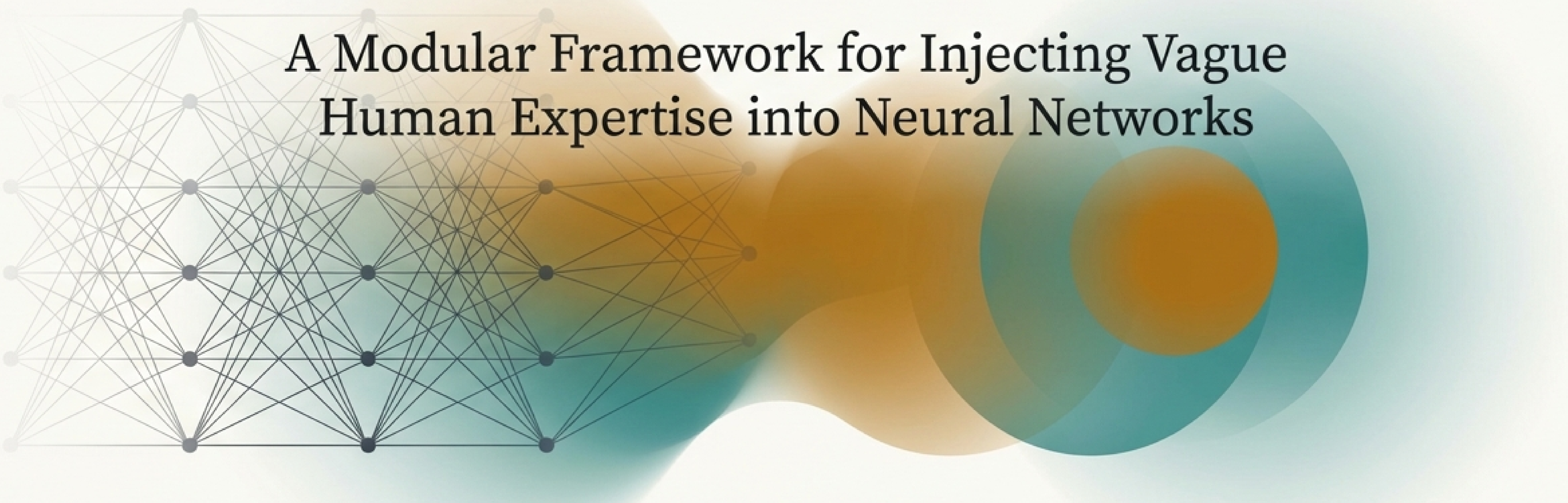
Fuzzy-Knowledge-Informed Machine Learning (FKIML)

A Modular Framework for Injecting Vague Human Expertise into Neural Networks



Amit K. Shukla[1], Vagan Terziyan[2], Olena Kaikova[2]
[1] University of Vaasa, FINLAND
[2] University of Jyväskylä, FINLAND

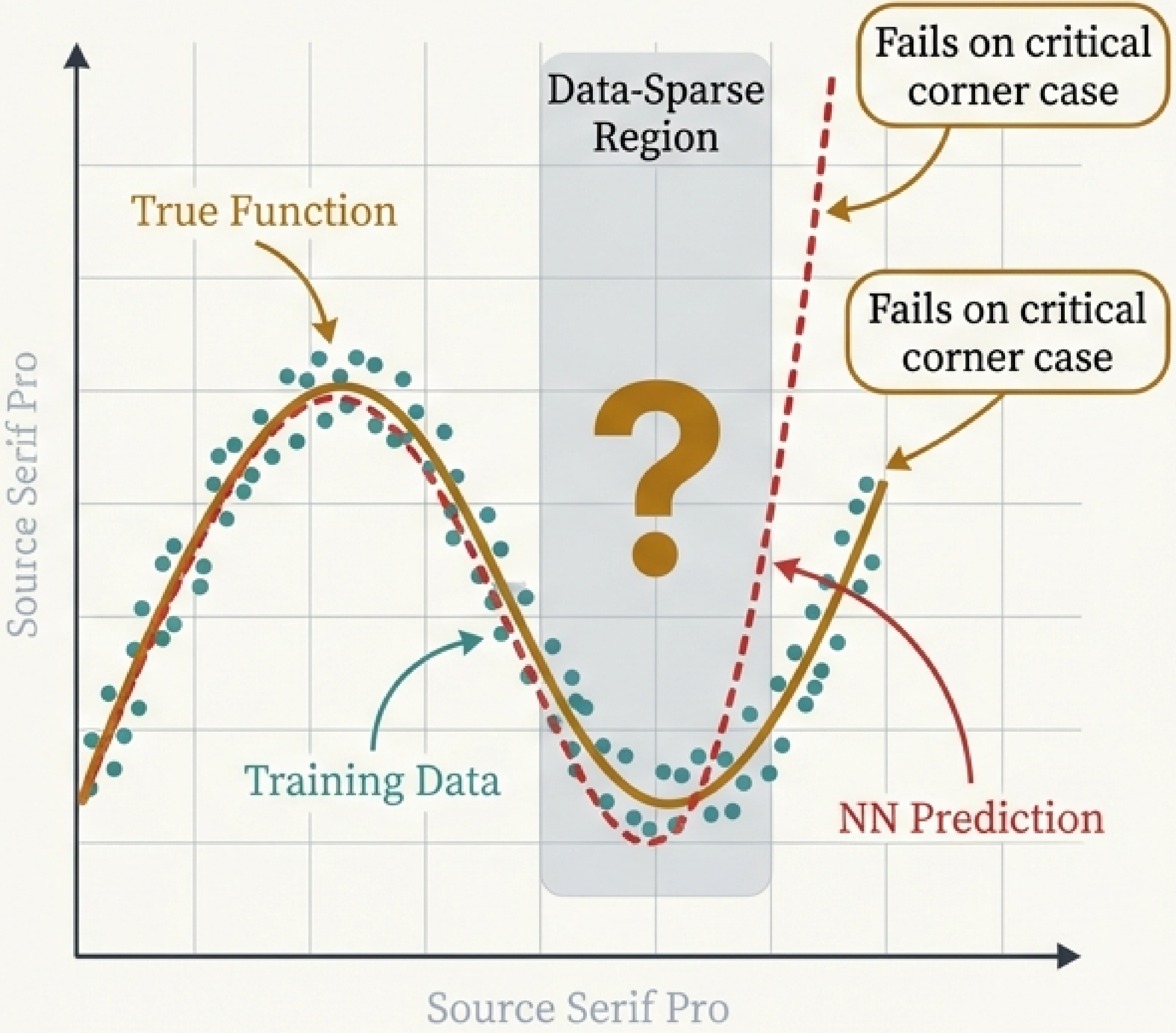
Purely Data-Driven Models Are Brittle Where Knowledge is Critical

Despite their power, standard neural networks face persistent challenges when trained exclusively on empirical data, especially in high-stakes domains.

Data Scarcity: Models often fail on rare but critical events (e.g., safety-critical “corner cases”) that are underrepresented in training data.

Lack of Interpretability: “Black box” models can drift into unrealistic or unsafe decision regions without adhering to known domain principles.

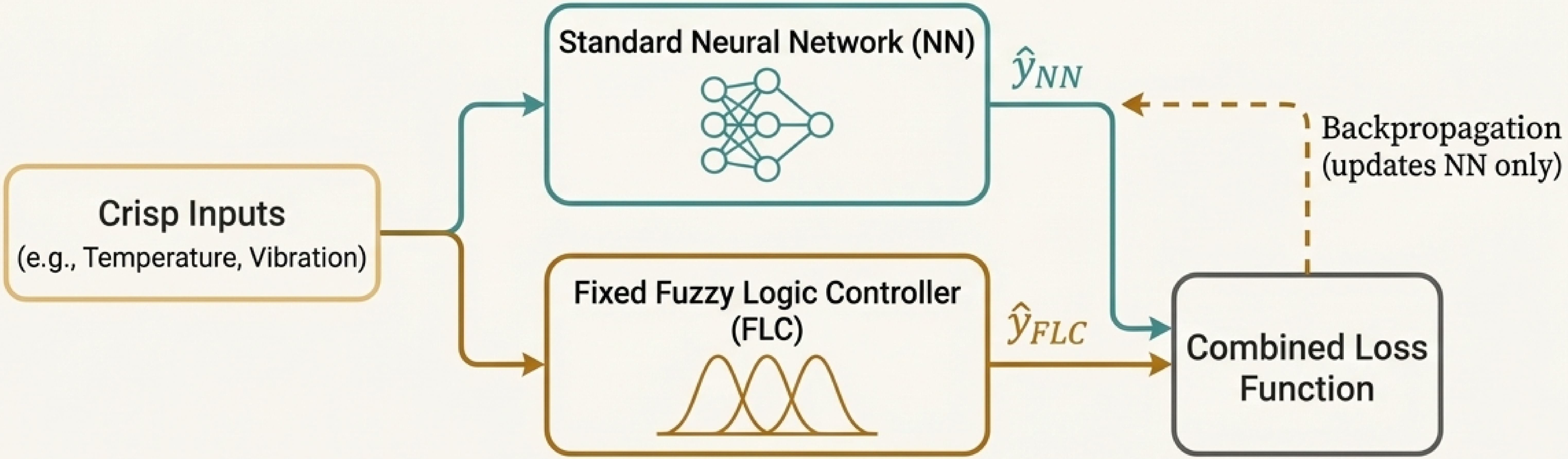
The Need for Expertise: How can we inject vague, heuristic, and linguistically-expressed human knowledge (“*high temperature*”, “*moderate risk*”) into a rigid numerical framework?



NotebookLM

FKIML: A Standard Neural Network Guided by an External Fuzzy "Teacher"

Instead of embedding fuzzy logic inside the network, FKIML uses a conventional Neural Network regularized by an external, fixed Fuzzy Logic Controller (FLC). The FLC acts as a source of prior knowledge, and the NN learns to align its predictions with both the data and the FLC's expert rules.


Standard Neural Network (NN)
$\hat{y}_{NN}$
Backpropagation (updates NN only)
Crisp Inputs
(e.g., Temperature, Vibration)
Fixed Fuzzy Logic Controller (FLC)
$\hat{y}_{FLC}$
Combined Loss Function

FKIML

Modular design. The NN remains a standard architecture. The FLC is a separate, non-trainable knowledge source.

ANFIS (Classic Neuro-Fuzzy)

Integrated design. Fuzzy rules and membership functions are embedded as trainable layers within the network architecture.

Situating FKIML in the Knowledge-Informed Machine Learning Landscape

KIML integrates domain knowledge to improve model robustness and efficiency. FKIML extends this paradigm to handle knowledge that is vague, linguistic, and imprecise—a common form of human expertise.

| Approach	| Type of Knowledge	| Enforcement	| Example |
|PINNs	Hard Physics (PDEs)	Hard/Soft Penalty	|Heat Equation |
|Logic-Informed	Symbolic Logic	Logical Loss	|First-Order Logic Rules |
|Bayesian Deep Learning	Probabilistic Priors	Distributional	| Gaussian Priors on Parameters |
| FKIML (Proposed)	**Fuzzy / Vague Knowledge**	**| Soft Constraints**	**| Linguistic Rules, Vague Beliefs |**

The Mechanism: An Integrated Loss Function Balancing Data and Knowledge

Data Fidelity. The standard supervised loss. Measures how well the NN output $\hat{y}\_{NN}$ matches the ground truth y_true.

Knowledge Alignment. The fuzzy regularization term. Measures the deviation between the NN output $\hat{y}\_{NN}$ and the FLC's expert output $\hat{y}\_{FLC}$.

$$L_{total} = L_{data} + \lambda * L_{fuzzy}$$

The Trade-Off Parameter. A hyper-parameter that controls the balance. λ=0 is a plain NN. A larger λ enforces stricter adherence to the fuzzy rules.

Gradients flow from both loss components, 'nudging' the NN's parameters to learn from both the data and the softly-encoded expert rules.

Two Complementary Strategies for Knowledge Integration

Option 1: Knowledge-Alignment Regularizer

The NN is trained to match the ground truth, while a second loss term penalizes it for disagreeing with the FLC. The FLC acts as a soft constraint or teacher.

$L_{total} = L_{NN}(\hat{y}_{NN}, y_{true}) + \lambda * L_{fuzzy}(\hat{y}_{NN}, \hat{y}_{FLC})$

Use Case:

Ideal when FLC encodes safety margins or heuristics that must be respected, even if they aren't perfectly accurate.

Option 2: Pseudo-Label Mixing

A blended target is created by interpolating between the ground truth and the FLC's output. The NN is then trained on this single 'pseudo-label'.

$y_{target} = (1\text{-}\beta) * y_{true} + \beta * \hat{y}_{FLC}$ where $\beta = \text{map}(\lambda)$.

Then: $L_{total} = \text{MSE}(\hat{y}_{NN}, y_{target})$

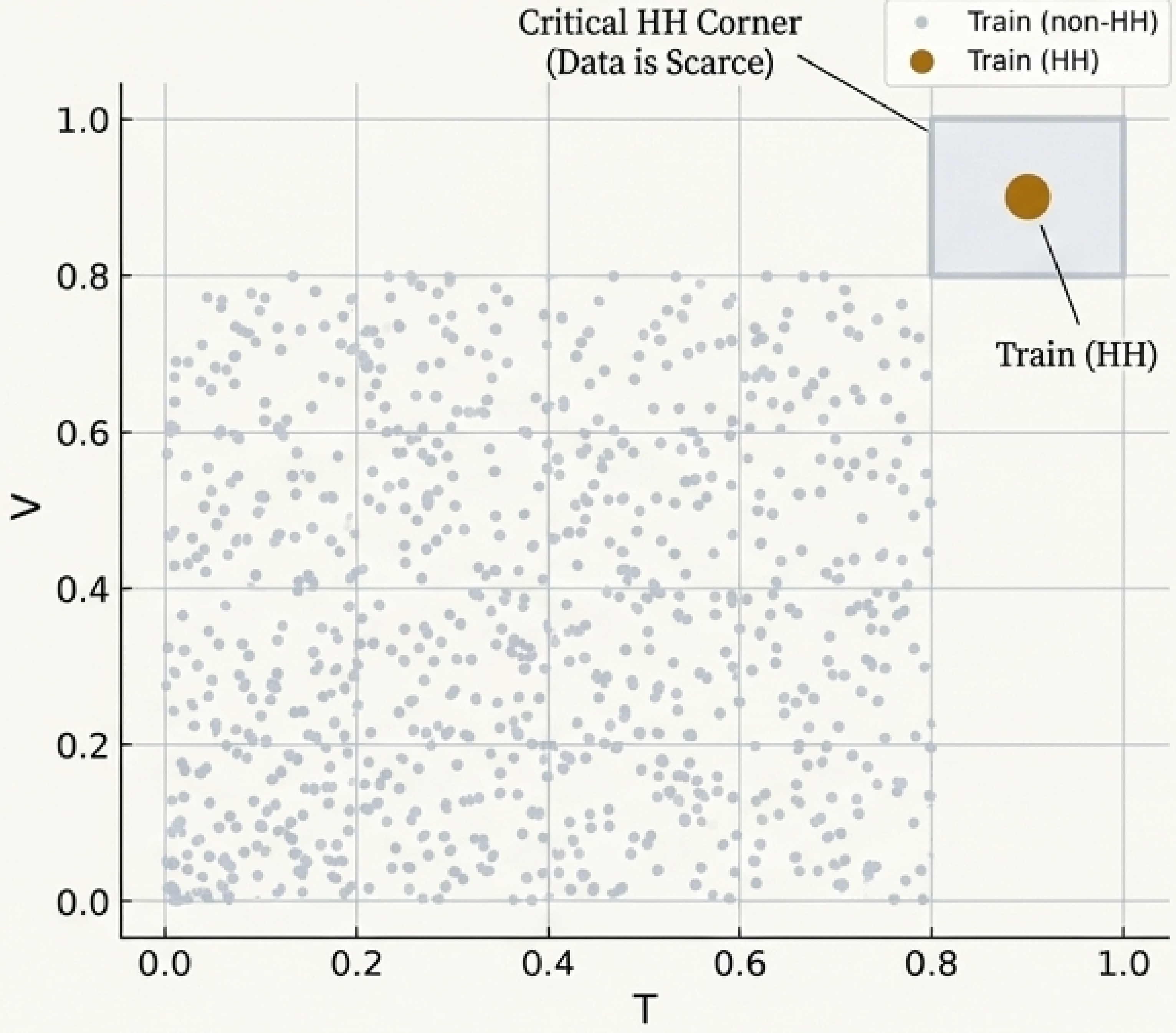
Use Case:

Offers a very direct and interpretable way to balance trust between empirical data and expert rules.

Proof-of-Concept: Predicting Machine Risk with Sparse, Critical Data

Experiment Setup

- **Task:** A synthetic dataset where a model must predict machine `Risk’ from `Temperature (T)` and `Vibration (V)`.
- **The Challenge:** The "High-High" (HH) corner (where T > 0.8 and V > 0.8) represents a critical, high-risk state. However, training data in this region is intentionally made sparse.
- **The Hypothesis:** A plain NN will underfit the HH region due to lack of data, while FKIML can use a fuzzy rule to improve performance on these critical cases.



Capturing Expert Knowledge in a Multi-Rule Fuzzy Logic Controller

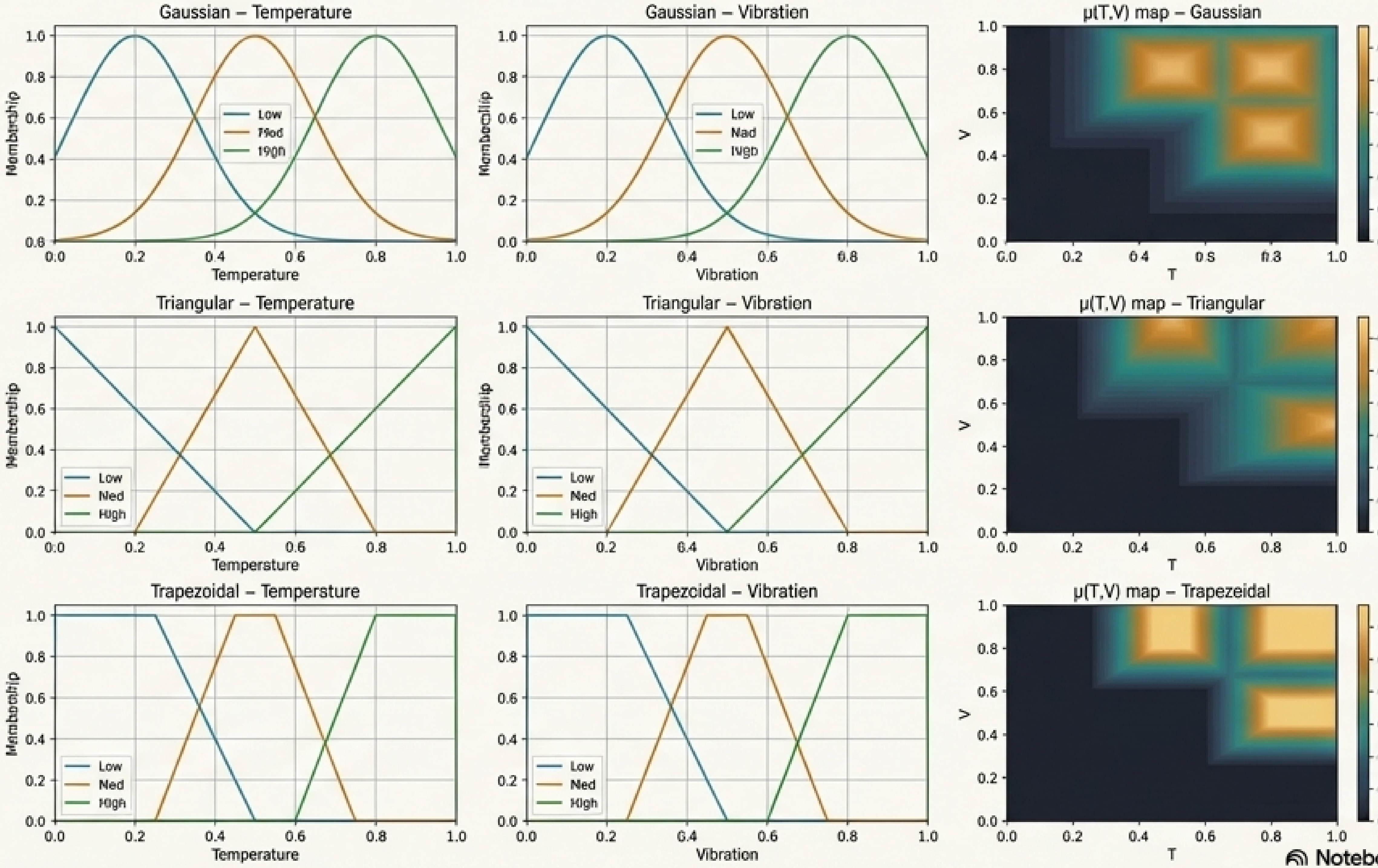
The Fuzzy Rule Base

Expert intuition is encoded into a Mamdani-type FLC with three rules:

- **Rule 1:** IF T is **High** AND V is **High**, THEN Risk is **High**.
- **Rule 2:** IF T is **Medium** AND V is **High**, THEN Risk is **Medium**.
- **Rule 3:** IF T is **High** AND V is **Medium**, THEN Risk is **Medium**.

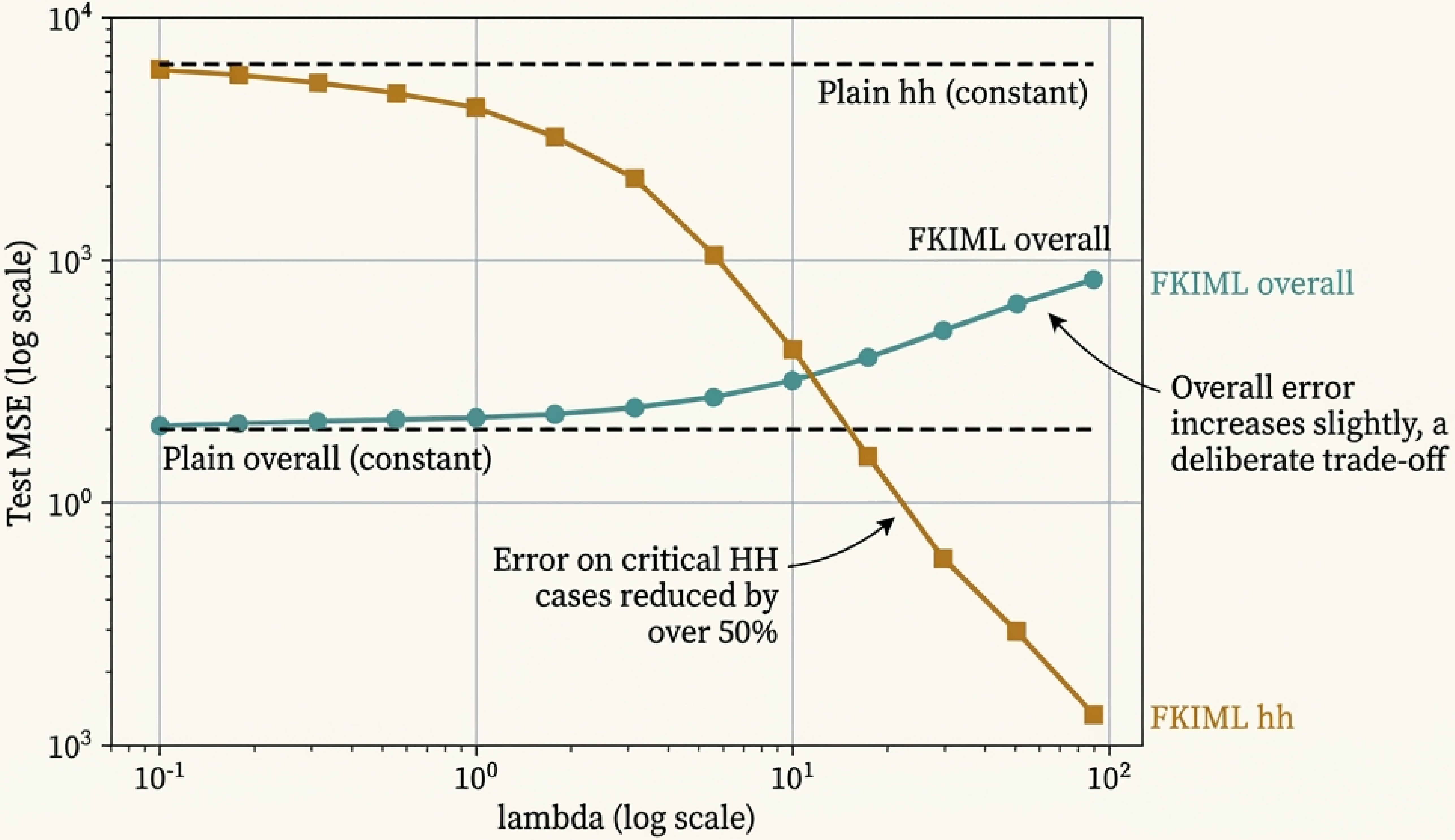
Membership Functions (MFs)

Linguistic variables like ‘Medium’ and ‘High’ are defined by mathematical functions. We evaluated three families:



The Verdict: FKIML Sacrifices Minor Global Accuracy for Major Gains on Critical Cases

As the knowledge-weighting parameter λ` increases, the error on the critical “High-High” (HH) cases drops significantly, while the overall error on all data increases only modestly. This demonstrates a controllable trade-off.

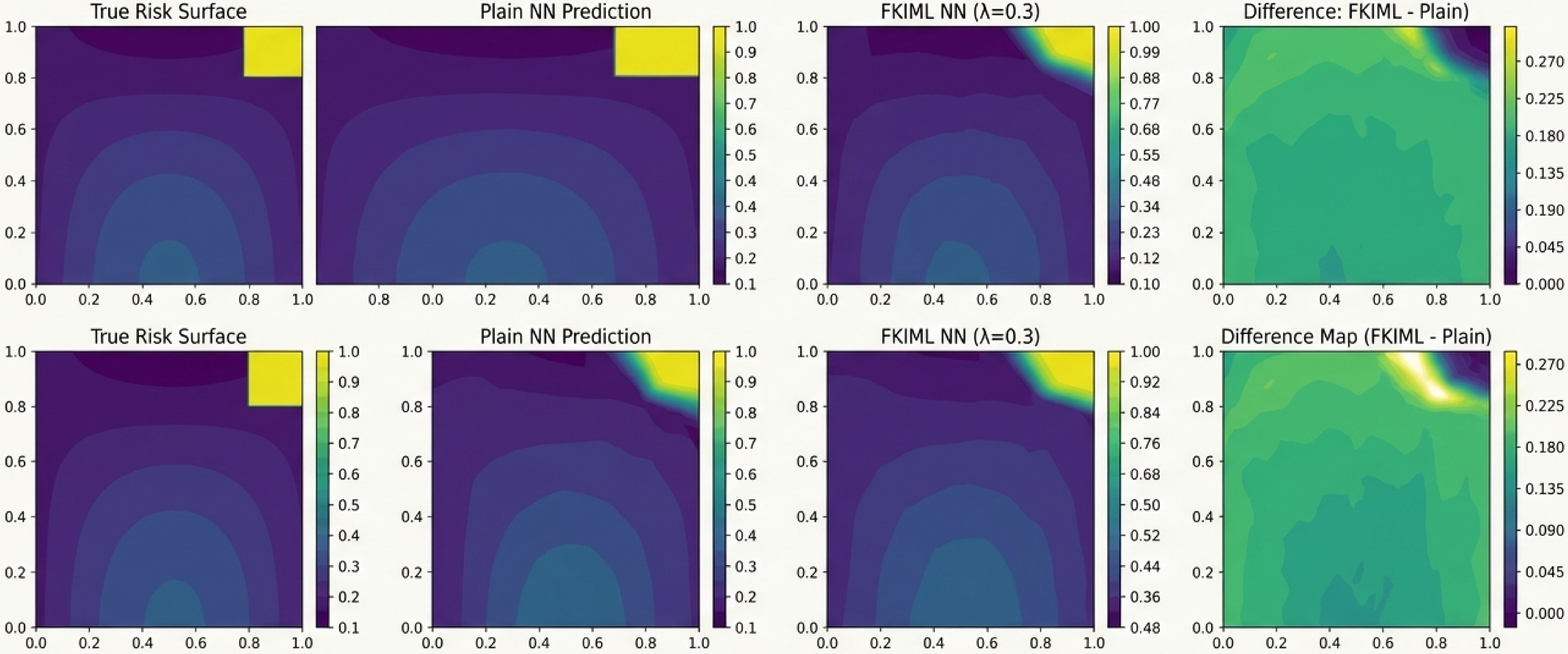


Overall error increases slightly, a deliberate trade-off

Error on critical HH cases reduced by over 50%

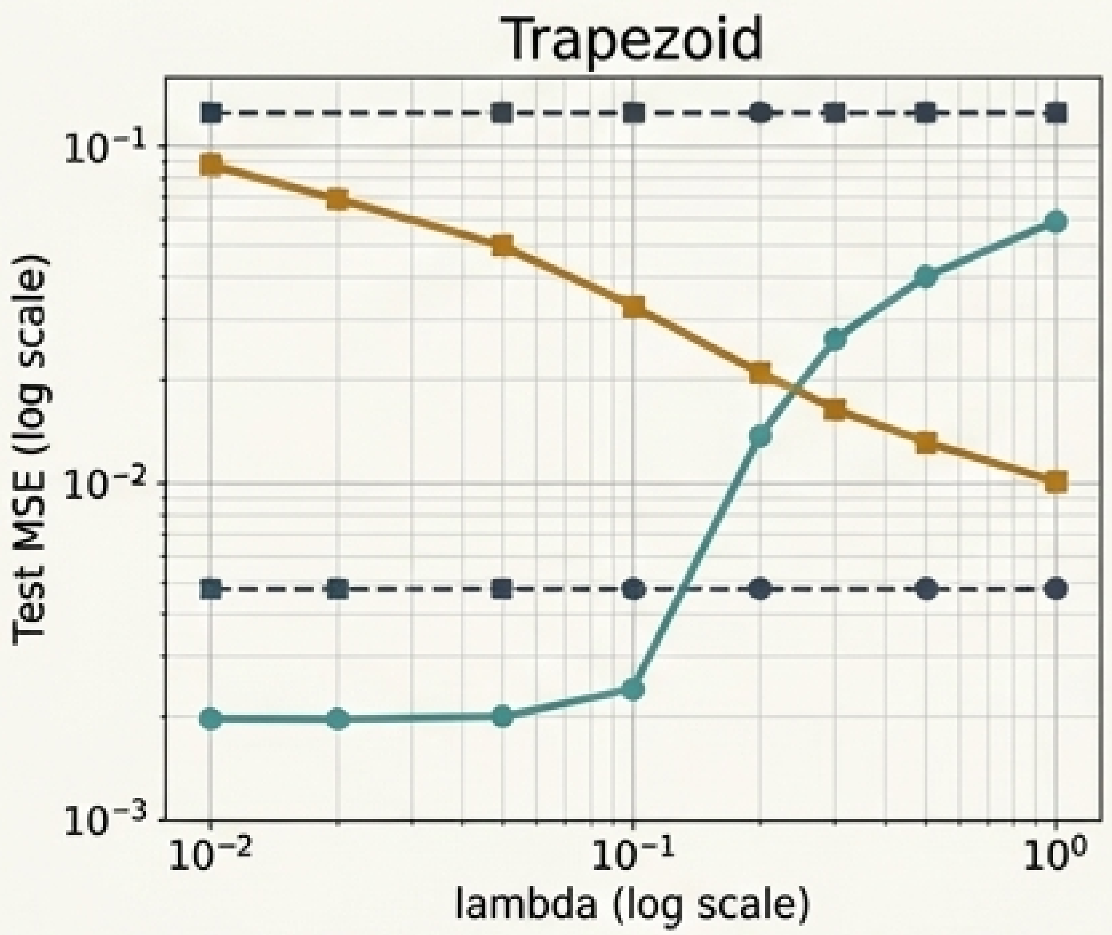
Visualizing the Impact: FKIML Learns the Critical Risk Corner

A plain NN, trained on sparse data, fails to capture the sharp increase in risk in the HH corner. FKIML, guided by the fuzzy rule, correctly learns this critical behavior.

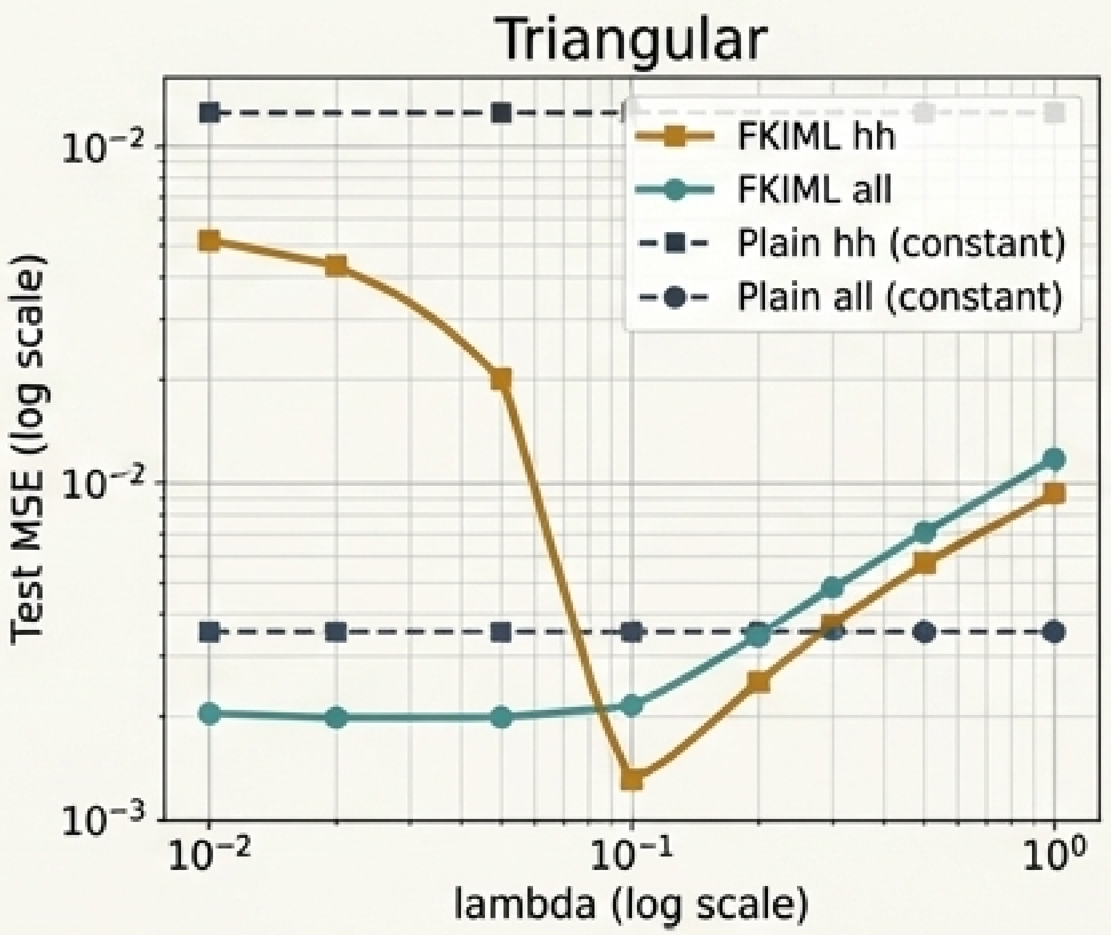


Membership Function Design Is Key to a Favorable Trade-Off

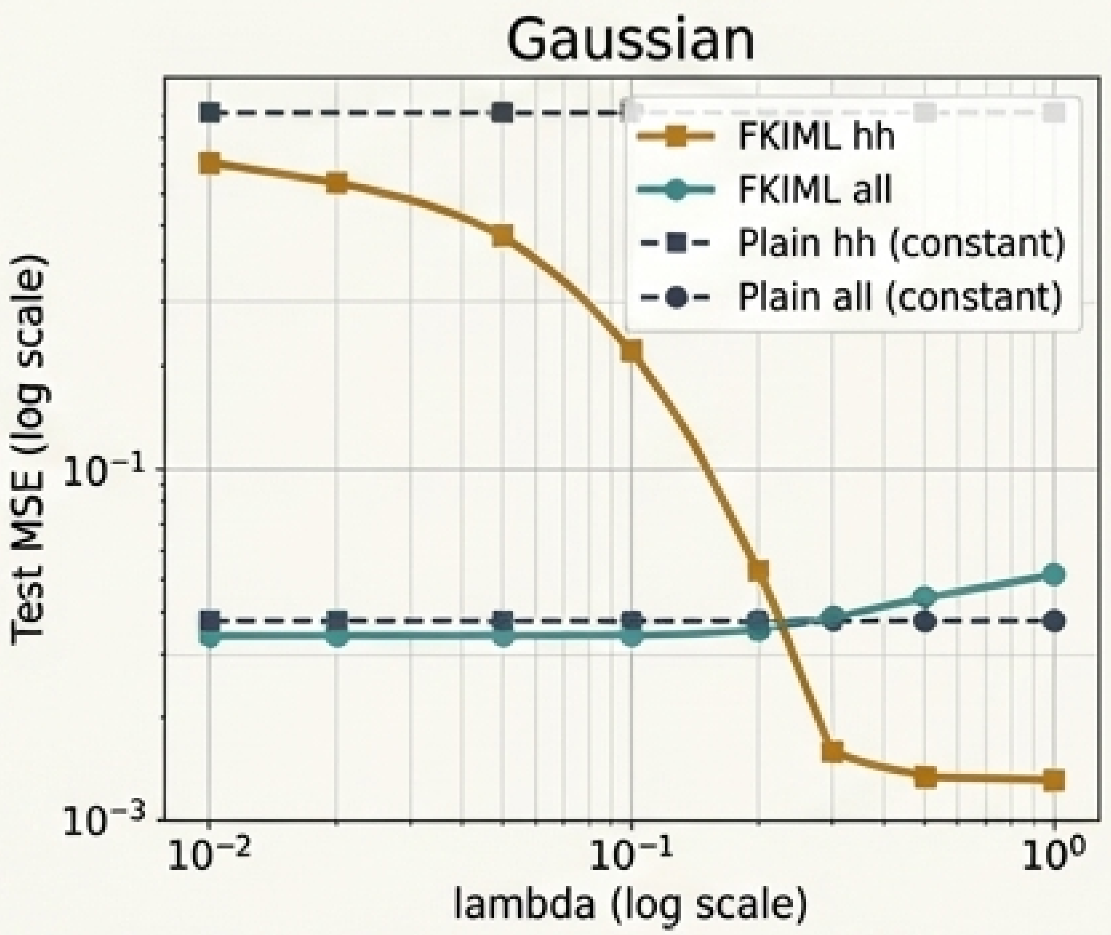
The shape of the membership functions (MFs) impacts the learning dynamics. Piecewise-linear (Triangular, Trapezoidal) and smooth (Gaussian) MFs provide different trade-offs between global accuracy and rule consistency.



Good for when HH performance is the only priority.



Offers a balanced ‘sweet spot’ with strong early gains.



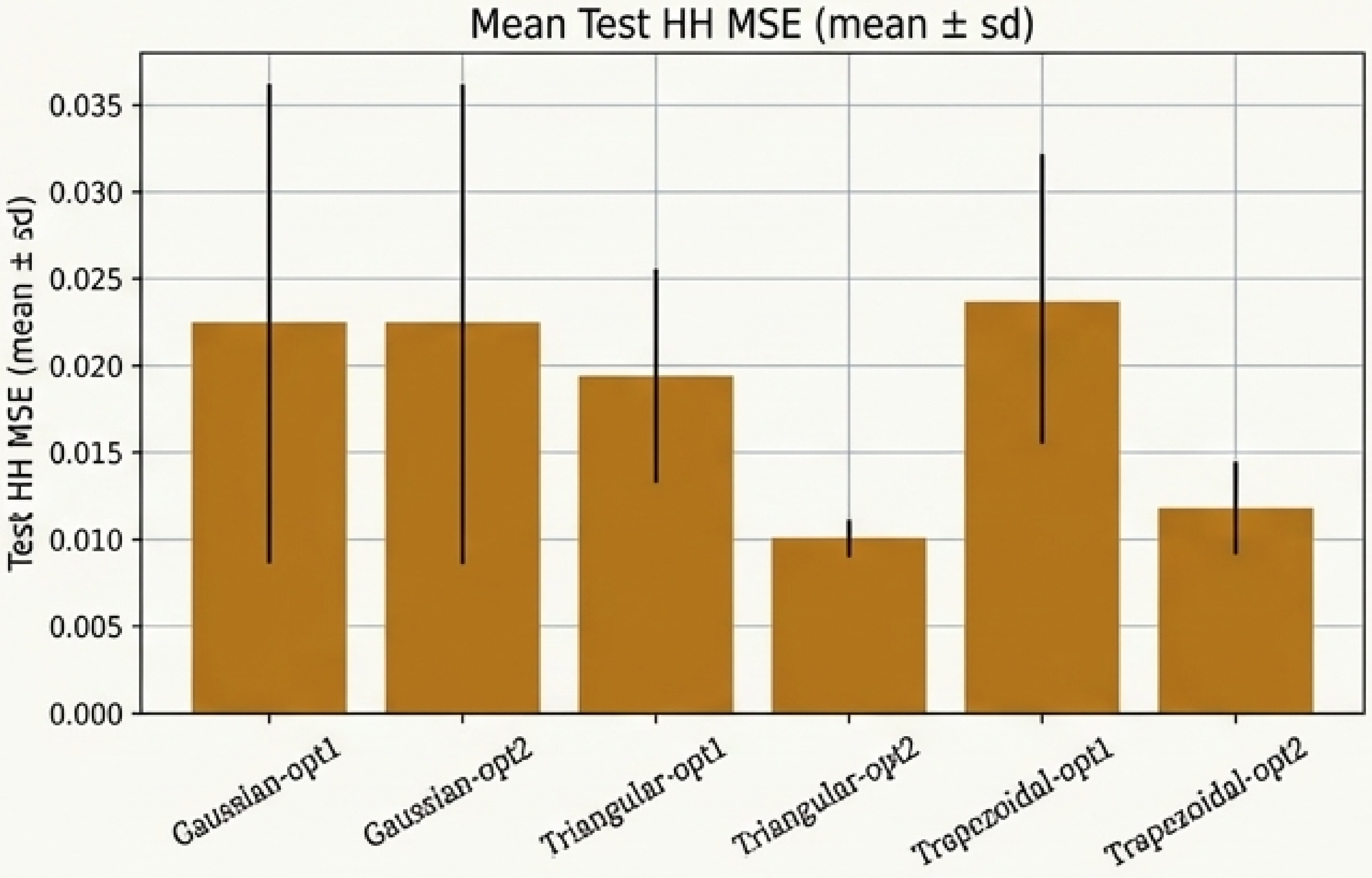
Provides the most focused guidance for the best trade-off in this experiment.

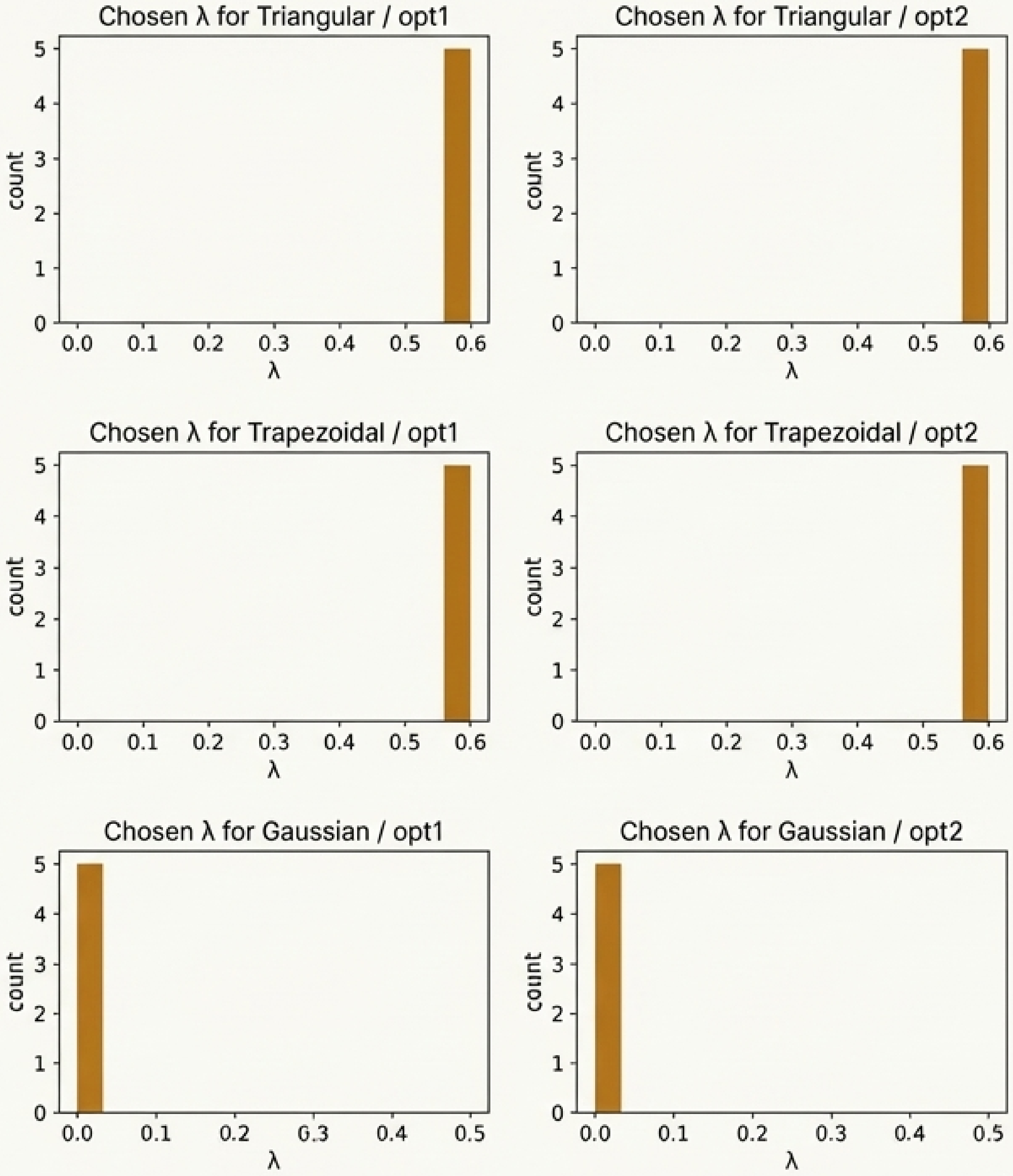
In these experiments, Triangular and Trapezoidal MFs provided the most promising trade-offs, showing strong improvement on critical cases while remaining robust.

NotebookLM

Robustness Check: Multi-Seed Analysis Confirms Consistent Gains

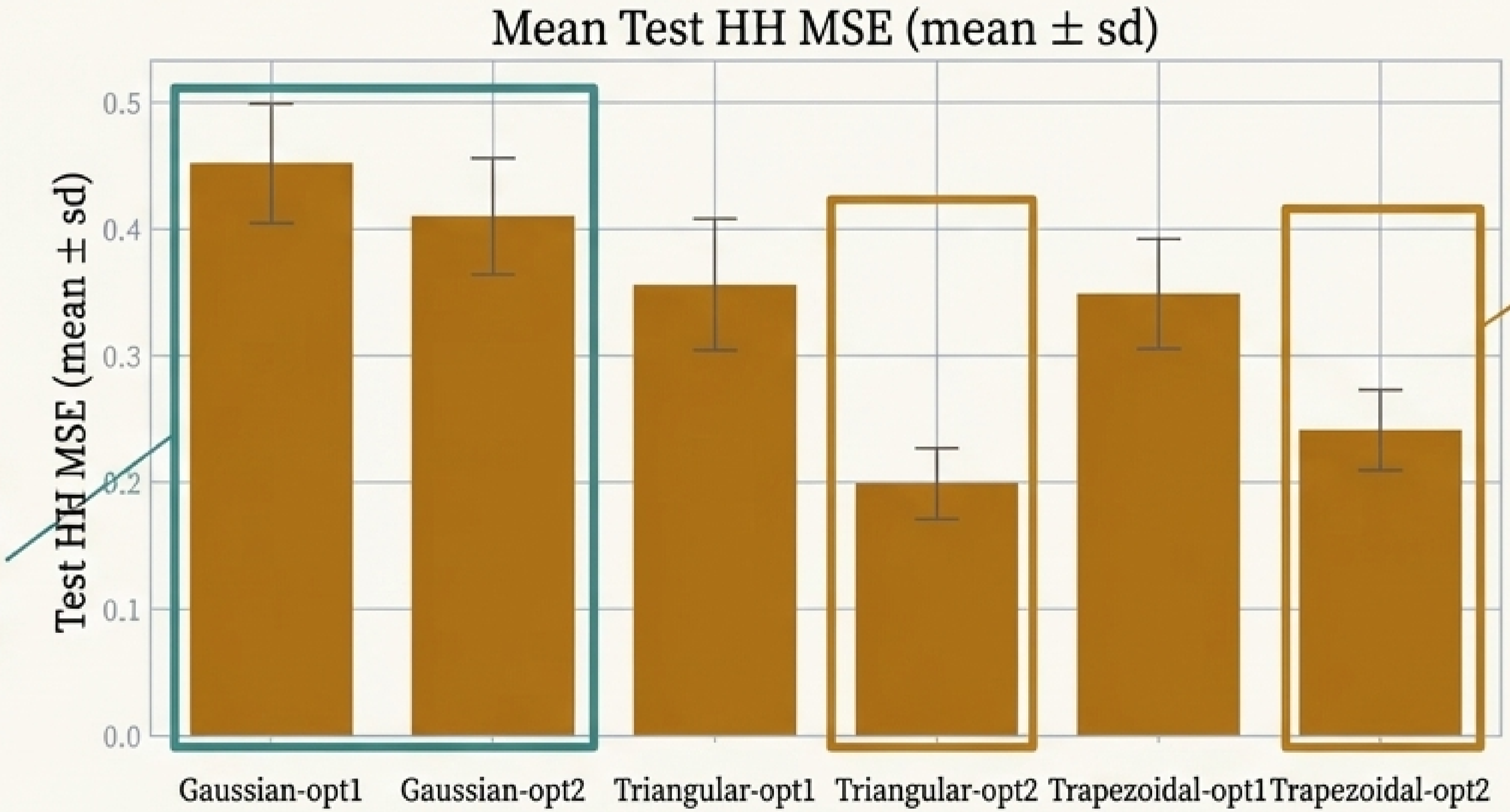
Protocol: To ensure results are statistically reliable, the entire experiment—from data sampling to λ selection via a validation set—was repeated across five independent random seeds. The improvement in HH-region performance is consistent, and the optimal λ is consistently non-zero.





Loss Formulation Matters: Pseudo-Label Mixing (Opt 2) Excels for Critical Cases

- **Option 1 (Consistency Loss):** Provides a gentle regularization but is less effective at forcing the model to learn the HH corner, especially with Gaussian MFs.
- **Option 2 (Pseudo-Label Mixing):** More aggressively steers the NN towards the FLC's target in rule-active regions. This consistently results in lower HH error, particularly with Triangular and Trapezoidal MFs.



Gaussian MFs were less effective in this multi-rule setup, with the selection process often defaulting to λ=0 (the plain NN).

Option 2 consistently yields the lowest error on the critical HH region across multiple seeds.

Summary: FKIML is a Simple, Modular, and Powerful Framework

Modular & Simple: Integrates with any standard NN via a single auxiliary loss term. No complex architectural changes are needed.

Interpretable by Design: Expert knowledge is kept in a separate, fixed FLC, remaining transparent and auditable, unlike in ‘black-box’ integrated systems.

Tunable Trade-Off: The `λ parameter provides explicit, granular control over the balance between fitting the data and adhering to expert rules.

Proven Effective: Demonstrates statistically significant improvement in robustness on rare, knowledge-critical ‘corner cases’ where data-driven models typically fail.

A Clear Alternative to ANFIS: Offers a more flexible and modern approach to blending neural and fuzzy systems, compatible with deep learning pipelines.

The Horizon: Applications and Future Research

Potential Applications

FKIML provides a promising bridge between imprecise human reasoning and numerical AI in domains rich with heuristic knowledge.

Healthcare: Integrating clinical guidelines (*“older patients usually require…”*).

Industrial Control: Encoding soft safety rules (*“avoid high pressure when possible”*).

Environmental Modeling: Capturing expert assessments (*“algae growth tends to accelerate…”*).

Reinforcement Learning: Using fuzzy rewards to reflect vague preferences.

Open Research Questions

- How can we best compose and weight multiple, potentially conflicting fuzzy rules?
- Can FKIML frameworks be made fully architecture-agnostic (e.g., for Transformers, GNNs)?
- How can we develop adaptive λ schedules for more efficient training?