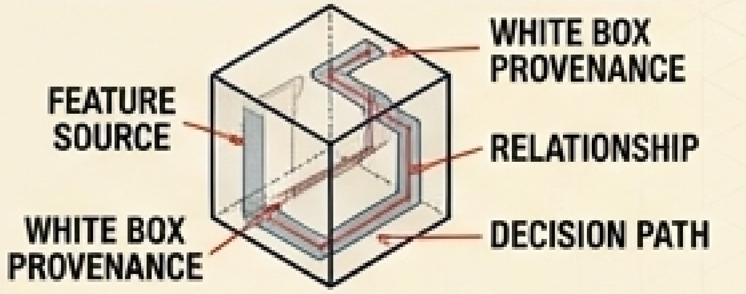
Mechanism — Explicit Symbolic Features



Transparency — Opaque Black Box



Transparency — White Box Explainable Provenance

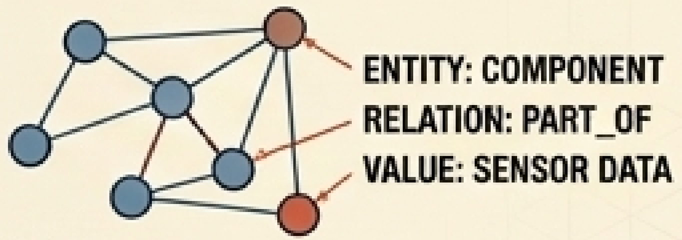


Representation — Mathematically Embedded Parameters

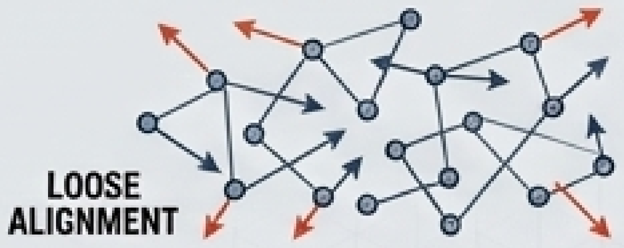
$$E = \frac{1}{n}\sum_{i=1}^{n}(x_i + n_j)^2 = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 8 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 3 & 0 \\ 0 & 0 & 1 & 5 & 2 & 1 \\ 0 & 0 & 0 & 5 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

PARAMETERS

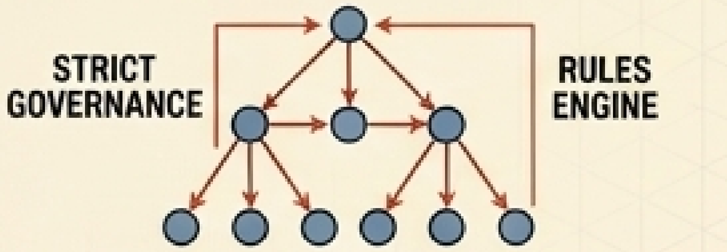
Representation — Semantically Grounded Graph Structure



Ontology — Weakly Aligned

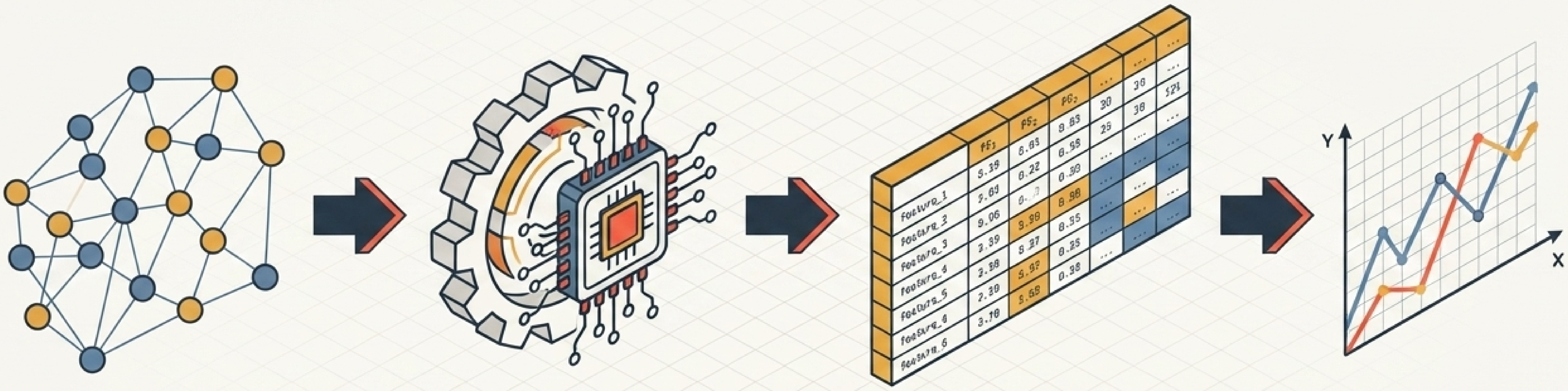


Ontology — Strictly Governed



Current methods bury knowledge inside unreadable vector weights.
KGIML transforms knowledge into readable predictive features prior to training.

Introducing KGIML: A Data-Centric, Neuro-Symbolic Frameworkwerk



Raw Knowledge Graph

KGIML Engine

Enriched Contextual Feature Space

Standard Machine Learning Model

Definition

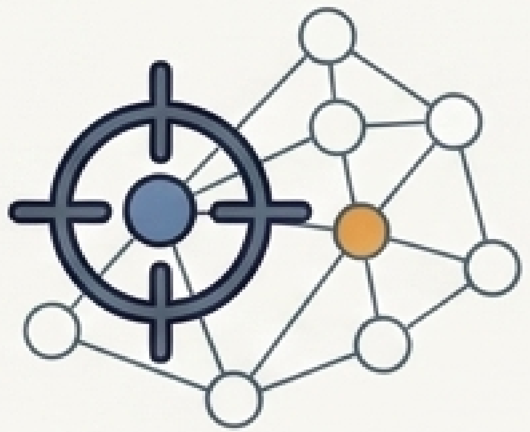
Knowledge-Graph-Informed Machine Learning (KGIML) constructs new numeric features for graph nodes by systematically propagating and aggregating information through object-property relations.

Core Objective

Enhance the numeric data associated with a node by deriving multiple layers of contextual attributes from the surrounding graph structure, entirely prior to model training.

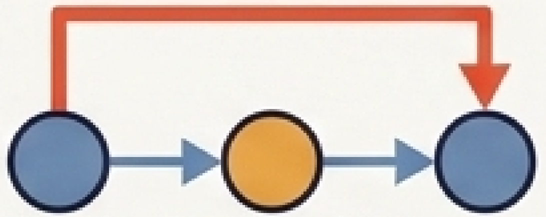
The KGIML Architecture

Step 1: Select



Identify **systemic structural relations** and propagation-friendly datatype attributes.

Step 2: Compose

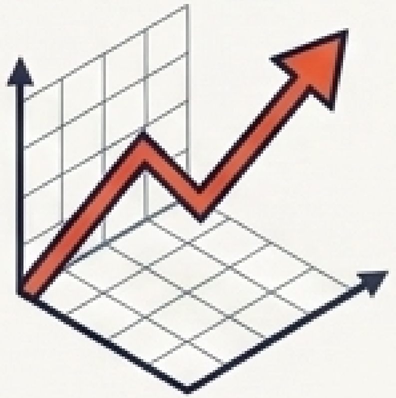


Generate **multi-hop object property pathways** dynamically using recursive SPARQL.

Step 3: Aggregate

Construct **new contextual attributes** by aggregating values along paths using a bias-controlled Lehmer mean.

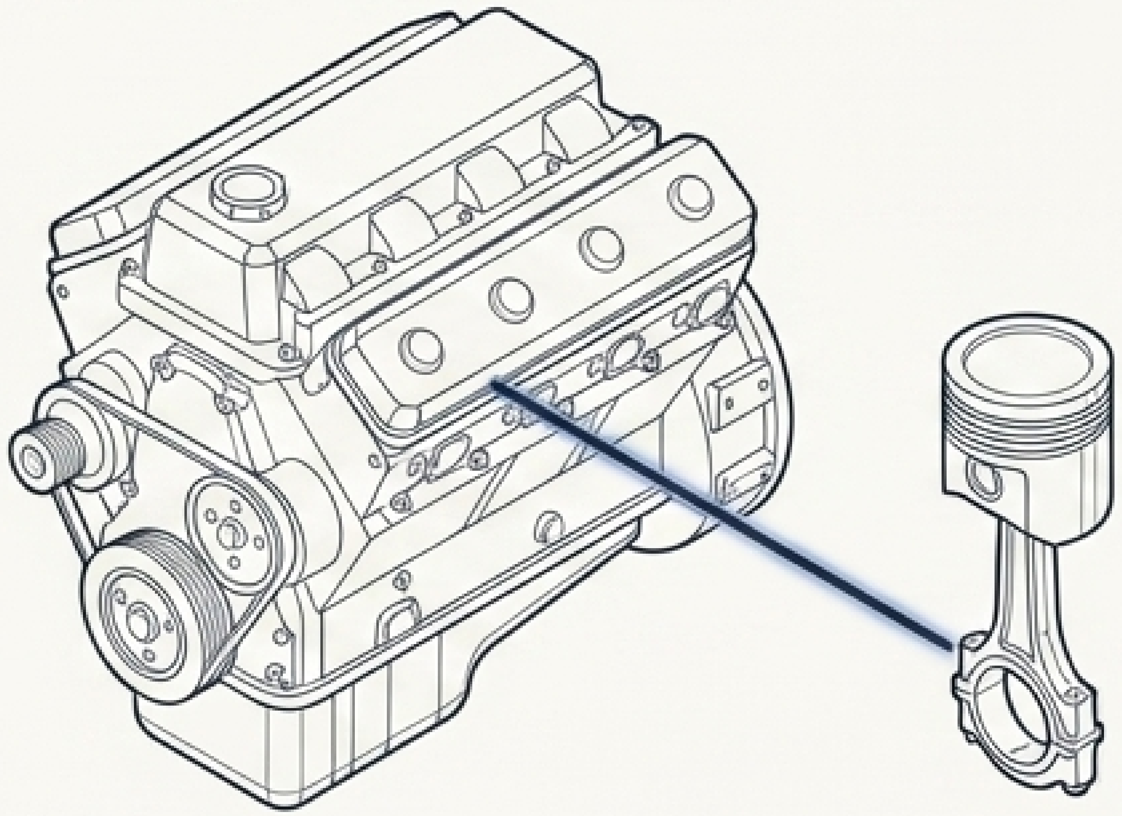
Step 4: Predict



Train **standard ML models** on this newly enriched, semantically grounded feature space.

Step 1: Selecting the Right Propagation Channels

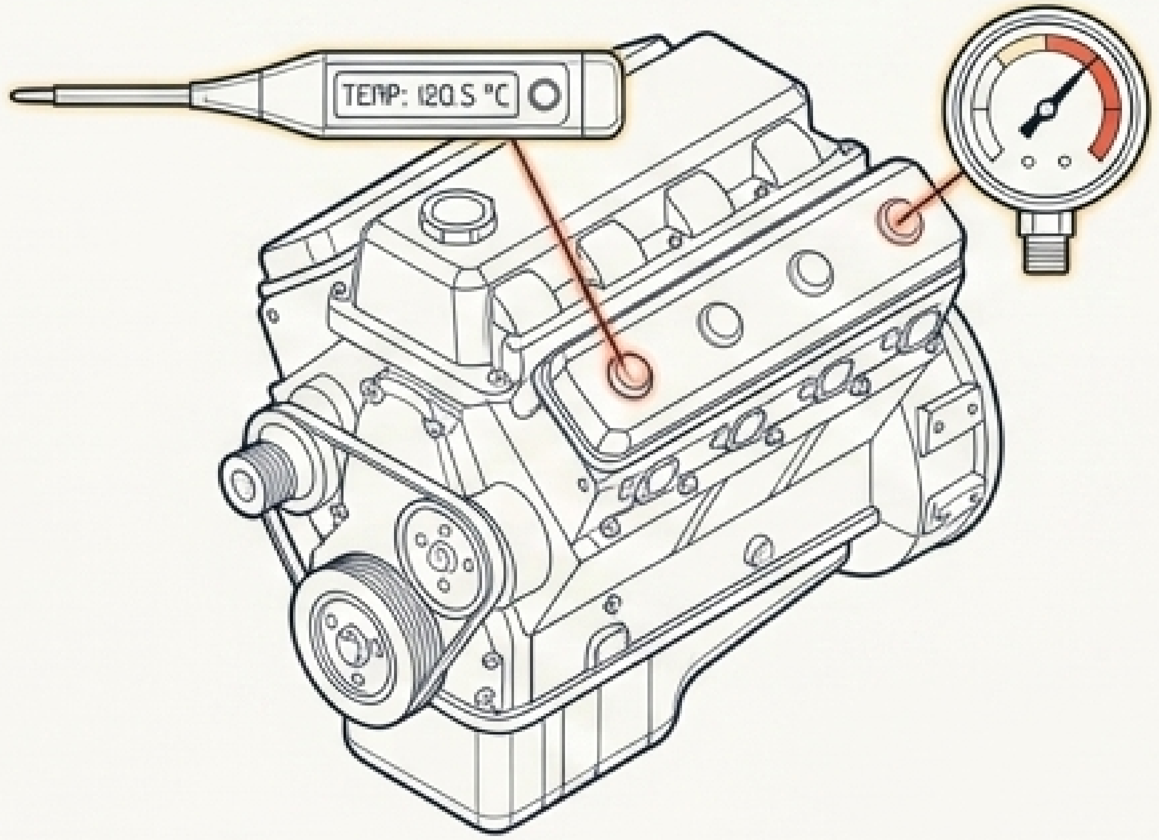
A. Structural Relations



Must capture **systemic dependencies**.

```
hasPart, isPartOf, getsInputFrom,
isPhysicallyConnectedTo
```

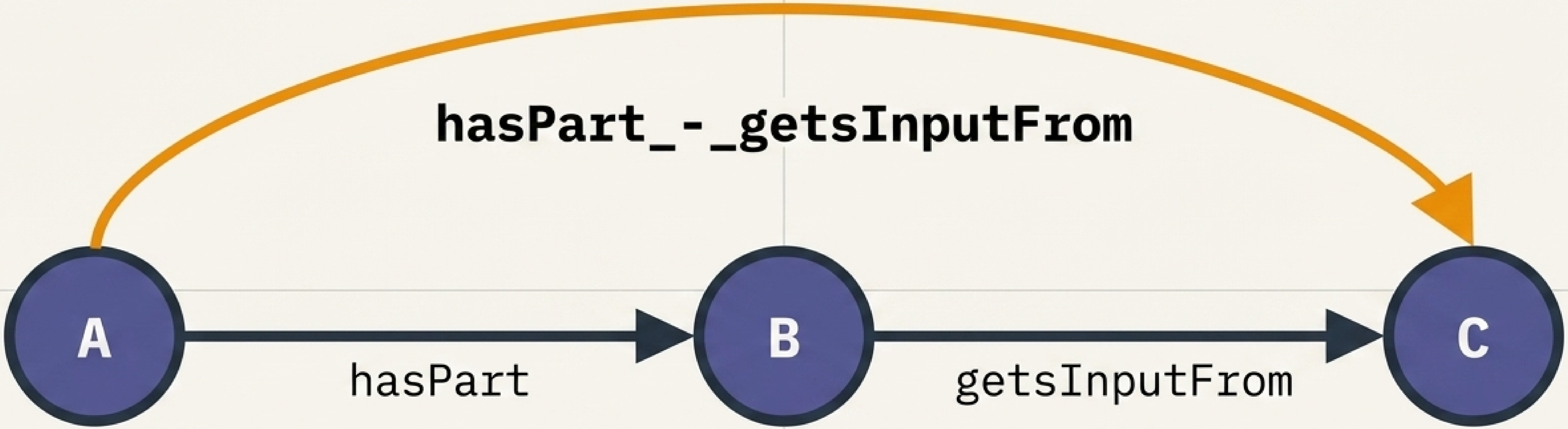
B. Datatype Attributes



Must have **physical continuity** and **quantitative interpretability** to meaningfully diffuse across structures.

```
Temperature, Pressure, Flow Rate
```

Step 2: Multi-Hop Relation Composition



The Innovation

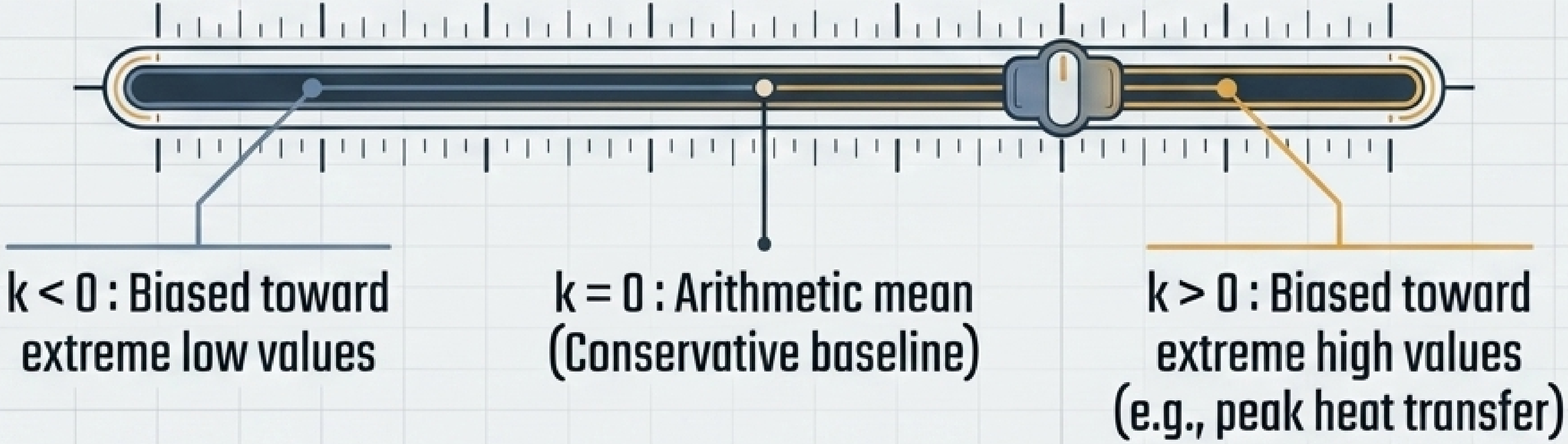
KGIML dynamically generates composite k-hop object properties via recursive SPARQL self-queries. It discovers deep contextual relationships defining the structural neighborhood of any entity.

The Mechanism

It does not permanently alter the core schema by statically adding combinations. At iteration t, it traverses chains of length t, materializes the paths, and uses the enriched graph for iteration t+1.

Step 3: Contextual Attribute Aggregation via Lehmer Mean

$$L_k(V) = \frac{\sum v^{k+1}}{\sum v^k}$$

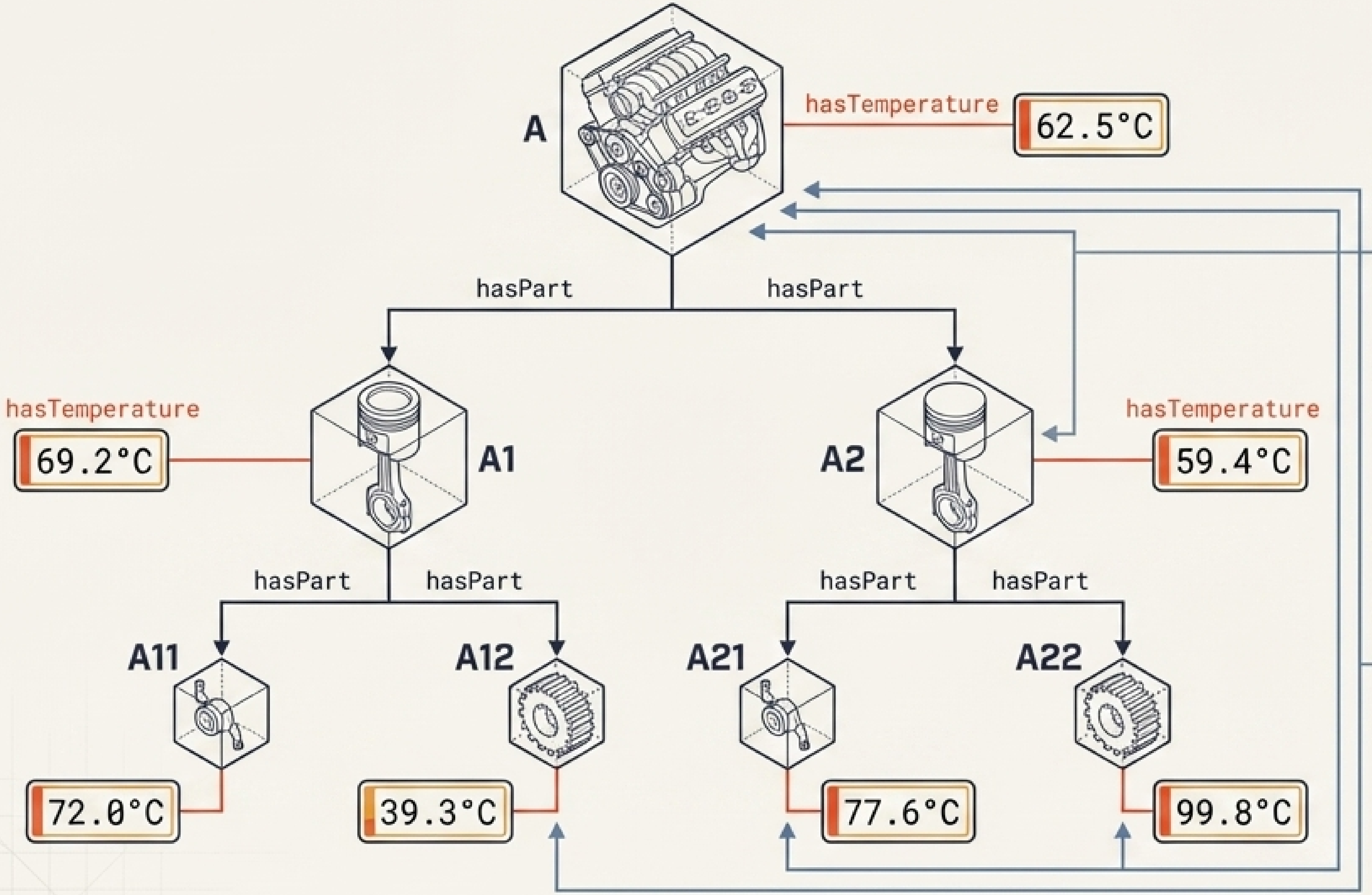


Semantic Consistency

'k' parameters are strictly defined directly within the ontology to avoid combinatorial explosion.

```
k_hasPart = 2.0
```

Feature Generation in Action: The Propagation Mechanism



1-Hop Enrichment:

Sub-components have local temperatures of 69.2°C and 59.4°C. Applying a Lehmer mean (k=2) generates a new contextual feature for parent node A:

`hasPart_-_hasTemperature = 65.0°C`

2-Hop Enrichment:

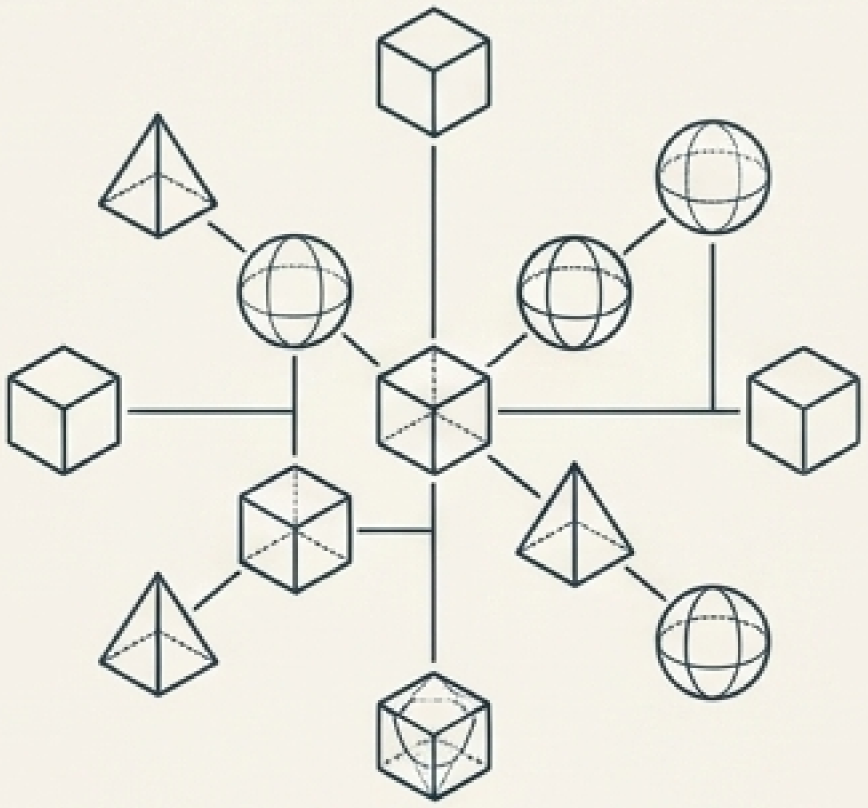
Pulling from deeper components (72.0°C, 39.3°C, 77.6°C, 99.8°C), KGIML generates a 2-hop feature for node A:

`hasPart_-_`
`_hasPart_-_hasTemperature = 83.5°C`

Takeaway: Explicit contextual features are materialized for machine learning without runtime graph traversal.

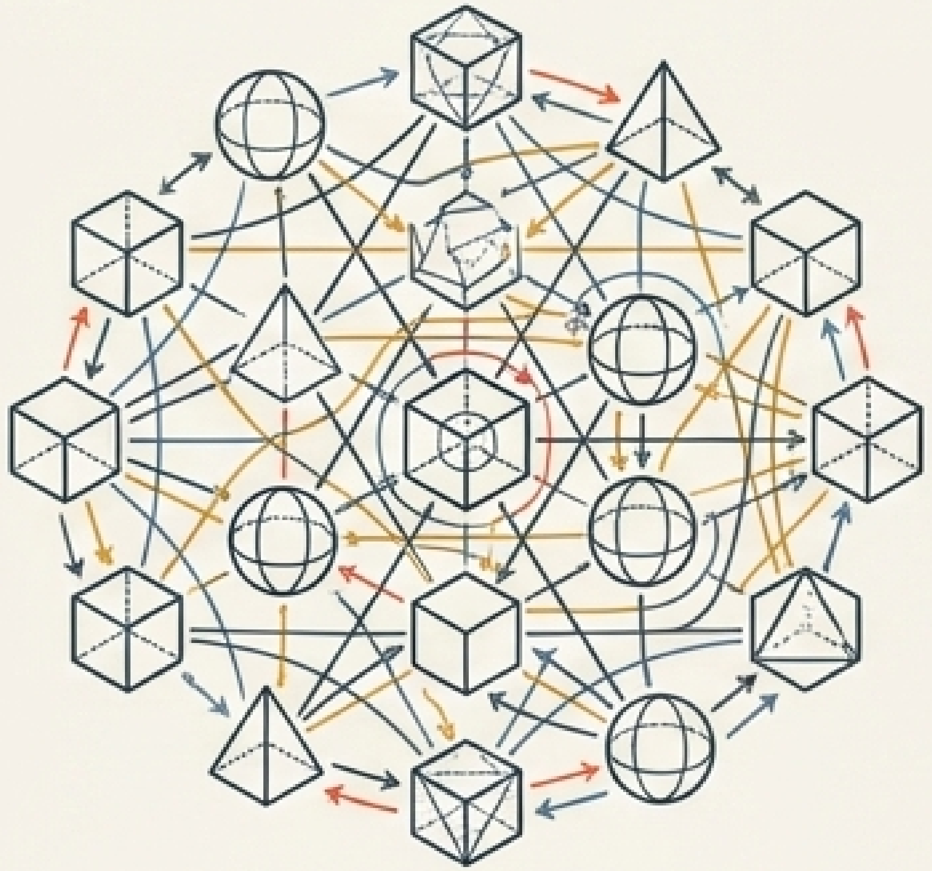
Experimental Proof: Validating the Paradigm

Baseline Knowledge Graph



~50,000 Triples

Enhanced Knowledge Graph

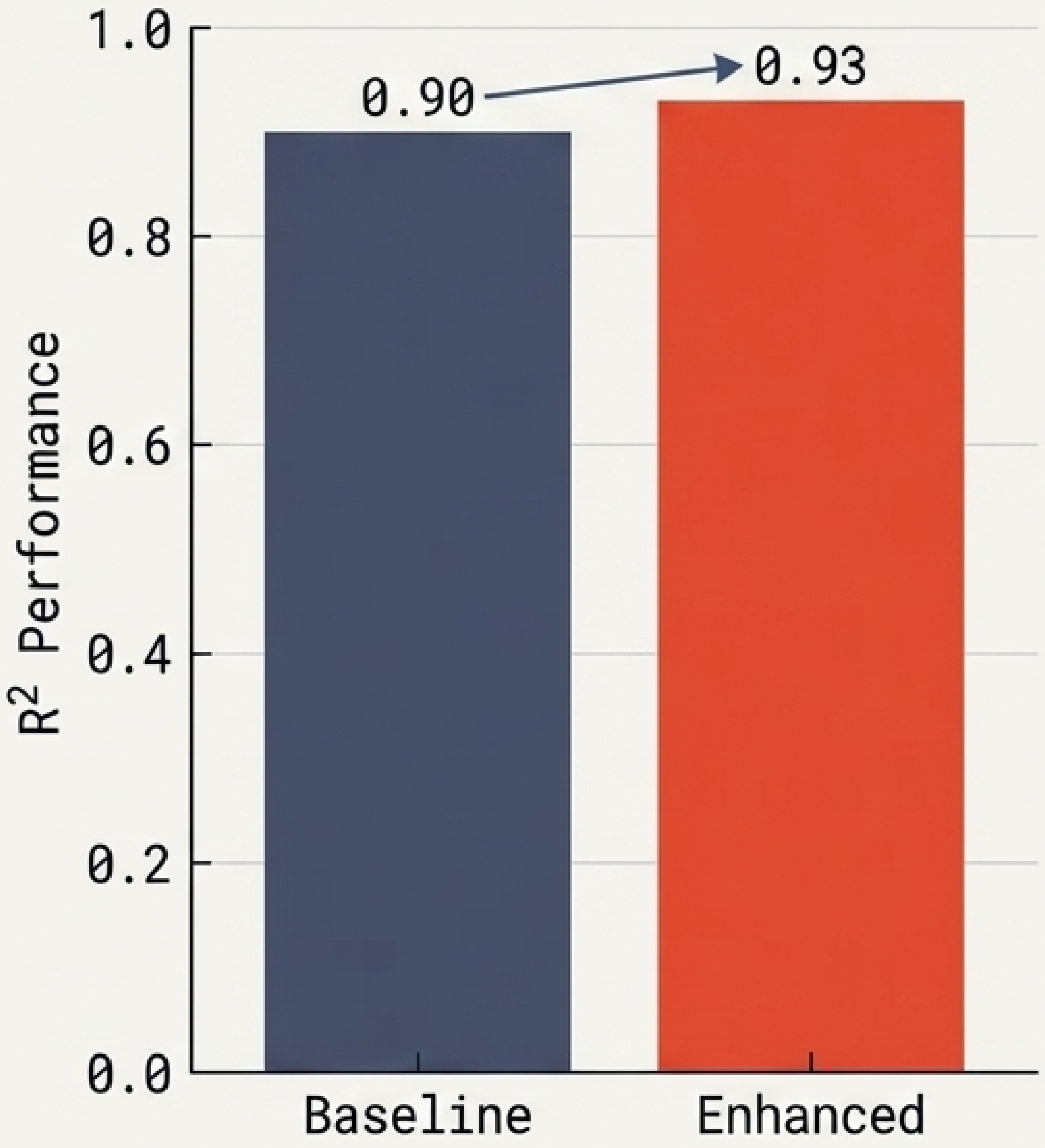


~621,000 Triples (After 3 Iterations)

The Setup

- **Scale:** 5,000 synthetic industrial object instances with complex relational structures.
- **Method:** Evaluated over 10 independent runs using Ridge Regression models.
- **Objective:** Determine if KGIML-injected structural bias is recoverable and statistically useful.

Experiment 1: Capturing Multi-Hop Compositional Signals



Target Objective

Predicting hasFailureRisk as a function of multi-hop relational compositions (e.g., getsInputFrom, providesOutputTo).

The Result

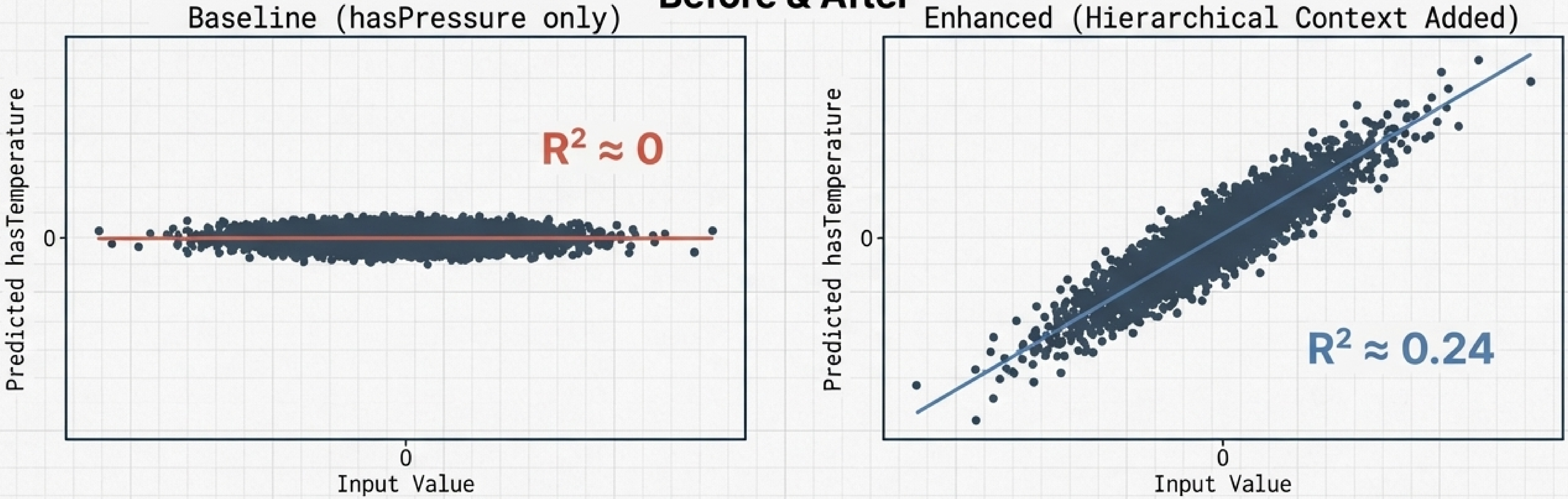
The Enhanced model dramatically outperformed the baseline. R^2 performance jumped from ~0.90 to ~0.93, with a consistent, significant reduction in Mean Squared Error (MSE).

The Insight

Multi-hop relational compositions are detectable. KGIML successfully encodes non-trivial dependencies that are ready for downstream ML exploitation.

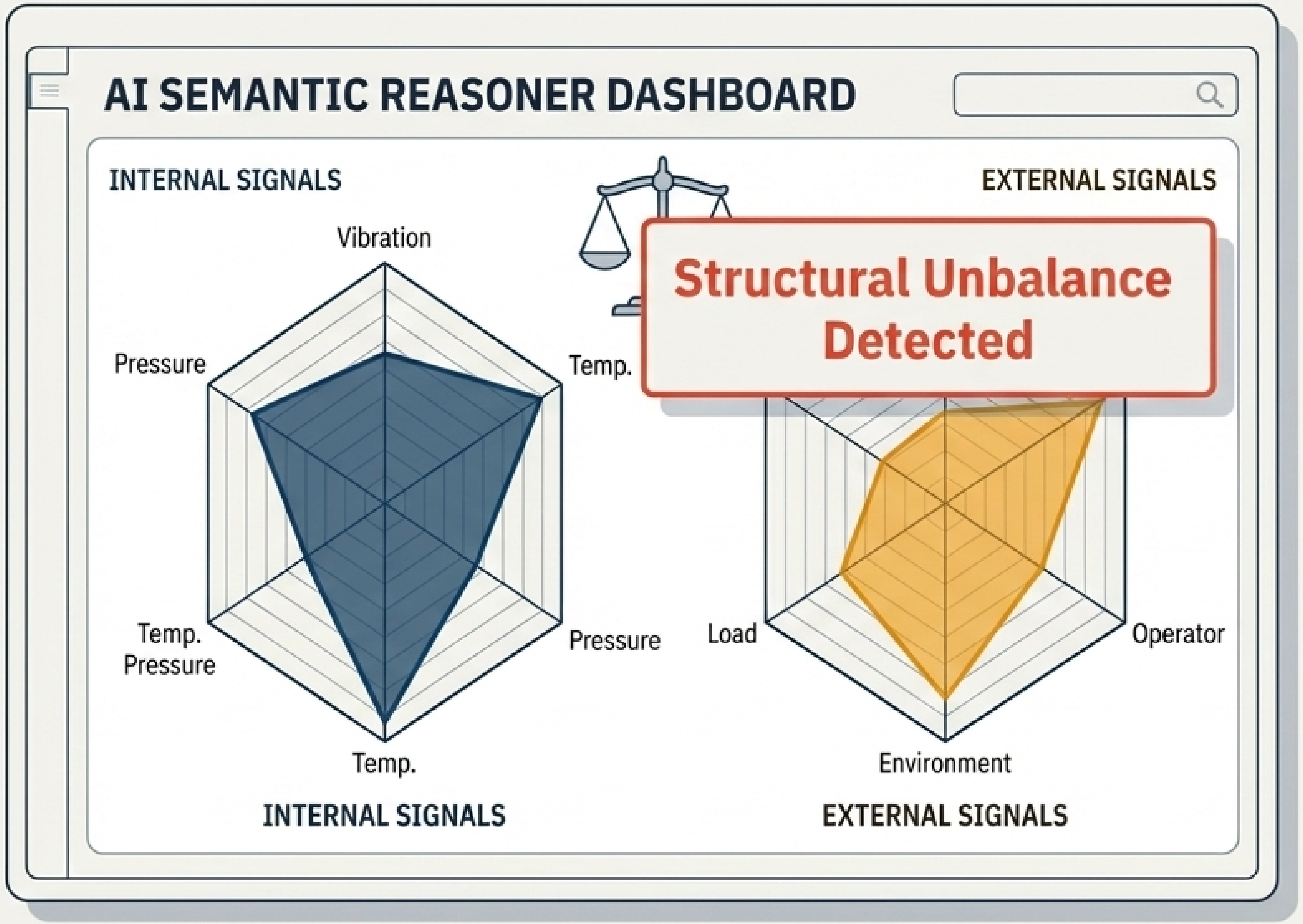
Experiment 2: Unlocking Hierarchical Dependencies

Before & After



Target Objective	The Result	Statistical Proof
Predicting hasTemperature using only hasPressure versus adding hierarchical composite attributes.	The Baseline model failed entirely. The Enhanced model achieved stable, overwhelming predictive power.	The improvement is statistically overwhelming (p ≈ 6.1 × 10^-14). Explicit structural context is strictly required for accurate prediction.

Beyond ML: Rule-Driven Predictive Maintenance



The Application

KGIML empowers semantic rules (SWRL). By enriching KGs with contextual features, reasoning engines can monitor the balance between an object's internal structure and its external environment.

The Value

Discovers hidden unbalances automatically, generating highly explainable, discoverable alerts for predictive maintenance before physical failures occur.

A Universal Architecture for Relational Systems

Power Grids

Upstream transformer hierarchy predicts local voltage load fluctuations.

Logistics & Supply Chain

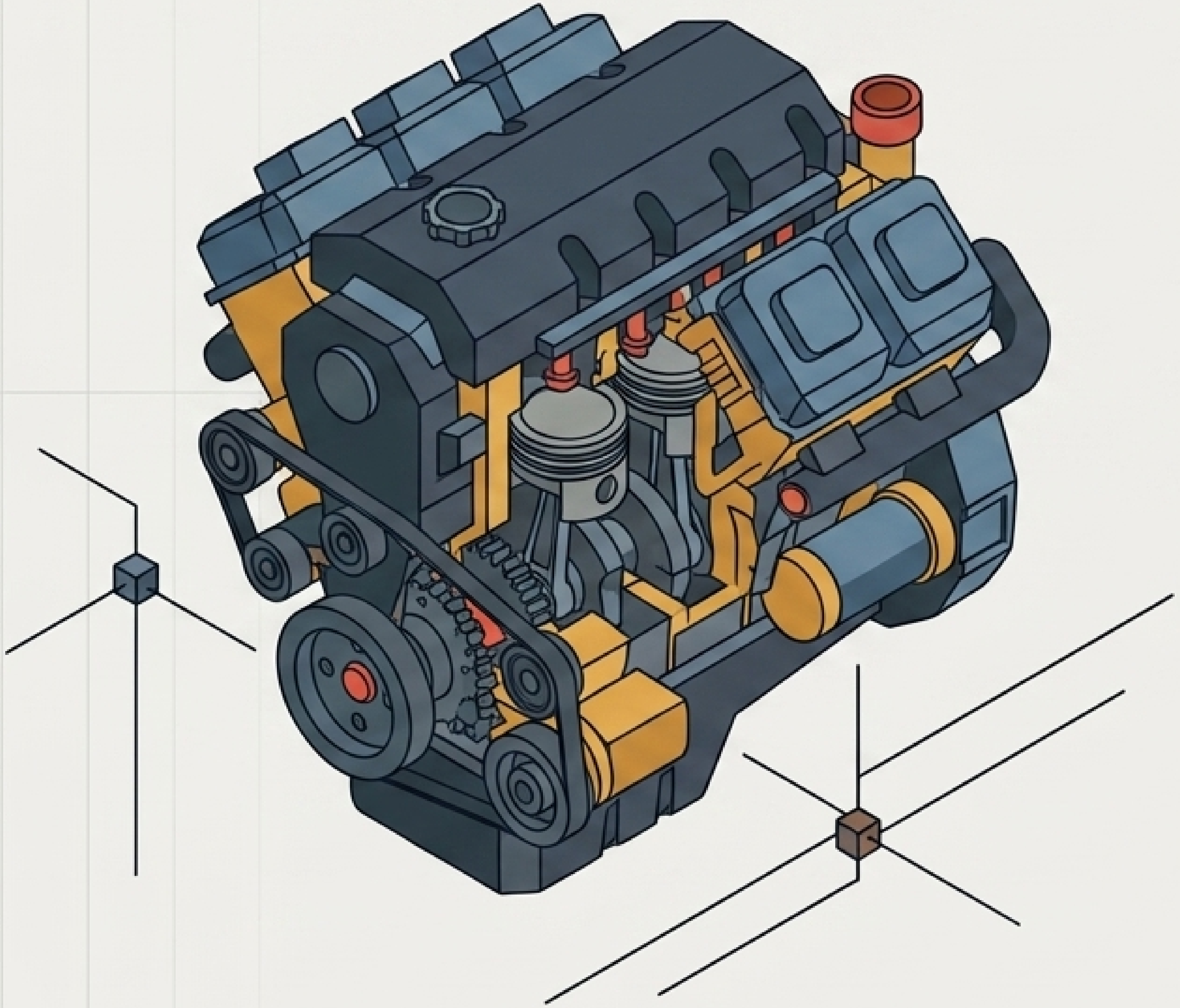
Route connectivity networks predict downstream hub delay metrics.

Smart Manufacturing

Component hierarchies predict system-wide vibration anomalies.

Synthesis: Wherever relational structures dictate physical outcomes, KGIML generates the semantic context required to predict them.

The Neuro-Symbolic Future of Industry 5.0



Explicit over Implicit

KGIML shifts AI from opaque statistical embeddings to explicit, explainable symbolic enrichment.

Domain Aligned

Provides human-in-the-loop validation, explainable feature provenance, and strict alignment with engineering ontologies.

The Final Takeaway

KGIML delivers the transparency, reliability, and domain-driven constraints required to trust AI with the infrastructure of the future.