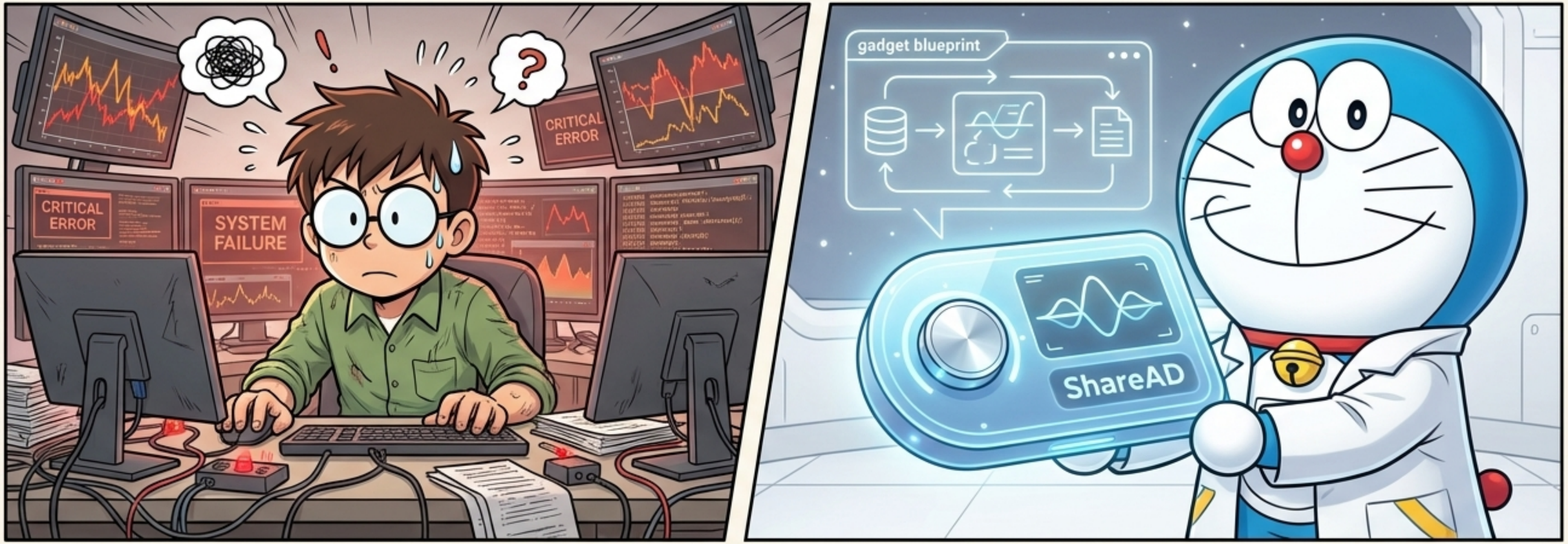
The Anomaly Detection Headache: Why Modern Systems Overwhelm Today's Tools

Introducing ShareAD: A "Pre-Train-and-Align" Framework from the Future



Based on: "On the Practicability of Deep Learning Based Anomaly Detection for Modern Online Software Systems: A Pre-Train-and-Align Framework," *ACM Transactions on Software Engineering and Methodology*, Vol. 34, 2025.

“Our Story” in Doraemon Blue

The Operator’s Dilemma

- The overwhelming challenge of anomaly detection in large-scale systems.
- Identifying the three critical “Practicality Gaps.”

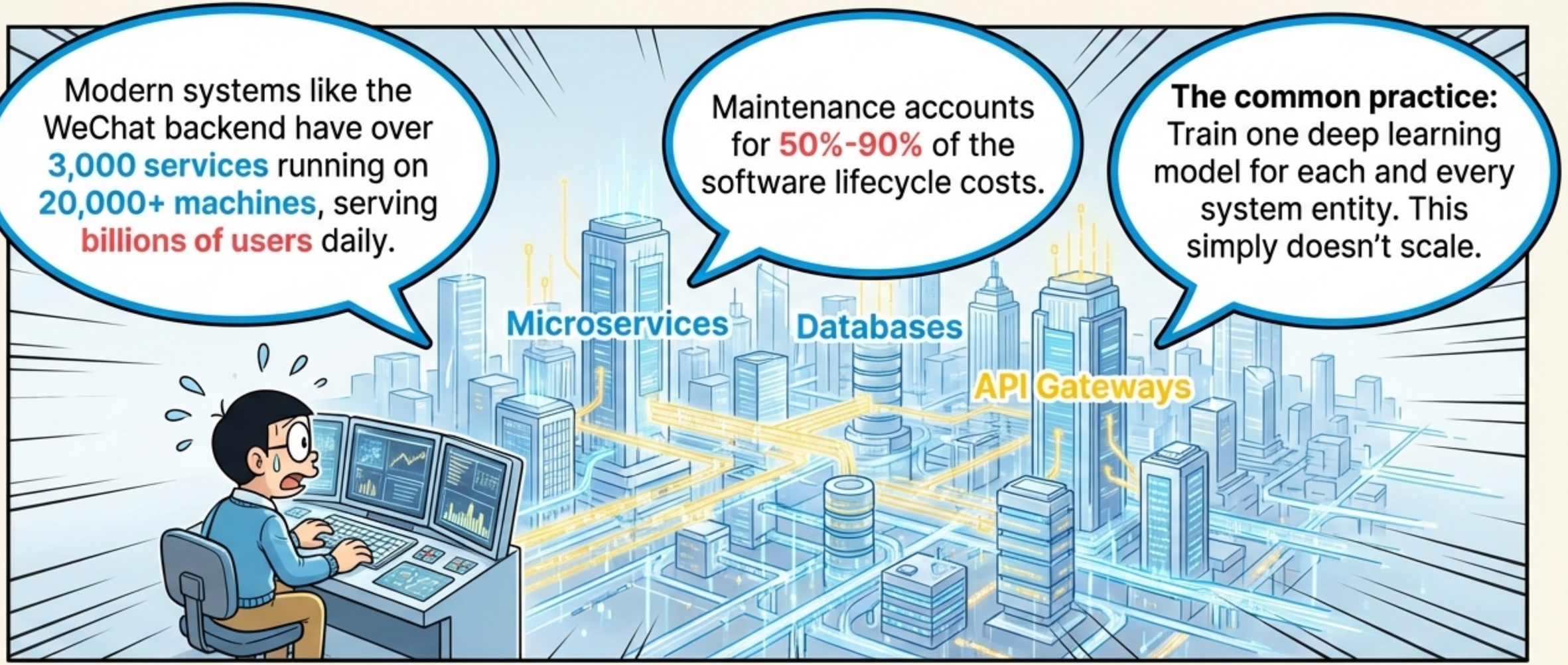
A Gadget from the Future

- Introducing ShareAD: The “Pre-Train-and-Align” paradigm.
- Deconstructing the gadget: The Pre-Train Engine and the Alignment Dial.

Proof in Action

- Putting ShareAD to the test: Experimental results.
- Demonstrating superior performance, efficiency, and practicality.

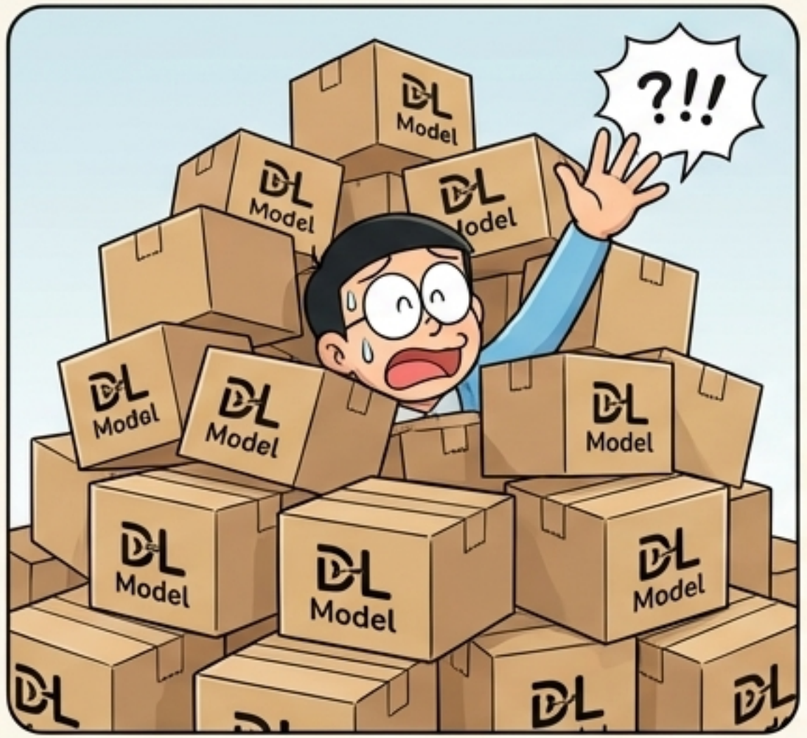
The Problem: Modern Systems are Too Big for Old Methods



This **"one-for-one" modeling approach**, while logical for small systems, creates massive practical problems at scale.

Three “Practicality Gaps” That Stop Deep Learning from Working in the Real World

The Scalability Gap



One-for-one modeling creates a huge cost for storing and running thousands of heavy deep learning models. Online inference becomes a critical bottleneck.

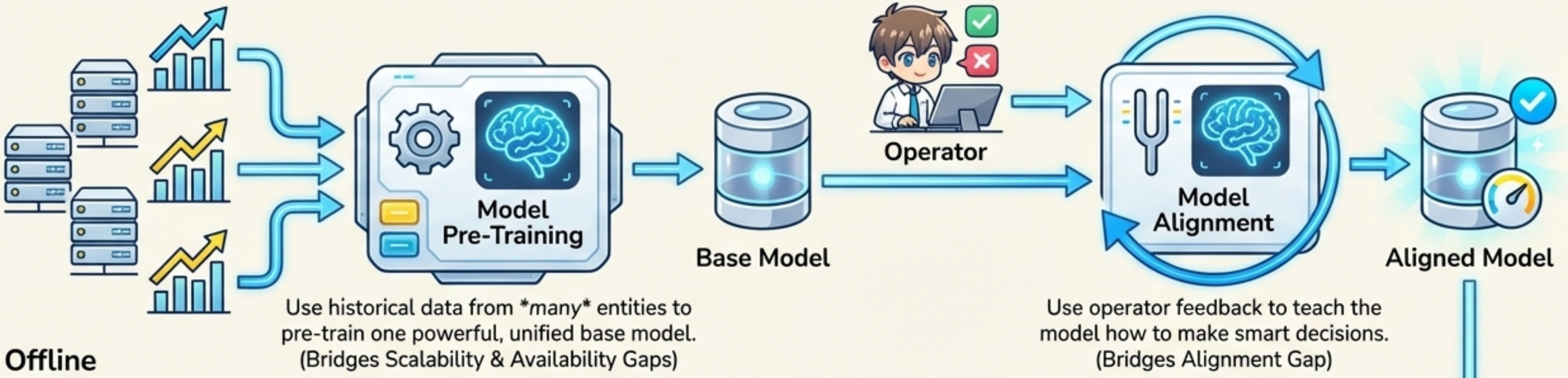
The Availability Gap

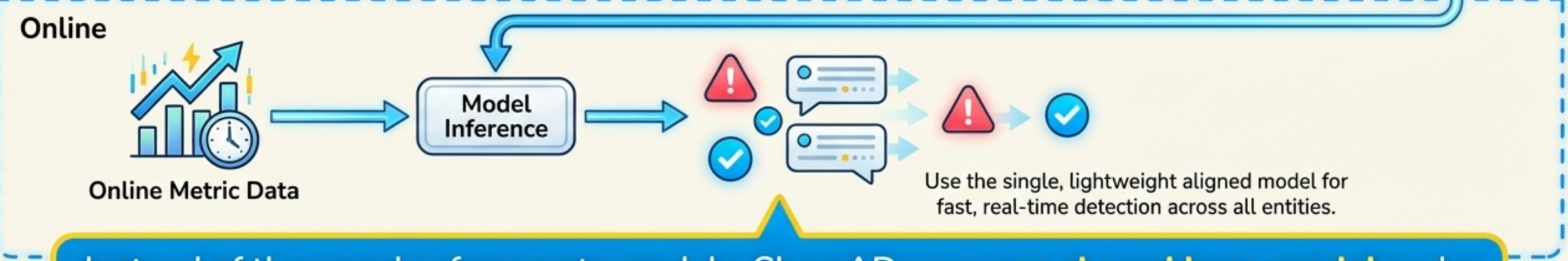
Modern systems are dynamic. Short-lived components like containers don't have enough historical data to train a dedicated model, leaving them unmonitored.

The Alignment Gap

Models output ambiguous **anomaly scores**, but operators need clear **decisions**. Existing thresholding methods fail to effectively align with operator preference.

The Solution: A New “Pre-Train-and-Align” Paradigm

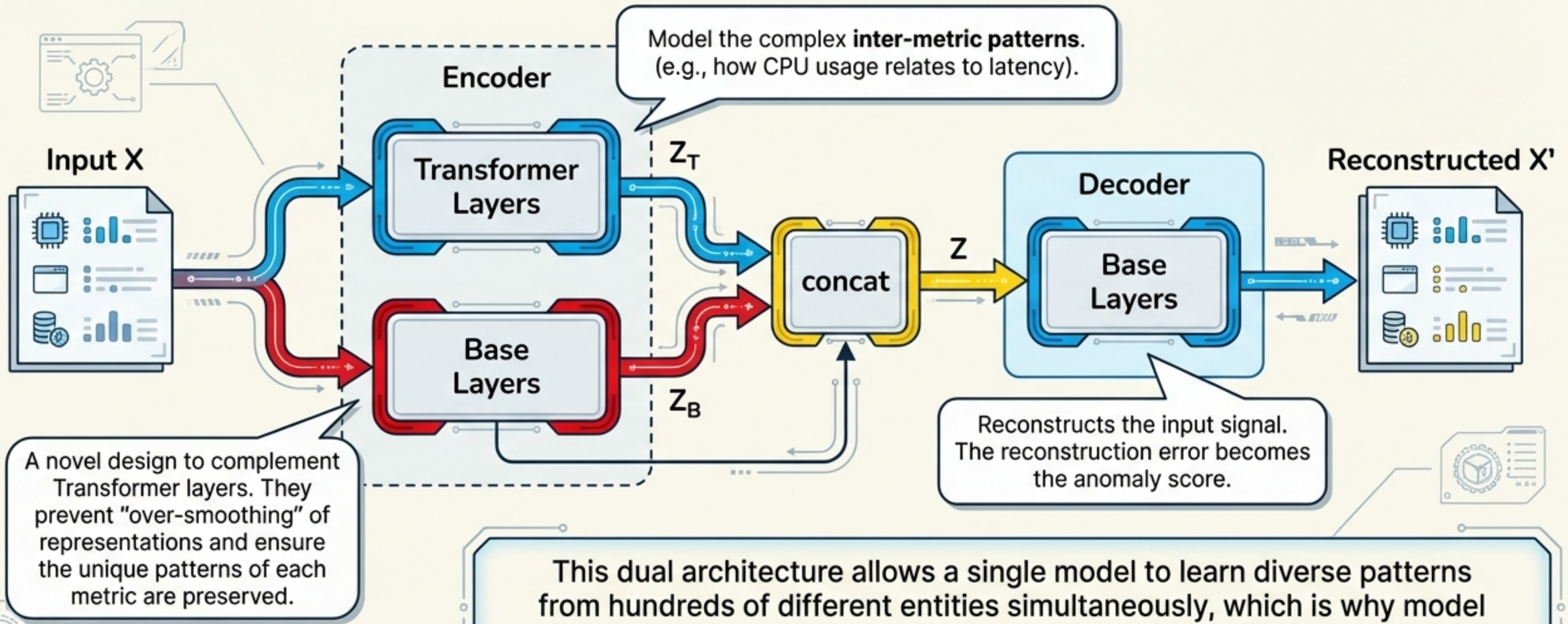




Instead of thousands of separate models, ShareAD uses **one shared base model** and a **lightweight aligner**, making it scalable, available, and aligned with operator needs.

Gadget Blueprint Part 1: The Pre-Train Engine

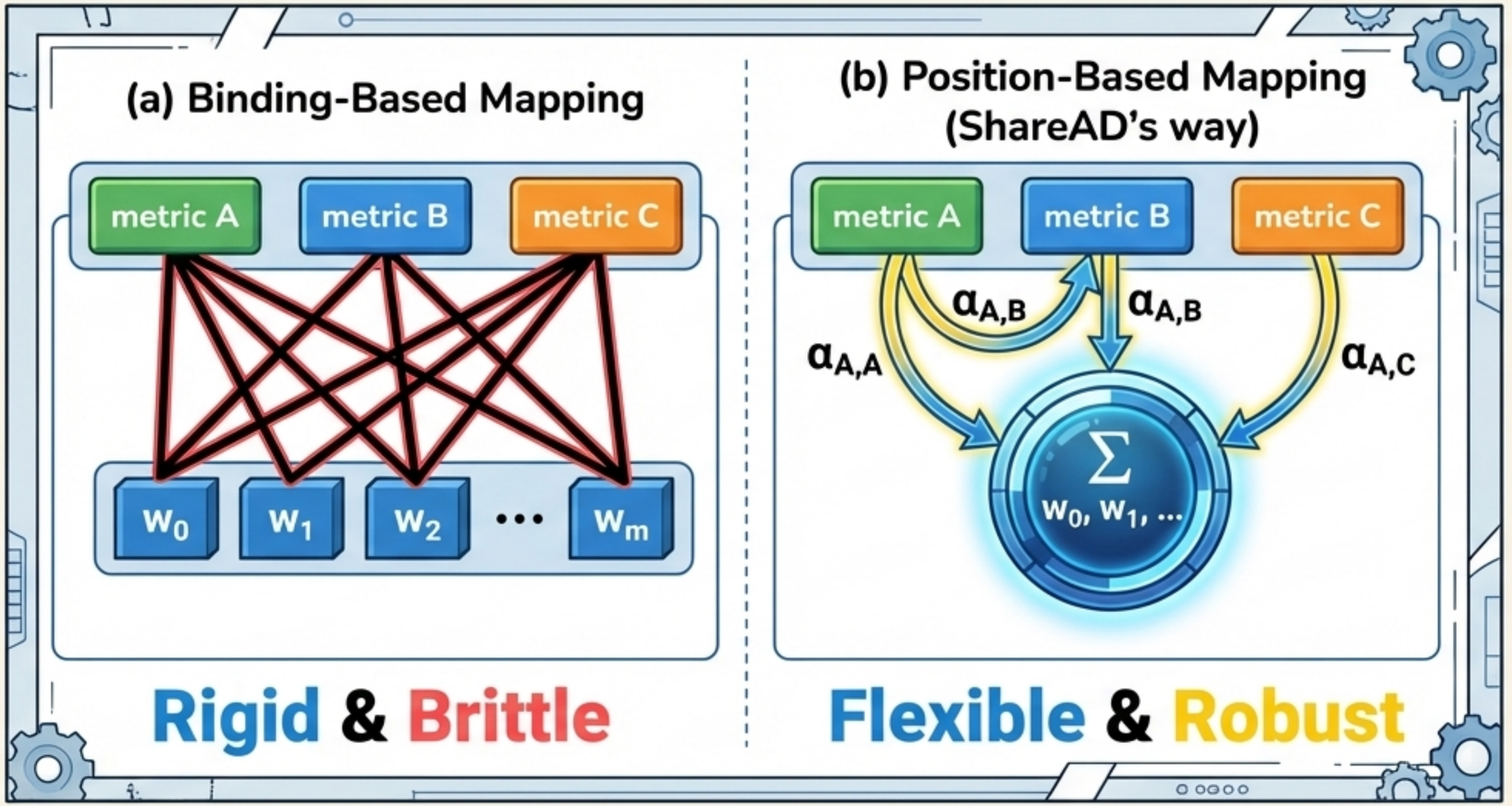
A Novel Architecture for Effective Model Sharing



This dual architecture allows a single model to learn diverse patterns from hundreds of different entities simultaneously, which is why model sharing works.

The Secret: Position-Based Mapping for Ultimate Flexibility

Traditional models bind each input metric to a specific weight.
This is rigid and breaks when metrics are missing or new.

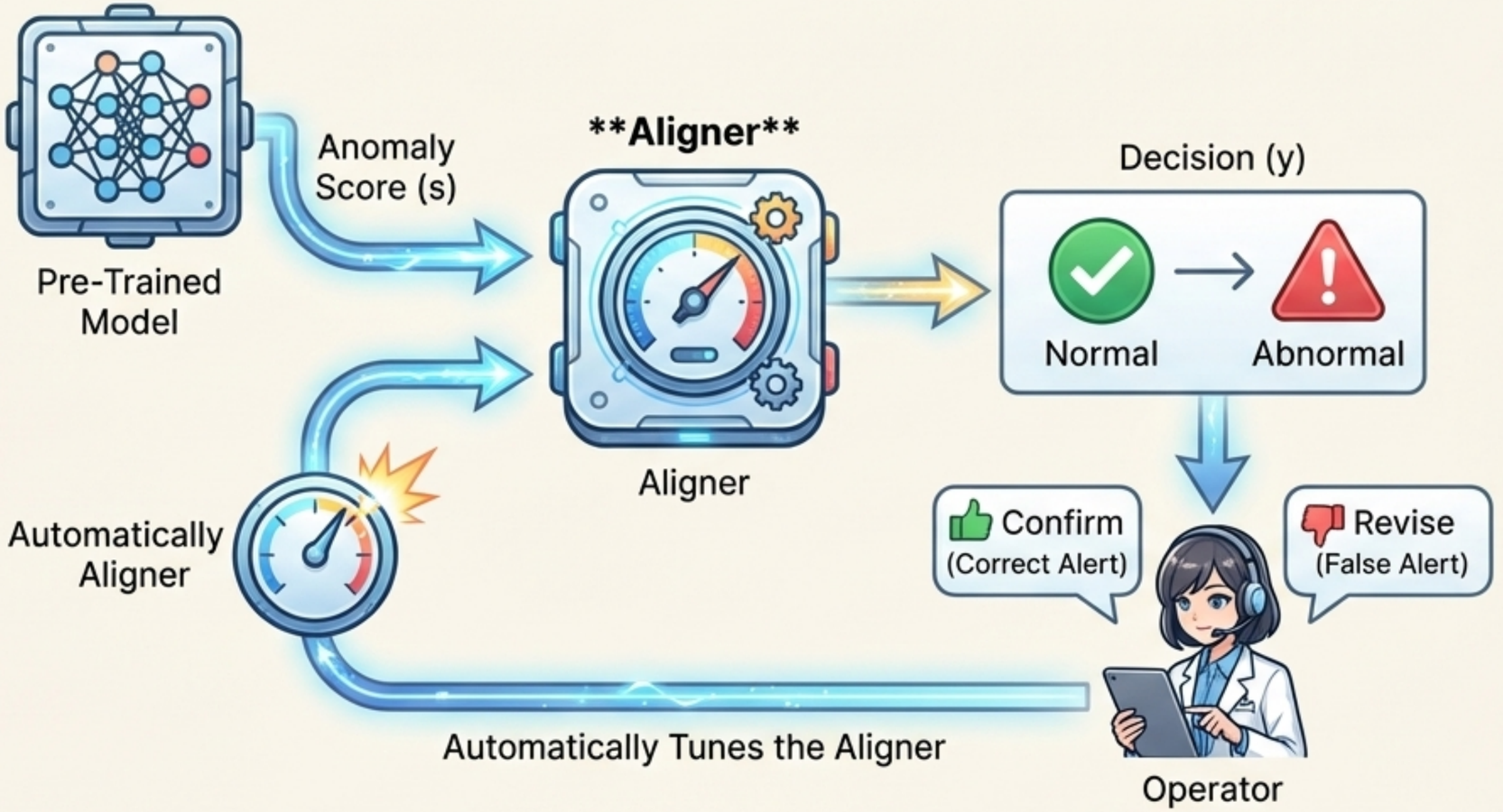


Key Benefits:

1. **Handles Diverse Patterns:** The model learns relationships, not fixed patterns for specific metrics.
2. **Handles Inputs with Variant Cardinalities:** It works even if entities have different sets of metrics, or if some metrics are missing. This is a critical feature for real-world practicality.

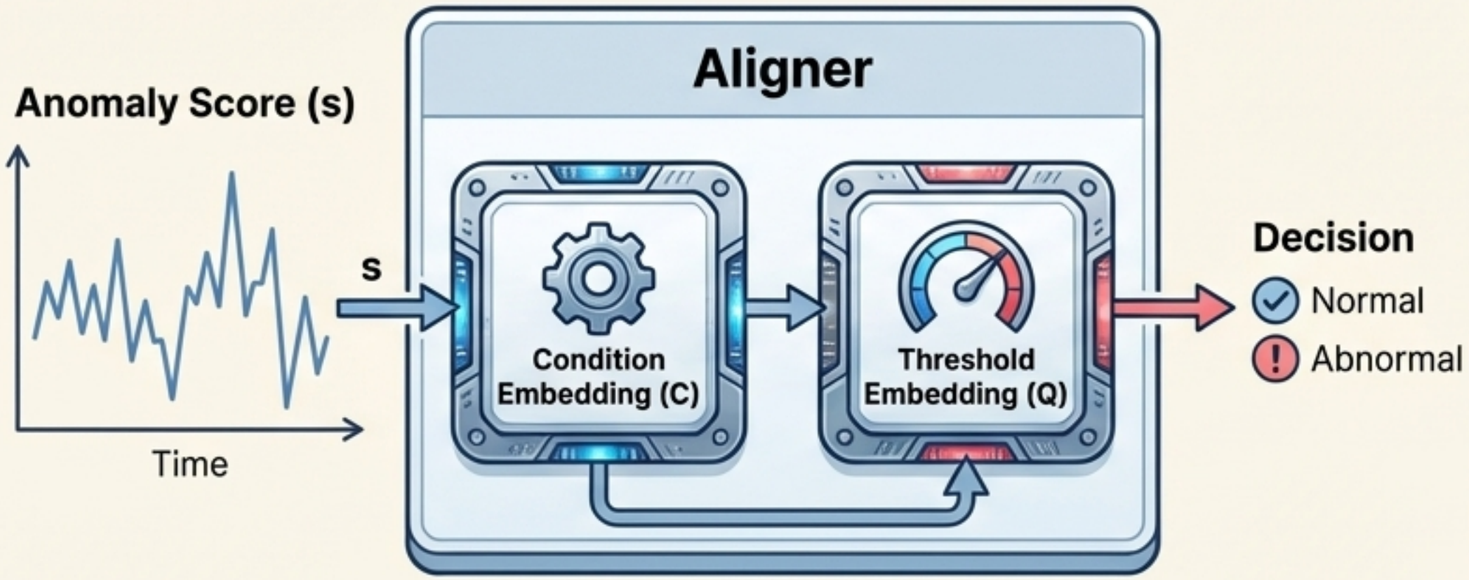
Gadget Blueprint Part 2: The Alignment Dial

Turning Ambiguous Scores into Operator-Friendly Decisions



The Aligner acts as a lightweight, personalized decision-maker for each entity. It learns operator preferences from their daily actions, eliminating the need for manual threshold tuning.

How the Aligner Makes Smart Decisions



1. Condition Embedding (C)

A learned **window size** for each entity.

It considers the **local context** of an observation. Is the current score high *relative to the recent past*? This makes it adaptive.

$$s_t - \max(s_{t-C_e} \dots s_{t-1})$$

2. Threshold Embedding (Q)

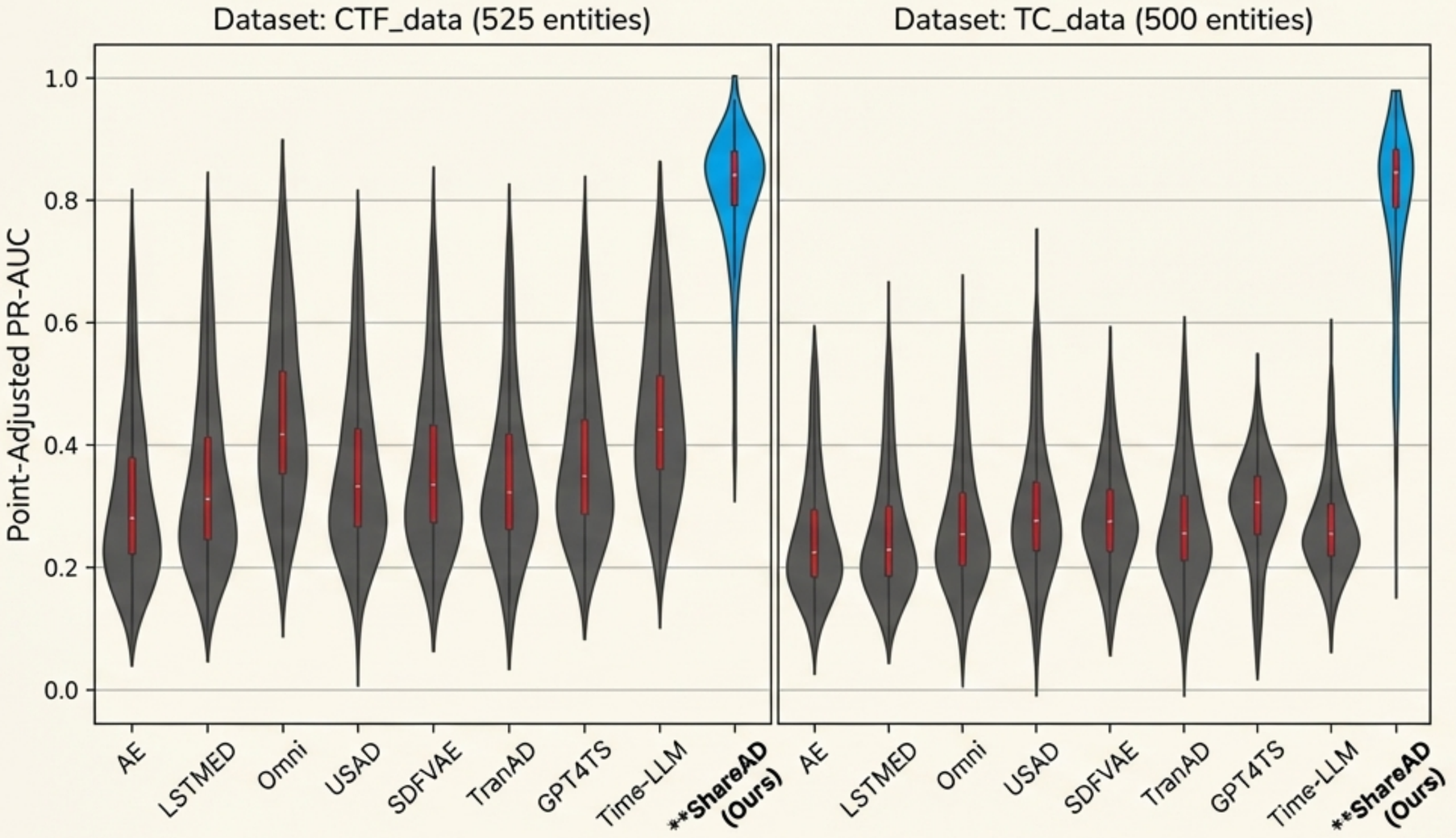
A learned **sensitivity level** for each entity.

It captures the **unique anomaly sensitivity** of each component. A critical database might have a lower threshold than a non-essential cache service.

$$\dots - Q_e > 0\,?$$

Learning Mechanism: These embeddings (C and Q) are optimized for each entity using Bayesian Optimization (TPE). This allows us to directly optimize for **non-differentiable**, real-world metrics like **Point-Adjusted F1-Score**, which is what operators truly care about.

The Proof, Part 1: ShareAD's Scoring Outperforms State-of-the-Art



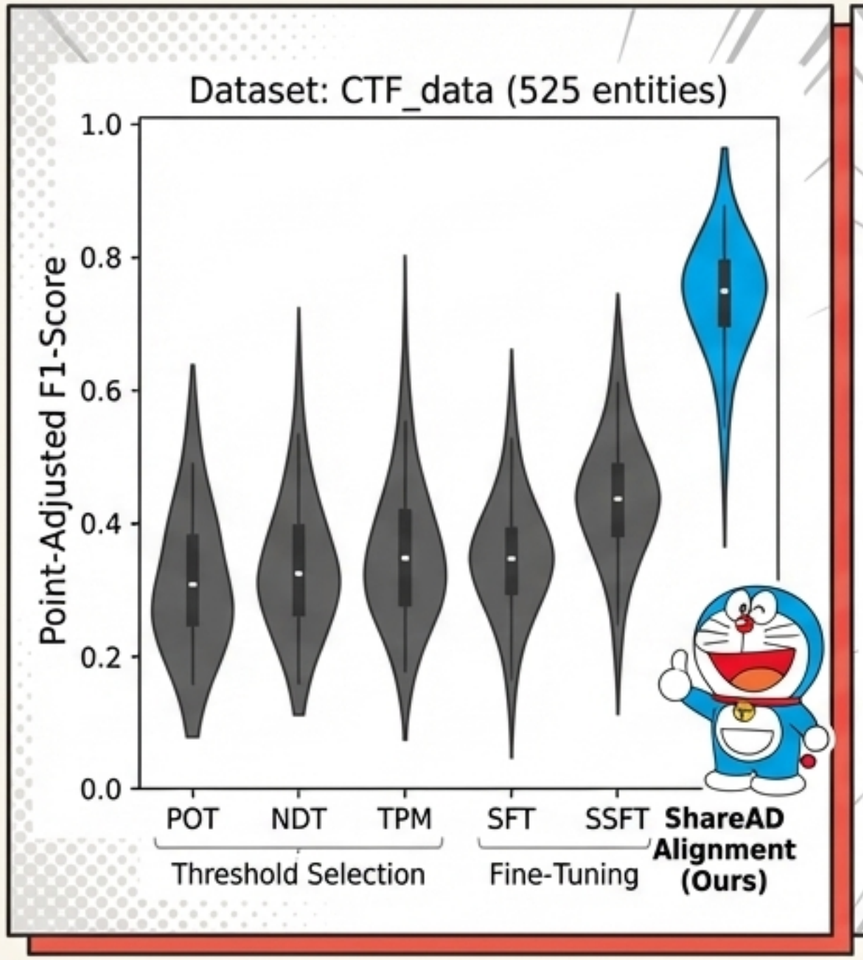
> On CTF_data, ShareAD achieves a relative improvement of **8.80%–93.06%** in Point-Adjusted PR-AUC compared to baselines.

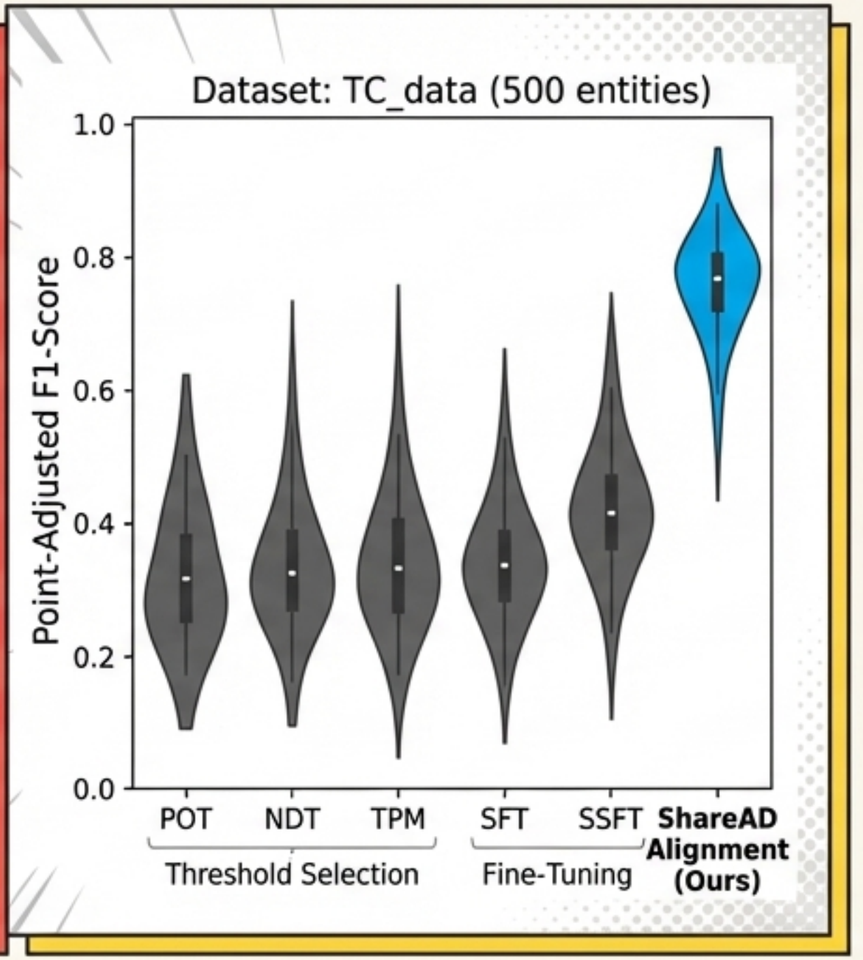
> On TC_data, the improvement is **6.66%–201.19%**.

> Performance is also more stable, with less variance across different hyper-parameter settings.

ShareAD's pre-trained model provides a more effective and reliable foundation for anomaly detection.

The Proof, Part 2: Aligned Decisions Beat Traditional Thresholding





ShareAD Alignment achieves a relative F1-score improvement of **20.99%–135.29%** on CTF_data.

On TC_data, the improvement is **15.73%–175.04%**.

Unlike baseline methods which are often unstable, ShareAD provides consistently high performance by directly learning operator preferences.

The model alignment technique effectively bridges the gap from scores to accurate, operator-friendly decisions.

The Proof, Part 3: Designed for Real-World Speed and Scale

Fast Detection

< 0.1 ms
per detection on GPU

< 1 ms
per detection on CPU

Rapid Alignment

~1 second
to align per entity on GPU

~4 seconds
to align per entity on CPU

Tiny Footprint

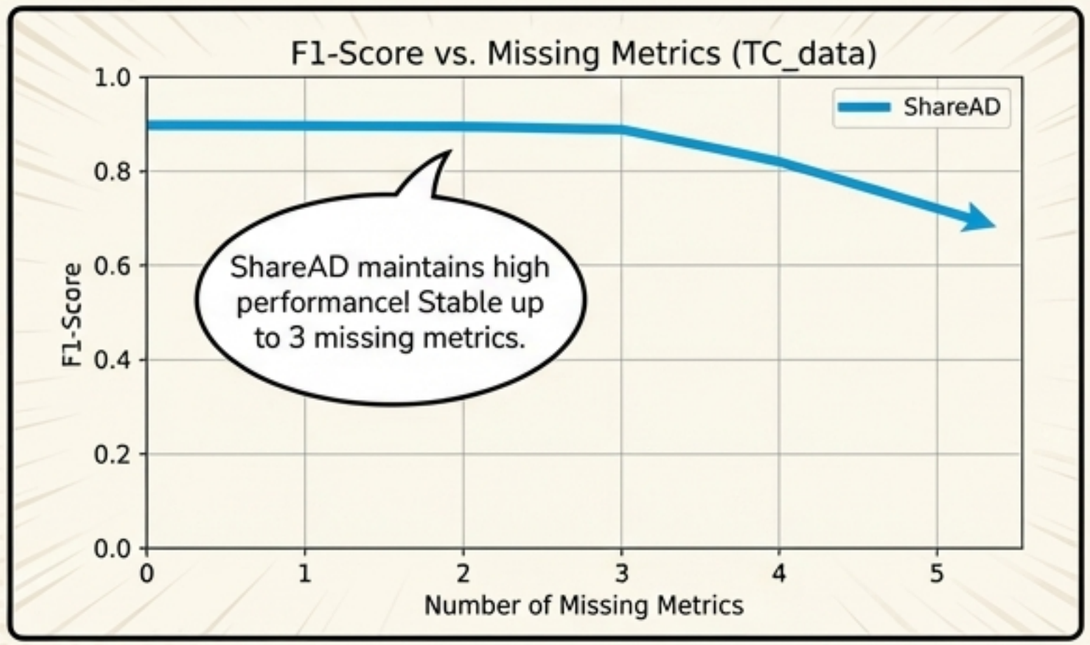
Pre-trained model is only
~400 KB

Aligner adds just
60 bytes
per entity

Bottom Line: ShareAD is lightweight and lightning-fast, making it practical to deploy, run, and continuously update across thousands of entities without requiring massive computational resources.

The Proof, Part 4: Excelling in Messy, Real-World Conditions

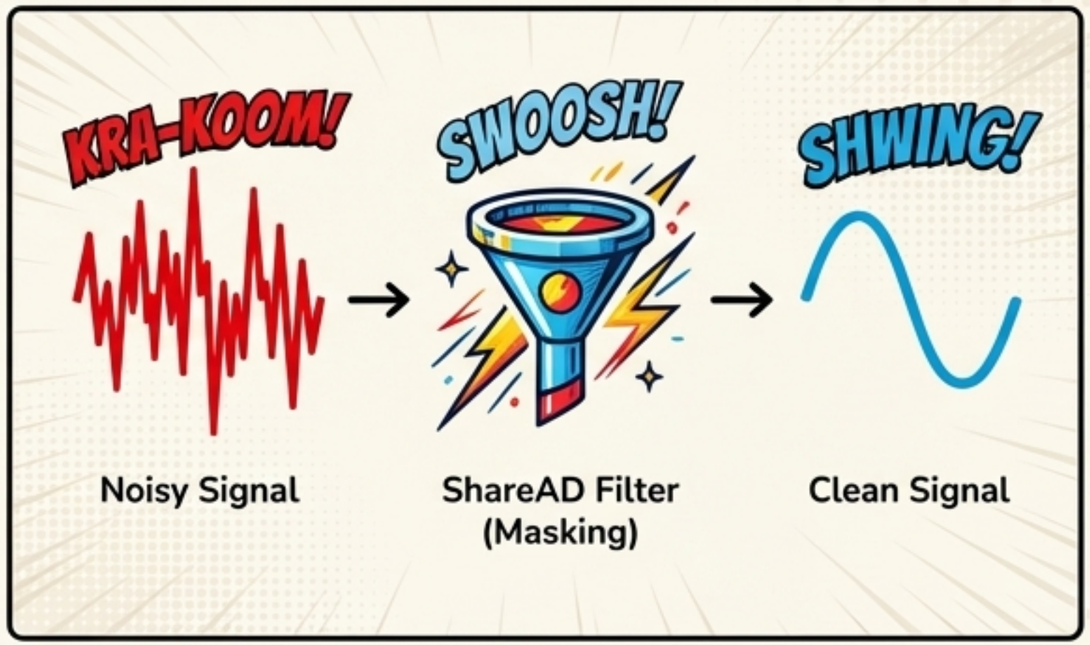
Handling Missing Metrics



ShareAD's position-based mapping allows it to maintain high performance even when metrics are permanently missing—a common issue that breaks traditional models.

- **Performance remains stable with up to 3 (out of 11) metrics missing on TC_data.**

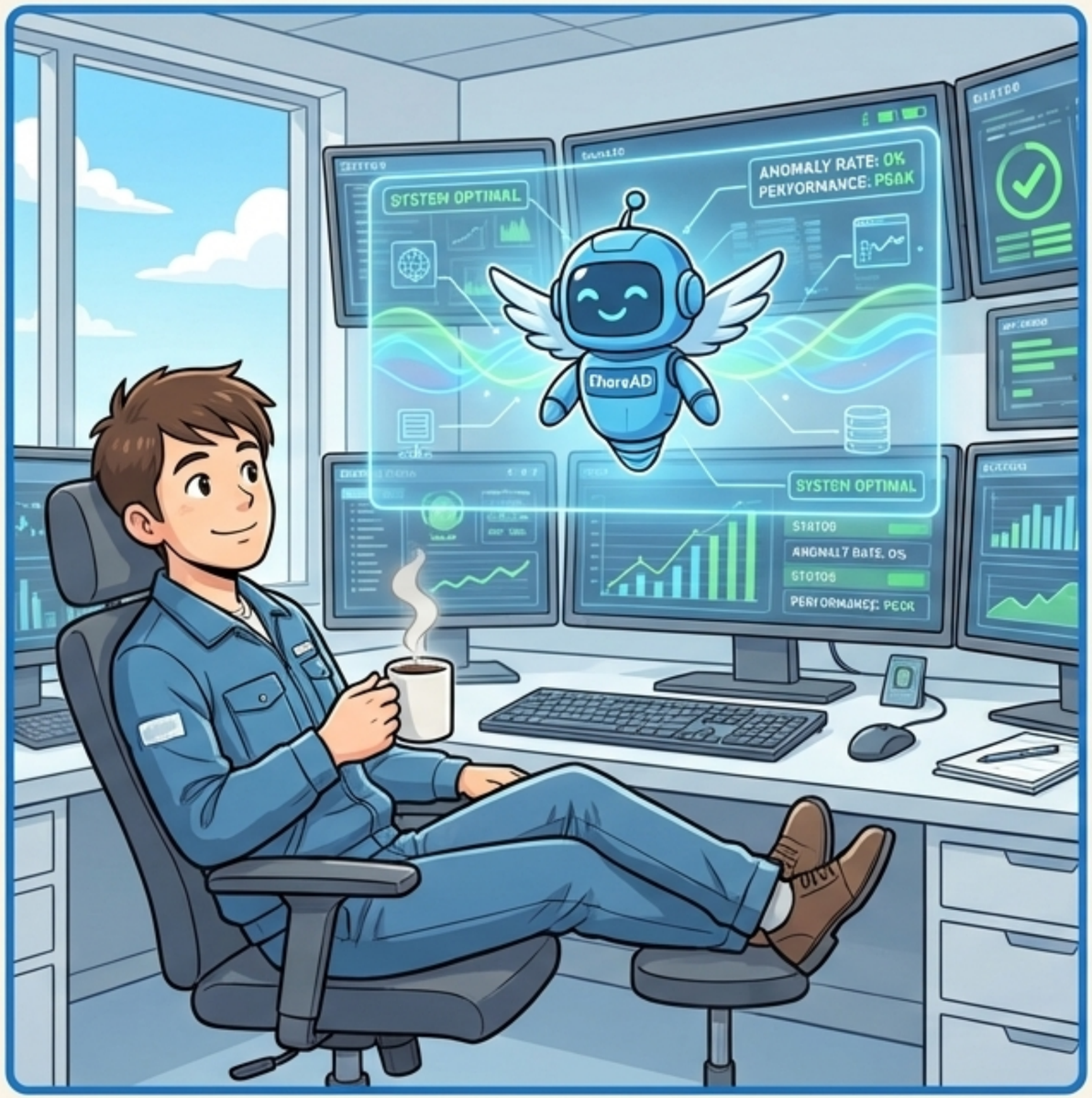
Masking Noisy Metrics



Operators can manually mask noisy metrics for specific entities. ShareAD adapts on-the-fly, improving accuracy.

- **Masking noisy metrics on select entities improved Point-Adjusted PR-AUC by an average of 0.32.**

From Overwhelmed Operator to Empowered Engineer

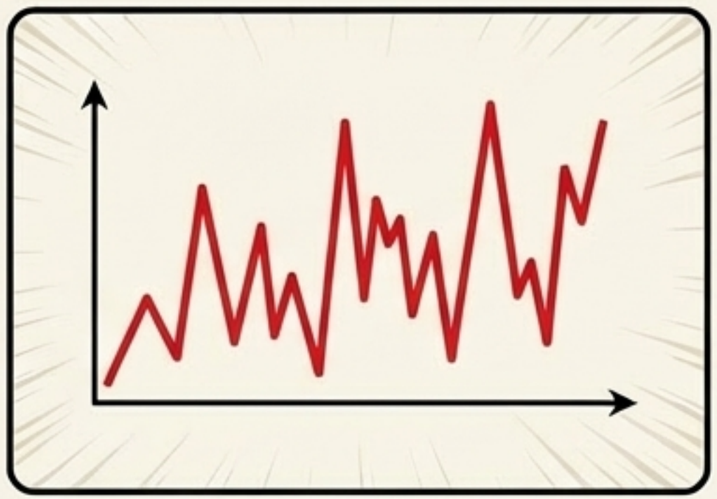


Summary of Key Innovations:

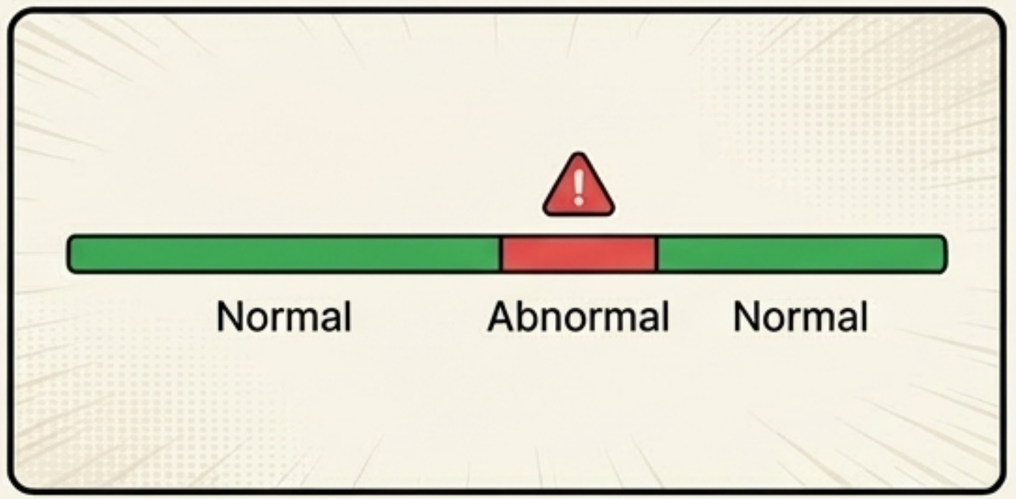
1. **A Shared Pre-Trained Model:** Bridges the **Scalability** and **Availability** gaps by replacing thousands of models with one.
2. **Novel Transformer + Base Layer Architecture:** Effectively learns diverse patterns from many entities.
3. **Lightweight, Feedback-Driven Aligner:** Bridges the **Alignment** gap by translating scores into decisions that learn from operator expertise.

ShareAD delivers a **practical, powerful**, and operator-centric solution for anomaly detection in modern, large-scale systems.

The Real Shift: From Ambiguous Scores to Aligned Decisions



The Old Way:
Model-Centric Scores



The ShareAD Way:
Human-Aligned Decisions

Effective anomaly detection is more than just a better model; it's a better framework. By shifting the focus from simply generating scores to learning **operator preferences** and making **aligned decisions**, ShareAD makes deep learning **practical and trustworthy** for real-world system maintenance.