

Energy Twin: A Possibilistic Neuro-Symbolic Framework for Early Warning Detection in Waste-to-Energy Boilers

Esteban Guerrero^{1,*†}, Andreas Brännström^{1,‡}, Rachele Carli^{1,†}, John Granström^{1,†},
Måns Kjellander^{2,†} and Juan Carlos Nieves^{1,†}

¹Department of Computing Science, Umeå University, Sweden

²Umeå Energi AB, Sweden

Abstract

Industrial anomaly detection systems commonly rely on statistical forecasting models that provide limited support for explicit uncertainty handling and symbolic reasoning. This becomes problematic in safety-critical environments where incomplete evidence, sensor degradation, and conflicting observations may coexist during early failure stages. This work presents the *Energy Twin*, a neuro-symbolic framework for possibilistic early warning in industrial boiler systems. The framework combines temporal neural representations with symbolic semantic reasoning to model epistemic degradation states under uncertainty. A Temporal Convolutional Network (TCN) extracts neural representations from multivariate telemetry streams, while a possibilistic layer computes *possibility* and *necessity* measures used to characterize *nominal*, *conflict*, and *ignorance* states. The resulting uncertainty signals are propagated into an ontology-based reasoning layer through symbolic sensor assertions. The symbolic component integrates OWL ontologies and SWRL rules to perform semantic verification of structural failure conditions over the inferred sensor evidence. In addition, gradient-based saliency is used to identify the temporal variables contributing most strongly to the detected degradation, enabling a dual explanation mechanism combining statistical attribution with symbolic diagnosis. A running example based on telemetry data from a real waste-to-energy facility in Northern Sweden illustrates the operational behavior of the framework under evolving degradation conditions. The results show how possibilistic reasoning can support early warning (>24 hours), uncertainty characterization, and semantic root-cause analysis within a unified neuro-symbolic architecture.

Keywords

Possibility Theory, Neuro-Symbolic AI, Digital Twins, Out-of-Distribution Detection, Knowledge Graphs

1. Introduction

Waste-to-energy (W2E) boiler systems operate under complex multivariate dynamics where undetected failures may produce significant operational and environmental consequences [1, 2, 3]. For this reason, industrial early warning systems commonly rely on deep learning models trained over multivariate telemetry streams to support predictive maintenance and anomaly detection. However, standard probabilistic architectures exhibit limitations when deployed under open-world conditions containing incomplete or previously unobserved evidence [4]. In such scenarios, softmax-based classifiers may produce overconfident predictions even when the observed system state deviates from the training distribution [5, 6]. This becomes problematic in safety-critical environments where competing failure explanations and uncertain sensor behavior may coexist during early degradation stages.

RuleML+RR'26: Companion Proceedings of the 10th International Joint Conference on Rules and Reasoning, August 24–26, 2026, Vilnius, Lithuania

*Corresponding author.

†This project is supported by VINNOVA under the project: “A trustworthy decision-support system for energy management at Umeå Energi”. Esteban was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) [proj. 570011330], and the Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society (WASP-HS) [proj. 570011505]

‡This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

✉ esteban.guerrero@umu.se (E. Guerrero)

ORCID 0000-0002-6035-800X (E. Guerrero); 0000-0001-9379-4281 (A. Brännström); 0000-0002-8689-285X (R. Carli); 0000-0003-4072-8795 (J. C. Nieves)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

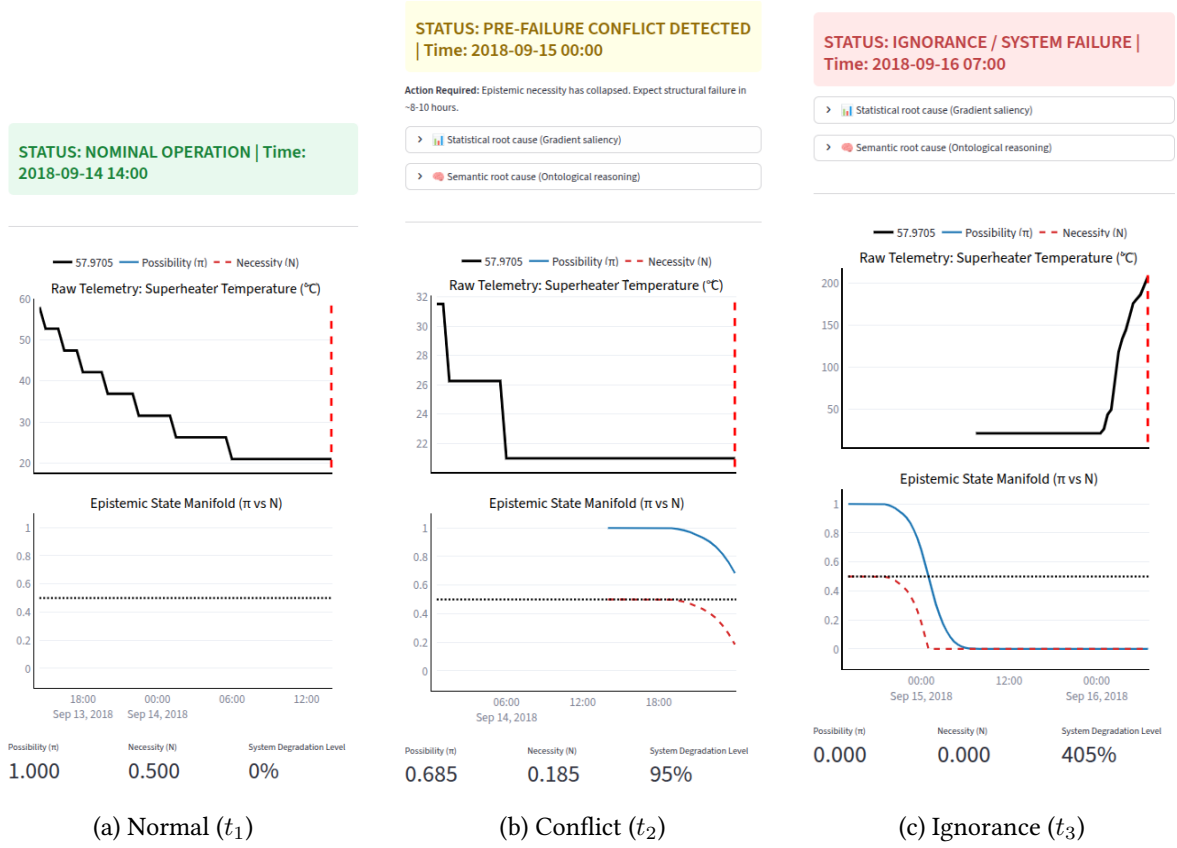


Figure 1: Part of the Energy Digital twin user interface capturing the evolution of possibilistic metrics across distinct epistemic states. From left to right: high certainty of normal operation t_1 , conflicting evidence triggering an early warning t_2 (Sept 14, 2018 19:00), and a complete collapse in possibility denoting an unobserved event t_3 (Sept 16, 2018 00:30). See that there is time gap between t_2 and t_3 showing an early warning alert when there is a conflict.

Possibility theory [7] provides an alternative uncertainty formalism capable of representing both plausibility and epistemic ignorance without enforcing normalized probabilistic competition between hypotheses. In particular, *possibility* and *necessity* measures allow uncertain operational states to be represented explicitly when the observed evidence is insufficient to support a reliable diagnosis.

This work presents the *Energy Twin*, a possibilistic neuro-symbolic framework for industrial early warning in W2E boilers. A Temporal Convolutional Network (TCN) extracts the neural representations from multivariate telemetry streams, while a possibilistic layer computes possibility and necessity measures used to characterize normal, conflict, and ignorance states. The resulting uncertainty signals are propagated into an OWL/SWRL knowledge base representing the boiler sensor topology and operational dependencies. In addition, the framework combines gradient-based saliency with symbolic ontology reasoning to provide both statistical and semantic explanations of anomalous system behavior.

The main contributions of this work are: 1) a possibilistic neuro-symbolic architecture for industrial early warning, 2) a symbolic uncertainty propagation mechanism connecting neural telemetry representations with ontology-based reasoning, and 3) a dual explanation layer combining gradient-based attribution with semantic root-cause analysis.

Running example To demonstrate the operational dynamics of the Energy Twin framework, we introduce a running example extracted from a real-world W2E facility situated in Northern Sweden. The system monitors continuous multivariate time-series data across critical boiler subsections, including flue gas lines, pressure valves, and thermal induction zones.

As shown in Figure 1a, during steady-state operation (t_1), the sensor measurements closely match the patterns that the neural network learned as normal behaviour. Consequently, the possibility of the

healthy state is $\Pi(\omega_{\text{healthy}}) = 1.0$ and its necessity is $N(\omega_{\text{healthy}}) = 1.0$, indicating complete confidence that the boiler is operating normally.

At time step t_2 (Figure 1b), the sensor measurements begin to differ from the learned normal behaviour. Although the system is still compatible with normal operation, alternative explanations also become plausible. As a result, the possibility of the healthy state decreases to $\Pi(\omega_{\text{healthy}}) = 0.50$, while the possibility of an anomalous state rises to $\Pi(\omega_{\text{anomaly}}) = 0.81$. Because competing explanations overlap, the necessity score for the healthy state precipitously drops to $N(\omega_{\text{healthy}}) = 0.0$. This divergence indicates a state of high feasibility for failure but low certainty, correctly triggering a preventative early warning without artificially forcing a definitive classification.

At time step t_3 (Figure 1c), the sensor measurements differ substantially from anything observed during training. Instead of forcing an overconfident classification, the framework identifies this situation as **epistemic ignorance**, signalling that it is facing an unknown operating condition. The system then analyses the neural network gradients to determine which sensors contributed most strongly to this warning.

2. The possibilistic early warning framework

This section formalizes the main components of the Energy Twin framework, defining how multivariate telemetry streams are mapped into possibilistic states and integrated with symbolic reasoning.

Definition 1 (Multivariate sensor time-series). Let $S = \{s_1, s_2, \dots, s_M\}$ be a set of M physical sensors monitoring the W2E boiler. A multivariate time-series window is represented as

$$X = [x_1, x_2, \dots, x_T] \in \mathbb{R}^{M \times T},$$

where each vector $x_t \in \mathbb{R}^M$ contains the sensor measurements at time step t .

The telemetry windows are embedded into a latent representation space used to compute possibilistic uncertainty measures.

Definition 2 (Possibility and necessity measures). (adapted from [8])

Let $z = f_\theta(X) \in \mathbb{R}^D$ denote the latent embedding generated by a Temporal Convolutional Network (TCN) parameterized by θ . Let $\mu_i \in \mathbb{R}^D$ be the prototype associated with an operational state $\omega_i \in \Omega$. The distance between z and μ_i is computed as: $d_i = \|z - \mu_i\|_2$. Then, the possibility measure associated with ω_i is defined as:

$$\Pi(\omega_i) = e^{-\gamma d_i}, \quad (1)$$

where $\gamma > 0$ is a scaling parameter.

The corresponding necessity measure is defined as:

$$N(\omega_i) = \min \left(\Pi(\omega_i), 1 - \sup_{j \neq i} \Pi(\omega_j) \right). \quad (2)$$

For the normal operational state ω_0 , small distances to the prototype μ_0 produce high possibility values. Conversely, when the latent representation remains distant from all known prototypes, the possibility values decrease uniformly, leading to a low necessity values across all operational states. The separation between possibility (Π) and necessity (N) allows the framework to distinguish between conflicting evidence and previously unobserved operational conditions.

Proposition 1 (Epistemic characterization of anomalies). Let Ω be the finite set of known operational states containing the normal state ω_0 , and let $\tau_N, \tau_\Pi \in (0, 1]$ be decision thresholds. Given a telemetry window X , the epistemic state is characterized as follows:

1. **Conflict state (degradation).** If $\Pi(\omega_0) \geq \tau_\Pi$ and $N(\omega_0) < \tau_N$, the operational state remains compatible with the normal manifold while simultaneously competing with alternative hypotheses.
2. **Ignorance state (OOD anomaly).** If $\sup_{\omega \in \Omega} \Pi(\omega) < \tau_\Pi$, the observed telemetry window lies outside the known operational regions.

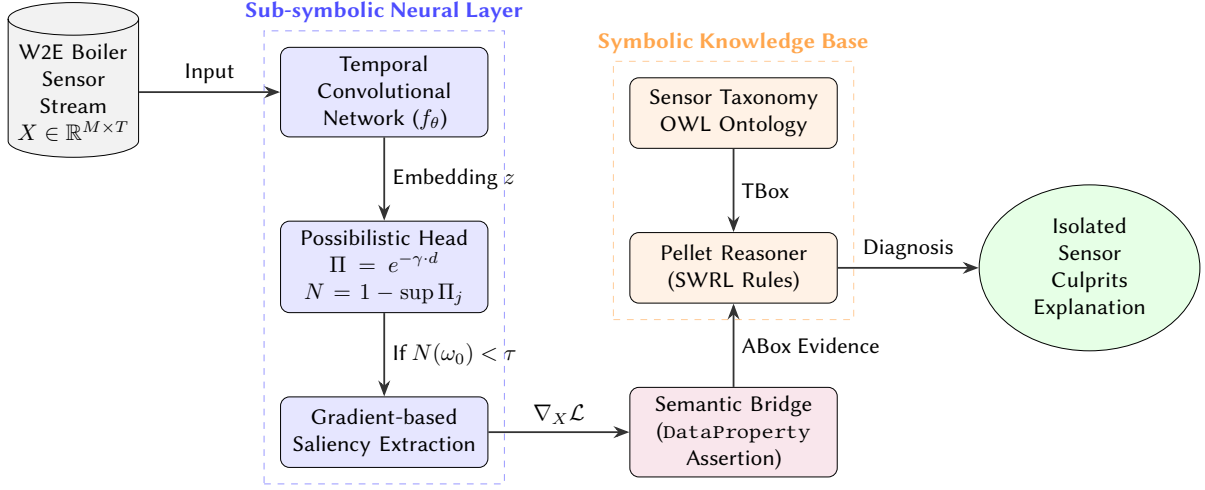


Figure 2: The end-to-end neuro-symbolic pipeline of the Energy Twin framework.

Early warning via possibilistic measures. The possibilistic measures provide a continuous characterization of operational deviations. In the running example (Figure 1b), the necessity measure $N(\omega_0)$ decreases before the possibility measure $\Pi(\omega_0)$, generating a conflict state prior to complete failure detection. In the retrospective experiments over historical W2E telemetry data, this transition appeared several hours before ($>24h$) the operational failure thresholds were reached. Since these observations are based on offline analysis, prospective evaluation in live industrial environments remains future work.

Explainability via neuro-symbolic evidence propagation. Beyond early warning detection, the framework provides a dual explanation mechanism connecting sub-symbolic neural evidence with symbolic semantic reasoning. First, sensor importance scores are computed over the multivariate sensor inputs, producing a ranked attribution of the variables contributing most strongly to the epistemic degradation state. As shown in Figure 3, these saliency scores localize the temporal variables responsible for the anomaly emergence. Second, the normalized attribution scores are injected into the OWL ontology as symbolic assertions linked to the physical boiler hierarchy. The Pellet reasoner subsequently evaluates SWRL rules over these assertions, isolating the specific sensors and chemical subsystems involved in the epistemic ignorance state. This neuro-symbolic explanation layer enables the system not only to detect uncertainty, but also to provide a structured and verifiable account of the physical evidence underlying the warning condition.

2.1. Implementation pipeline

The practical instantiation of the framework follows an end-to-end pipeline structured across a sub-symbolic neural network and a symbolic semantic model (Figure 2). The sequence operates as follows:

1. **Feature extraction:** The TCN processes the sliding window of multivariate time-series X , producing an internal representation of the sensor data.
2. **Epistemic assessment:** The possibilistic head maps the embedding to structural scores Π and N . If necessity drops below a tuned warning threshold τ , an early warning condition is registered.
3. **Explanation generation:** For any certified anomaly or ignorance state, the system calculates backward gradients $\nabla_X \mathcal{L}$ of the anomaly score relative to the inputs (see Figure 3a). This localizes which variables deviate most severely from normal (nominal) boundaries.
4. **Ontology mapping:** The normalized sensor attribution scores are parsed by an ontology mapping and injected into the OWL ontology as structural facts via the assertion of `tcnSaliencyScore` data properties.



(a) Sensor importance scores are ranked over the multivariate sensor streams.

(b) Ontology-based symbolic explanation obtained through SWRL reasoning.

Figure 3: Dual explanation mechanism of the Energy Twin framework. The statistical explanation layer identifies the temporal variables contributing most strongly to the anomaly state, while the symbolic reasoning layer isolates the physical sensor culprits through ontology-guided semantic inference.

- 5. Symbolic inference:** The Pellet reasoner runs execution checks on human-engineered SWRL rules (e.g., verifying if a given sensor’s attribution score exceeds physical tolerances defined inside the ontology structure), culminating in a verifiable list of root-cause sensor culprits (see Figure 3b).

3. Discussion and future work

In this paper, we introduced the Energy Twin, a possibilistic neuro-symbolic framework designed to address the critical vulnerabilities of standard predictive maintenance systems in W2E facilities. By replacing the traditional softmax with a formal possibilistic head, our architecture provides a theoretically grounded mechanism to quantify epistemic ignorance and anticipatory conflict. This allows the system to emit anticipatory early warnings without forcing unwarranted, overconfident classifications on unobserved data. Furthermore, our framework establishes a strict neuro-symbolic bridge: continuous sub-symbolic saliency gradients are translated into discrete `DataProperty` assertions. These assertions are evaluated against a formalized OWL ontology utilizing the Owlready2 framework and the Pellet reasoner, enabling the deterministic deduction of localized sensor root-causes based on immutable physical hierarchies.

This architecture contrasts with the current state-of-the-art (SOTA) in deep learning anomaly detection. Contemporary leading approaches, such as attention-based anomaly transformers [9] or generative reconstruction frameworks [10], rely heavily on deterministic forecasting errors or rigid thresholding criteria. While highly accurate within their training distributions, these SOTA mechanisms treat all out-of-distribution variations as equivalent anomalies. They lack the epistemic vocabulary to logically differentiate between an early-stage known degradation (conflict) and a completely novel environmental transient (ignorance). By explicitly decoupling possibility from necessity, the Energy Twin framework preserves this vital ambiguity, aligning closely with the safety-critical requirements of industrial Knowledge Representation systems.

Moving forward, our research will focus on two primary theoretical extensions. First, we aim to investigate possibilistic uncertainty at a finer architectural granularity by mapping necessity and possibility measures to specific subsets of physical sensors rather than the global latent manifold, providing localized uncertainty metrics. Second, we intend to close the neuro-symbolic loop by integrating the

hierarchical inferences generated by the OWL ontology directly into the neural module's loss function. This will allow the ontological rules to dynamically guide the latent space representation during training, strictly penalizing physical impossibilities and ensuring the neural backbone natively adheres to the thermodynamic laws governing the W2E boiler.

Data availability. The industrial telemetry data and prototype infrastructure used in this work are not publicly available due to operational and confidentiality constraints associated with the collaborating W2E facility. Research access to selected materials may be considered upon reasonable request to the corresponding author.

Declaration on Generative AI. During the preparation of this work, the author(s) used GeminiAI, ChatGPT, and Copilot in order to: Grammar and spelling check, Paraphrase and reword, Improve writing style. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] Boiler Tube Failures in Municipal Waste-To-Energy Plants: Case Histories, volume CORROSION 1995 of *NACE International Annual Conference*, 1995. URL: <https://doi.org/10.5006/C1995-95561>. doi:10.5006/C1995-95561.
- [2] Z. Cui, J. Xu, W. Liu, G. Zhao, S. Ma, Data-driven modeling-based digital twin of supercritical coal-fired boiler for metal temperature anomaly detection, *Energy* 278 (2023) 127959. URL: <https://www.sciencedirect.com/science/article/pii/S0360544223013531>. doi:<https://doi.org/10.1016/j.energy.2023.127959>.
- [3] D. Wang, L. Jiang, M. Kjellander, E. Weidemann, J. Trygg, M. Tysklind, A novel data mining framework to investigate causes of boiler failures in waste-to-energy plants, *Processes* 12 (2024). URL: <https://www.mdpi.com/2227-9717/12/7/1346>. doi:10.3390/pr12071346.
- [4] S. Lu, Y. Wang, L. Sheng, L. He, A. Zheng, J. Liang, Out-of-distribution detection: A task-oriented survey of recent advances, *ACM Comput. Surv.* 58 (2025). URL: <https://doi.org/10.1145/3760390>. doi:10.1145/3760390.
- [5] J. Yang, K. Zhou, Y. Li, Z. Liu, Generalized out-of-distribution detection: A survey, *Int. J. Comput. Vis.* 132 (2024) 5635–5662. URL: <https://doi.org/10.1007/s11263-024-02117-4>. doi:10.1007/s11263-024-02117-4.
- [6] W. Liu, X. Wang, J. Owens, Y. Li, Energy-based out-of-distribution detection, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), *Advances in Neural Information Processing Systems*, volume 33, Curran Associates, Inc., 2020, pp. 21464–21475. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/f5496252609c43eb8a3d147ab9b9c006-Paper.pdf.
- [7] D. Dubois, H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Springer US, 2012. URL: <https://books.google.se/books?id=GEvUBwAAQBAJ>.
- [8] Anonymous, Epimax: Possibilistic learning from raw logits via energy-based margins, 2026. Under review.
- [9] J. Xu, H. Wu, J. Wang, M. Long, Anomaly transformer: Time series anomaly detection with association discrepancy, in: *International Conference on Learning Representations*, 2022. URL: https://openreview.net/forum?id=LzQQ89U1qm_.
- [10] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, M. Salehi, Deep learning for time series anomaly detection: A survey, *ACM Comput. Surv.* 57 (2024). URL: <https://doi.org/10.1145/3691338>. doi:10.1145/3691338.