ICML
2026
POSTER

A Risk Decomposition Framework for Pre-Hoc Fine-Tuning Prediction

Predict final fine-tuning performance by separating irreducible ambiguity from reducible optimization variance.

Yuxiang Luo · Chen Wang · Nan Tang
The Hong Kong University of Science and Technology (Guangzhou)

arXiv 2606.17649

CORE IDEA **Pre-hoc prediction is an uncertainty-resolution problem:** probing only helps when it reveals optimization information not already encoded in static data-model signals.

2 risk components	**$Kc^{-\alpha}$** uncertainty envelope	**3** predictability regimes	**1500** runs per probe depth

1 Why Pre-Hoc Prediction Is Hard

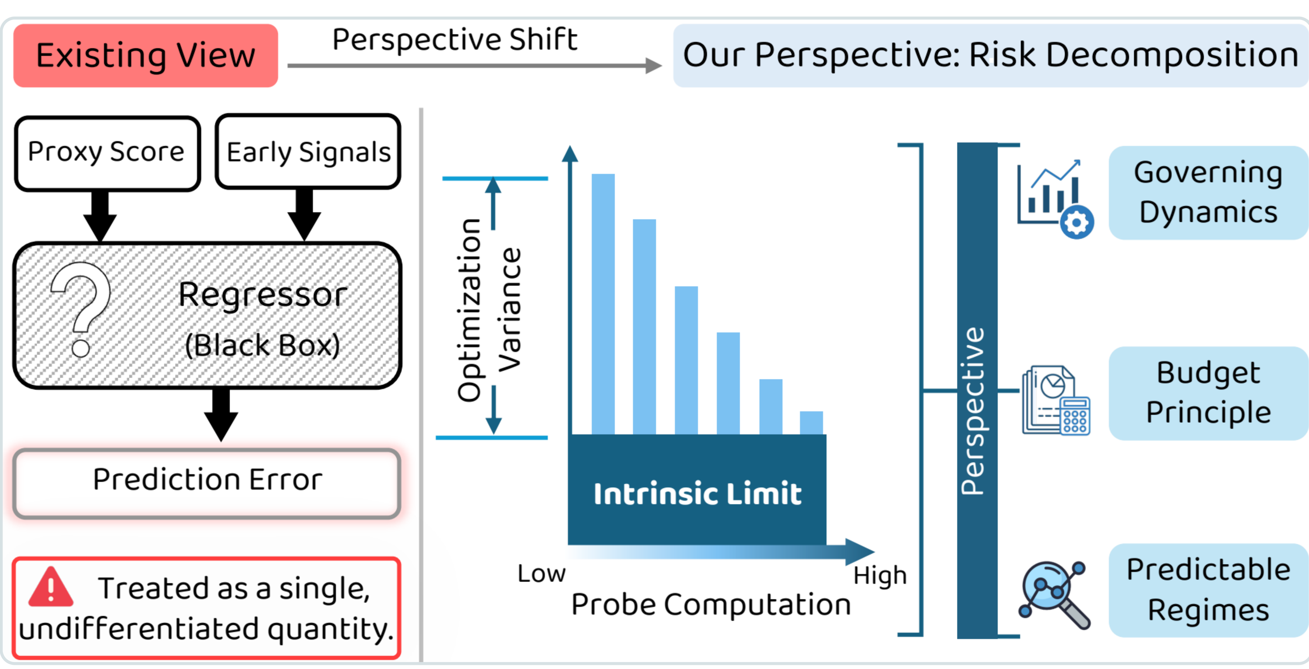
Fine-tuning LLMs is costly and uncertain: identical configurations can end in different outcomes because pre-training priors, dataset properties, and stochastic optimization interact.

- **Goal:** estimate the final performance before committing to full training.
- **Problem:** existing predictors treat error as one black-box quantity and tune probing depth heuristically.

Question: Which part of prediction error is fundamentally irreducible, and which part can computation actually resolve?

2 Perspective Shift

The paper moves from black-box regression to a structural view: prediction risk is governed by static data-model compatibility and dynamic information revealed by the optimization trajectory.



Paper Fig. 1: pre-hoc prediction decomposes a single error signal into an intrinsic limit plus optimization variance.

2b Task Setup

A fine-tuning task is T=(M,D,A): pretrained model, downstream dataset, and stochastic optimizer. Running T yields a scalar random outcome R.

RISK AT BUDGET C

$$L_T(c) = E[(f(I_c) - R)^2]$$

2c Main Contributions

- Formal decomposition of pre-hoc prediction risk.
- Asymptotic lower-bound envelope for uncertainty decay.
- Risk-cost equilibrium for budget-aware probing.
- Predictability phase diagram linking theory to empirical failures.

Practical use: continue, stop early, or prioritize another configuration.

3 Risk Decomposition

For the Bayes-optimal predictor with information set I_c, the prediction risk separates into a computation-independent floor and a finite-trajectory uncertainty gap.

PROPOSITION

$$L(c) = E[Var(R \mid I_\infty)] + E[Var(R \mid I_c) - Var(R \mid I_\infty)]$$

L_{int} irreducible intrinsic limit	**V_{opt}** reducible optimization variance	**I_c** information after budget c

4 Information Revelation

Dynamic probing can reduce $V_{opt}(c)$ only when the partial trajectory contains information about R beyond static priors X_s.

REQUIREMENT

$$I(R \,;\, X_d^{(c)} \mid X_s) > 0$$

If the trajectory is determined by static information, probing reveals no new signal and cannot improve prediction accuracy.

STATIC X_s: data-model descriptors **+** **DYNAMIC** $X_d^{(c)}$: early trajectory **=** **DECISION** continue, stop, or reprioritize

5 Uncertainty Cannot Vanish Arbitrarily Fast

ASYMPTOTIC ENVELOPE

$$V_{opt}(c) \geq K\, c^{-\alpha}$$

Under locally regular stochastic optimization, the effective decay rate alpha captures how quickly regime-relevant uncertainty is revealed.

5b Offline Calibration

For analysis and benchmark design, repeated lightweight probes estimate a surrogate uncertainty curve, then jointly fit alpha, K, and the empirical intrinsic floor.

PROBE depths c_i **>** **FIT** $L_{int} + Kc^{-\alpha}$ **>** **CHOOSE** c^*

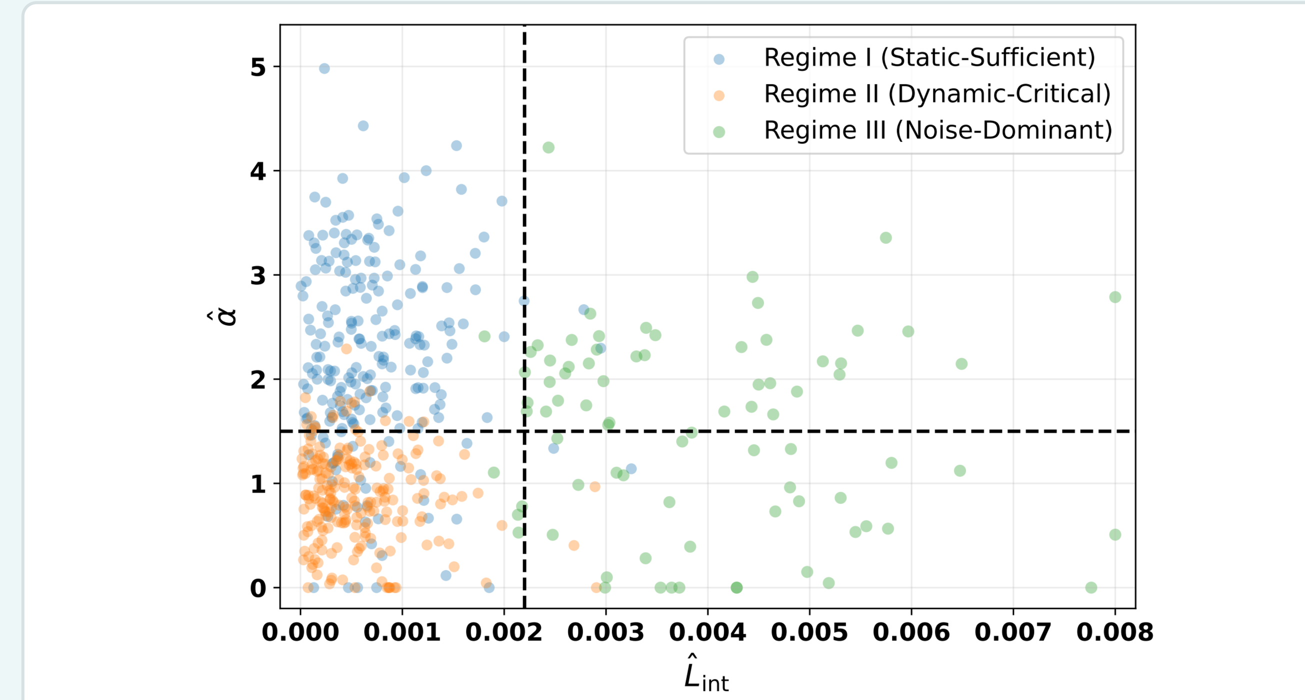
5c Which Signals Matter?

The framework predicts that static and dynamic signals are complementary only when dynamic information carries conditional signal beyond static priors.

Appendix evidence: dynamic signals help primarily in Dynamic-Critical regimes; static/noisy regimes show limited gains.

6 Predictability Phase Diagram

Tasks live in a two-dimensional phase space: intrinsic ambiguity L_{int} and effective decay rate α. A single scalar difficulty score misses the mechanism.



Paper Fig. 3: tasks separate by intrinsic floor and decay exponent into three regime-level structures.

7 Three Regimes

Static-Sufficient: rapid uncertainty collapse; static descriptors nearly saturate predictability.

Dynamic-Critical: low intrinsic ambiguity but slow information revelation; extended probing is valuable.

Noise-Dominant: high intrinsic floor; more probing has limited practical effect.

Failures of pre-hoc predictors can therefore be read as **regime mismatch** between estimator design, allocated probing budget, and task dynamics.

7b How to Read Failures

- **SST-2-like tasks:** shallow probing may be enough because information is already static.
- **GSM8K-like tasks:** shallow probing can be unstable because useful dynamics arrive late.
- **Noisy tasks:** deeper probing mostly refines noise around a large intrinsic floor.

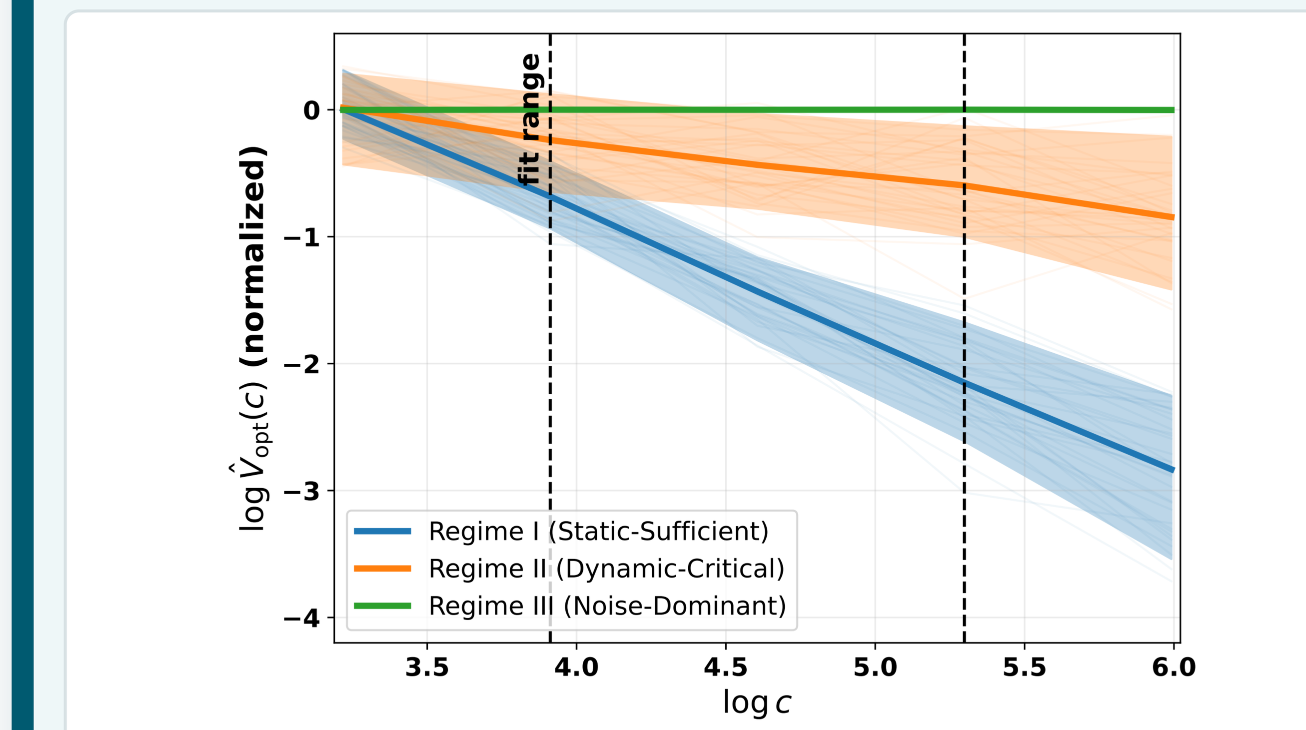
7c Hard Cases Localize

Appendix evidence: hard cases concentrate in Noise-Dominant regions or near phase boundaries.

8 Empirical Decay Across Regimes

The experiments estimate observable uncertainty proxies across repeated fine-tuning runs and probe depths.

- **Blue:** Static-Sufficient tasks shed uncertainty quickly.
- **Orange:** Dynamic-Critical tasks keep useful uncertainty over longer probe budgets.
- **Green:** Noise-Dominant tasks remain near an intrinsic floor.



Paper Fig. 2: normalized uncertainty follows an approximately linear log-log trend inside the fitted range, with regime-specific decay rates.

9 Budget-Optimal Probing

Because $V_{opt}(c)$ decays with diminishing returns, probing becomes an optimal stopping problem under computational cost.

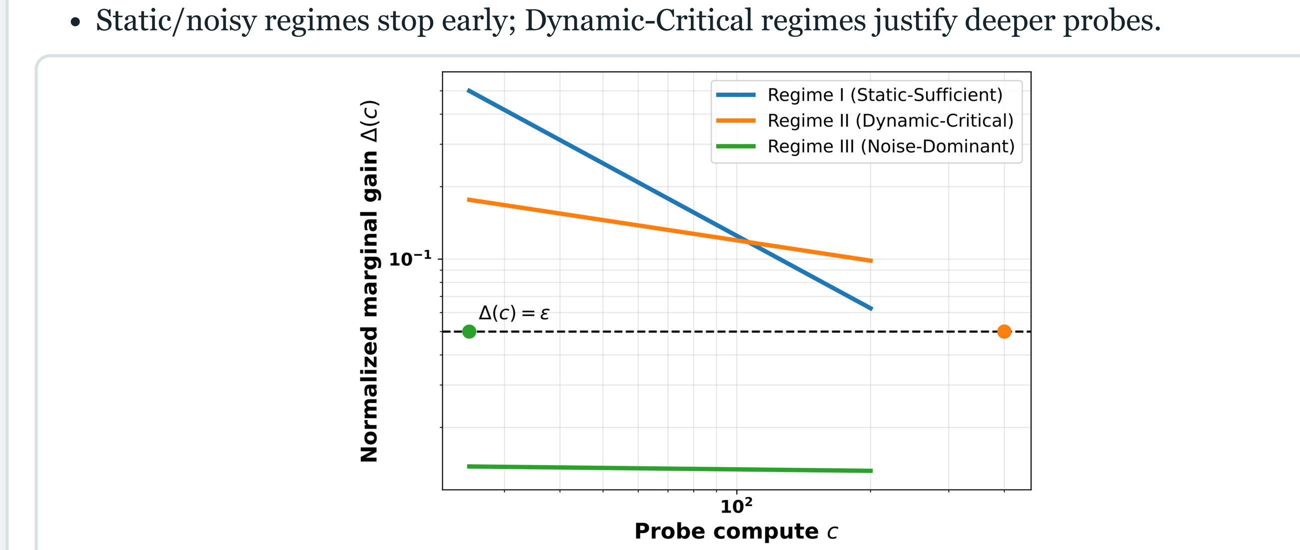
EQUILIBRIUM

$$|dV_{opt}(c)/dc| = \gamma\, C'(c)$$

LINEAR COST

$$c^* = (\alpha K / \gamma\lambda)^{1/(\alpha+1)}$$

- Probe while the marginal uncertainty drop is worth the extra compute.
- Static/noisy regimes stop early; Dynamic-Critical regimes justify deeper probes.



Paper Fig. 4: marginal gains decay at different rates, supporting regime-aware stopping rather than fixed-step heuristics.

Takeaways

Decompose. Prediction risk has an intrinsic floor plus a reducible optimization gap.

Probe selectively. Trajectory signals help only when they reveal information beyond static priors.

Diagnose regimes. L_int and alpha explain early success, delayed success, and intrinsic failures.

Allocate budget. Stop where marginal uncertainty reduction equals marginal compute cost.

Paper: arxiv.org/abs/2606.17649 · Contact: nantang@hkust-gz.edu.cn