PORTFOLIO · MACHINE LEARNING RESEARCH

VOL. 01

Youngjae *Cho.*

2026

ML Research Scientist · Trustworthy & Robust Learning

Preference Alignment / Vision-Language Models / Active Learning

leon5760@gmail.com

KAIST · Industrial & Systems Engineering

Profile.

I build learning systems that don't fall apart when supervision gets messy.

My work spans LLM preference alignment, vision-language prompt learning, active learning, and industrial anomaly detection — published at AAAI and ICML, with current work submitted to NeurIPS. I currently serve as an ML Research Scientist at Pyler under the Korean alternative military service program.

06	04	02	06+
Peer-reviewed & preprint papers	First-author works	Top-tier venues AAAI · ICML	Years in ML research

Education & Experience.

2025 – *Present*

ML Research Scientist

PYLER

Alternative military service · Continuing alignment & robustness research in industry.

2024 – *2025*

ML Research Scientist

AIV CO.

Alternative military service · Led applied research on industrial anomaly detection, with deployment via ONNX / TensorRT / Triton.

2022 – *2024*

M.S. Industrial & Systems Engineering

KAIST

Published at AAAI 2024 and ICML 2023 on prompt learning and active learning.

2017 – *2022*

B.S. Industrial & Systems Engineering

KAIST

Foundations in optimization, statistics, and machine learning.

Research Focus.

I work on making learning systems trustworthy: align them with human intent, choose what to learn from, and stay robust under noisy or shifted supervision.

PREFERENCE ALIGNMENT

01

GAPO · Geometric Anchoring

Geometry-aware DPO/SimPO objectives. Build a batch-conditioned pessimistic anchor and reweight by local stability.

NeurIPS 2026 · in submission

VISION-LANGUAGE

02

Adaptive Bayesian Prompts

Each input gets its own prior over prompts, learned variationally. Better generalization to novel classes and shifted domains.

AAAI 2024 · first author

ACTIVE LEARNING

03

Sharpness-Aware Acquisition

Score candidate samples by the loss geometry they induce. Bridges Sharpness-Aware Minimization and active learning.

ICML 2023 · co-first author

ANOMALY DETECTION

04

Background-Aware Defects

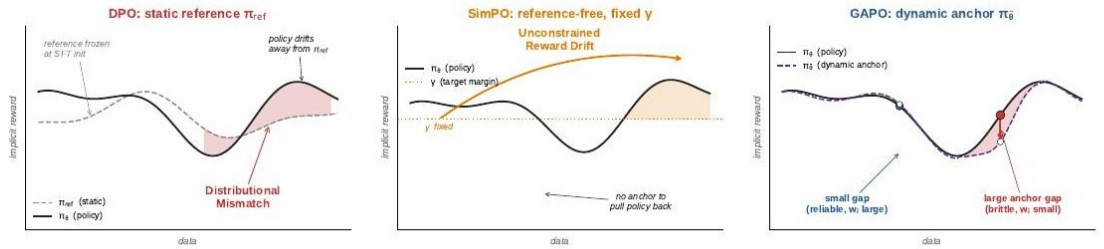
Generate realistic defects that respect background structure — for robust manufacturing inspection at scale.

preprint · first author

Learning Where It Matters.

NeurIPS 2026 · submitted

Geometric Anchoring for Robust Preference Alignment



(a) DPO: distributional mismatch with a static reference.

(b) SimPO: unconstrained reward drift with no anchor.

(c) GAPO: dynamic anchor with adaptive Anchor Gap.

FIGURE 1 · DPO's frozen reference creates distributional mismatch. SimPO has no anchor (unconstrained drift). GAPO's dynamic anchor tracks the policy and reweights brittle pairs.

THE PROBLEM

Margin-based losses like DPO and SimPO can't distinguish a genuinely hard pair from a brittle, noisy one — both produce large updates. Noisy labels and length bias quietly destabilize alignment.

THE IDEA · GAPO

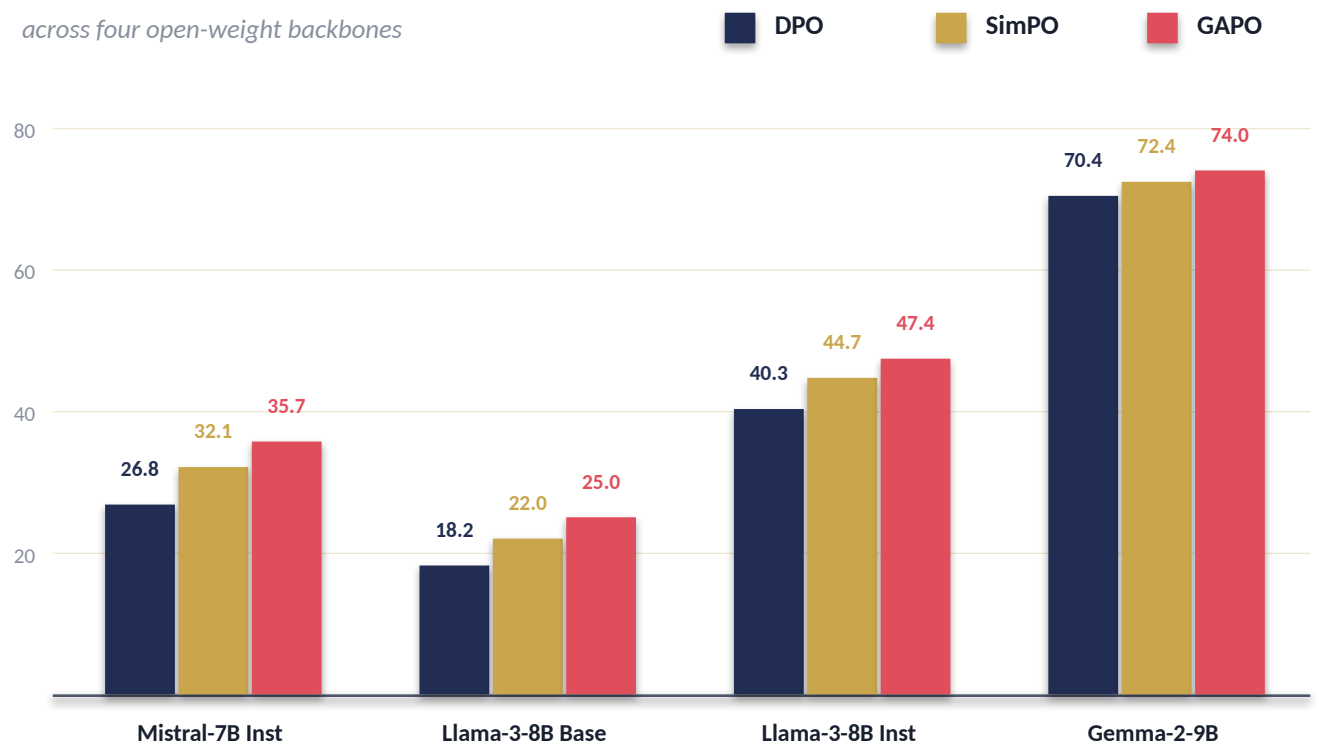
Build a batch-conditioned pessimistic anchor by perturbing the policy in the direction that maximally decreases the batch-average margin. Reweight each pair by its Anchor Gap — local brittleness — preserving the preference-gradient direction.

GAPO — Alignment Results.

Stronger alignment, preserved reasoning, and lower training cost than longer SimPO runs.

ALPACAEVAL 2.0 · LENGTH-CONTROLLED WIN RATE (%)

across four open-weight backbones

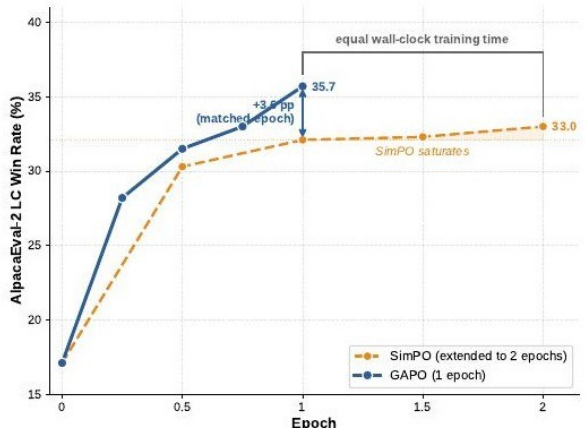


ALPACAEVAL 2 LC · vs SimPO

+3.6 pp

on Mistral-7B; gains hold across all four model families.

FIGURE · WALL-CLOCK MATCHED

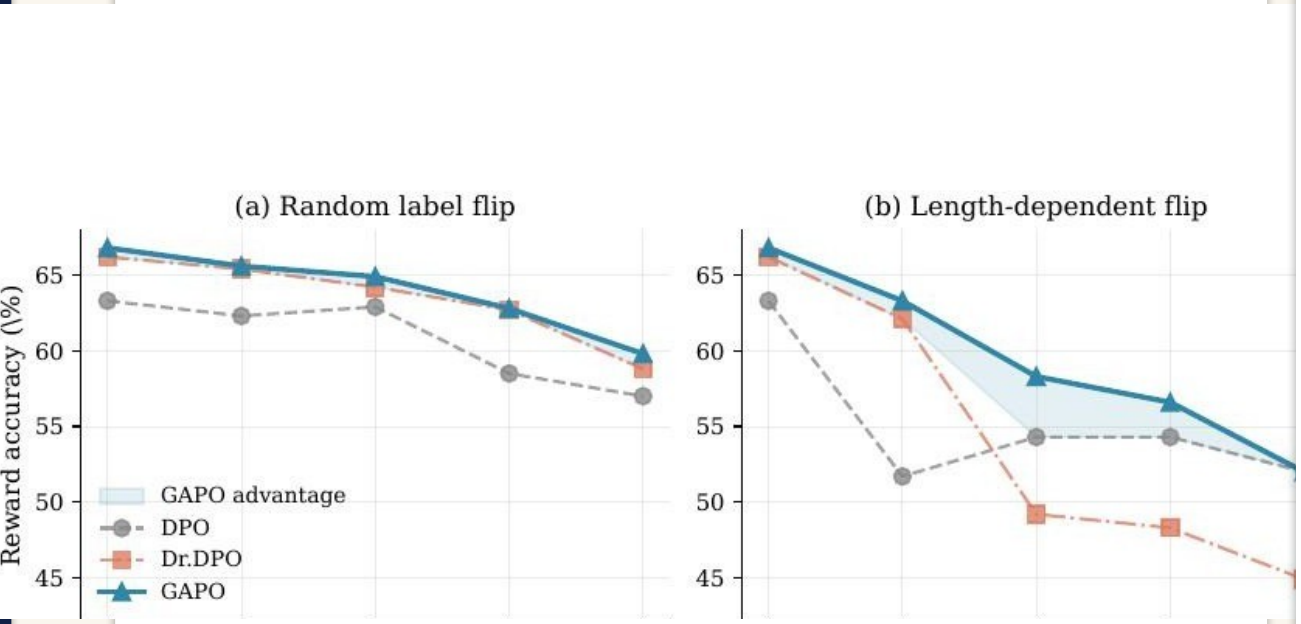


1 epoch of GAPO beats 2 epochs of SimPO at the same wall-clock budget.

GAPO — Robustness & Diagnostics.

Why does GAPO help? Brittle pairs are statistically the noisy ones — and the algorithm down-weights them.

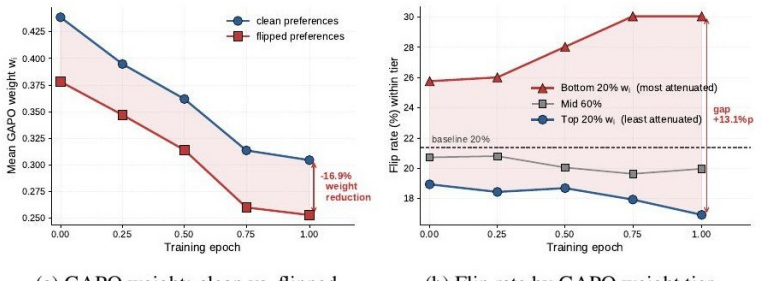
FIGURE 2 · ROBUSTNESS AGAINST LABEL NOISE



Flip rate (%) → 0 · 10 · 20 · 30 · 40

GAPO degrades smoothly under both random and length-dependent label flips — Dr.DPO collapses below DPO at 20%+ flip rate.

FIGURE 3 · ANCHOR-GAP REWEIGHTING SIGNAL



−16.9%

Mean weight reduction on flipped pairs by end of first epoch.

+13.1pp

Flip-rate gap between the bottom-20% and top-20% GAPO-weight tiers.

Make Prompts Adaptable.

AAAI 2024 · first author

Bayesian Modeling for Vision-Language Prompt Learning with Data-Dependent Prior

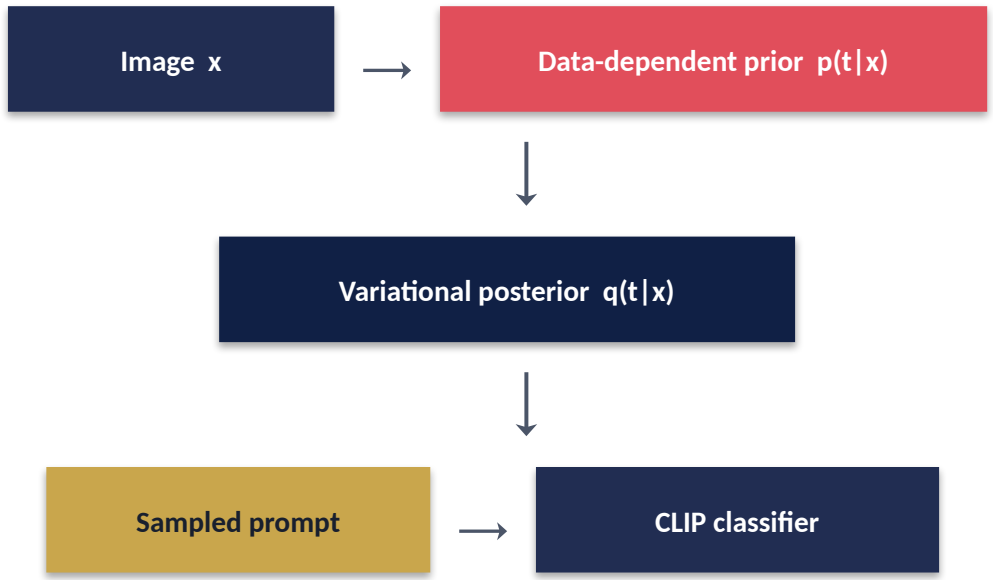
MOTIVATION

CLIP-style prompt learners (CoOp, CoCoOp) collapse the prompt to a single deterministic vector that is shared across all inputs. The prompt becomes fragile to distribution shift and limits generalization to unseen classes.

CONTRIBUTION

A Bayesian prompt-learning framework with a data-dependent prior: each input induces its own prior over prompts, and the posterior is trained variationally. The result is a prompt that adapts to the input image while remaining regularized — improving generalization to novel classes and shifted domains.

CONCEPT · ADAPTIVE PROMPT



Each image gets its own prompt distribution — not a shared point estimate.

SAAL — Sharpness-Aware Active Learning.

ICML 2023 · co-first

Selecting samples whose loss is robust to small perturbations.

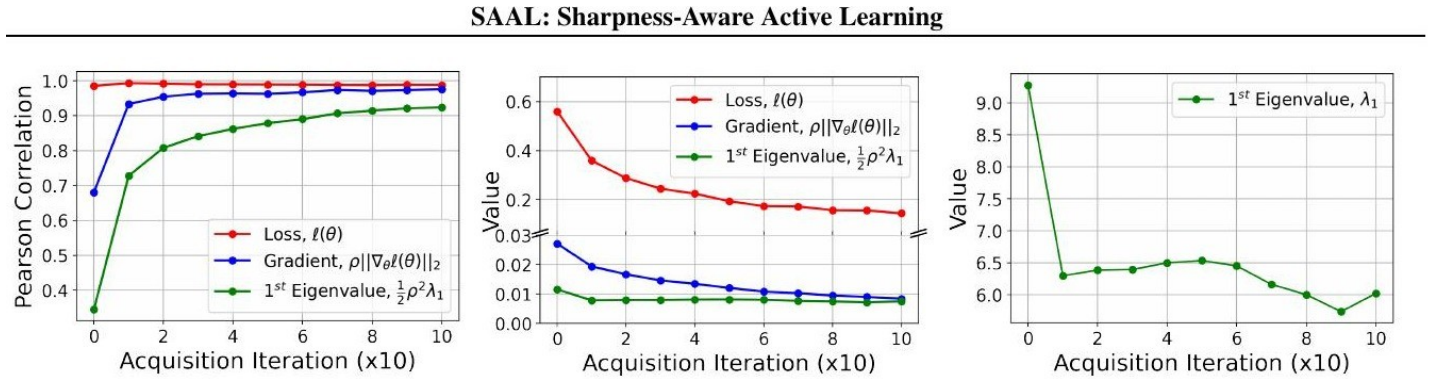


FIGURE · Correlation and magnitude of the proposed acquisition score across loss, gradient, and 1st Hessian eigenvalue.

PROBLEM

01

Classical acquisition functions (entropy, BALD, coreset) score samples by uncertainty alone, ignoring loss geometry.

IDEA

02

Score candidates by sharpness — the worst-case loss in a small parameter neighborhood — and prefer flatter samples.

IMPACT

03

Consistent gains across image and tabular benchmarks in low-budget regimes. Bridges SAM and active learning.

Selected Papers.

2025 PREPRINT

Learning Where It Matters: Geometric Anchoring for Robust Preference Alignment

Youngjae Cho, Jongsuk Kim, Ji-Hoon Kim

Under review at NeurIPS 2026 · first author

2025 PREPRINT

Background-Aware Defect Generation for Robust Industrial Anomaly Detection

Youngjae Cho, Gwangyeol Kim, Sirojbek Safarov, Seongdeok Bang, Jaewoo Park

first author

2024 AAAI

Make Prompts Adaptable: Bayesian Modeling for Vision-Language Prompt Learning

Youngjae Cho, HeeSun Bae, Seungjae Shin, YeoDong Youn, Weonyoung Joo, Il-Chul Moon

AAAI Conference on Artificial Intelligence · first author

2023 ICML

SAAL: Sharpness-Aware Active Learning

Yoon-Yeong Kim*, Youngjae Cho*, JoonHo Jang, Byeonghu Na, Yeongmin Kim, Kyungwoo Song, Wanmo Kang, Il-Chul Moon

International Conference on Machine Learning · co-first author

---

2022 ICML WS

Improving Group-based Robustness and Calibration via Ordered Risk and Confidence Regularization

S. Shin, B. Na, H. Bae, J. Jang, H. Kim, K. Song, Youngjae Cho, Il-Chul Moon

Workshop on Spurious Correlations, Invariance and Stability

2021 IEEE SMC

Predict Sequential Credit Card Delinquency with VaDE-Seq2Seq

Yeongmin Kim, Youngjae Cho, Hanbit Lee, Il-Chul Moon

IEEE Int. Conf. on Systems, Man, and Cybernetics

Technical Skills.

RESEARCH & MODELING

Frameworks I use to design, train, analyze.

PyTorch | TensorFlow | JAX

Strong fluency in PyTorch (primary). Comfortable porting research code between frameworks.

DEPLOYMENT & SERVING

Out of the notebook and into production.

ONNX | TensorRT | Triton

Experience deploying trained models with optimized runtimes for low-latency serving.

METHODS & TOPICS

Where my work has gone deep.

- **LLM Alignment** — DPO · SimPO · KTO · GAPO
- **Vision-Language** — CLIP · CoOp / CoCoOp · prompt learning
- **Active Learning** — Entropy · BALD · coreset · SAAL
- **Robustness** — Noisy labels · spurious correlations · calibration
- **Generative Models** — VAE · Diffusion · defect synthesis
- **Bayesian DL** — Variational inference · data-dependent priors

LET'S BUILD SOMETHING.

Thank you.

I'd love to talk about alignment, robustness, or any of the work here.

EMAIL

leon5760@gmail.com

CURRENT

ML Research Scientist · Pyler

EDUCATION

M.S. & B.S. — KAIST ISysE

— Youngjae Cho