

Learning from One Another: Toward Complementary Knowledge Transfer via Localized Preference Deltas

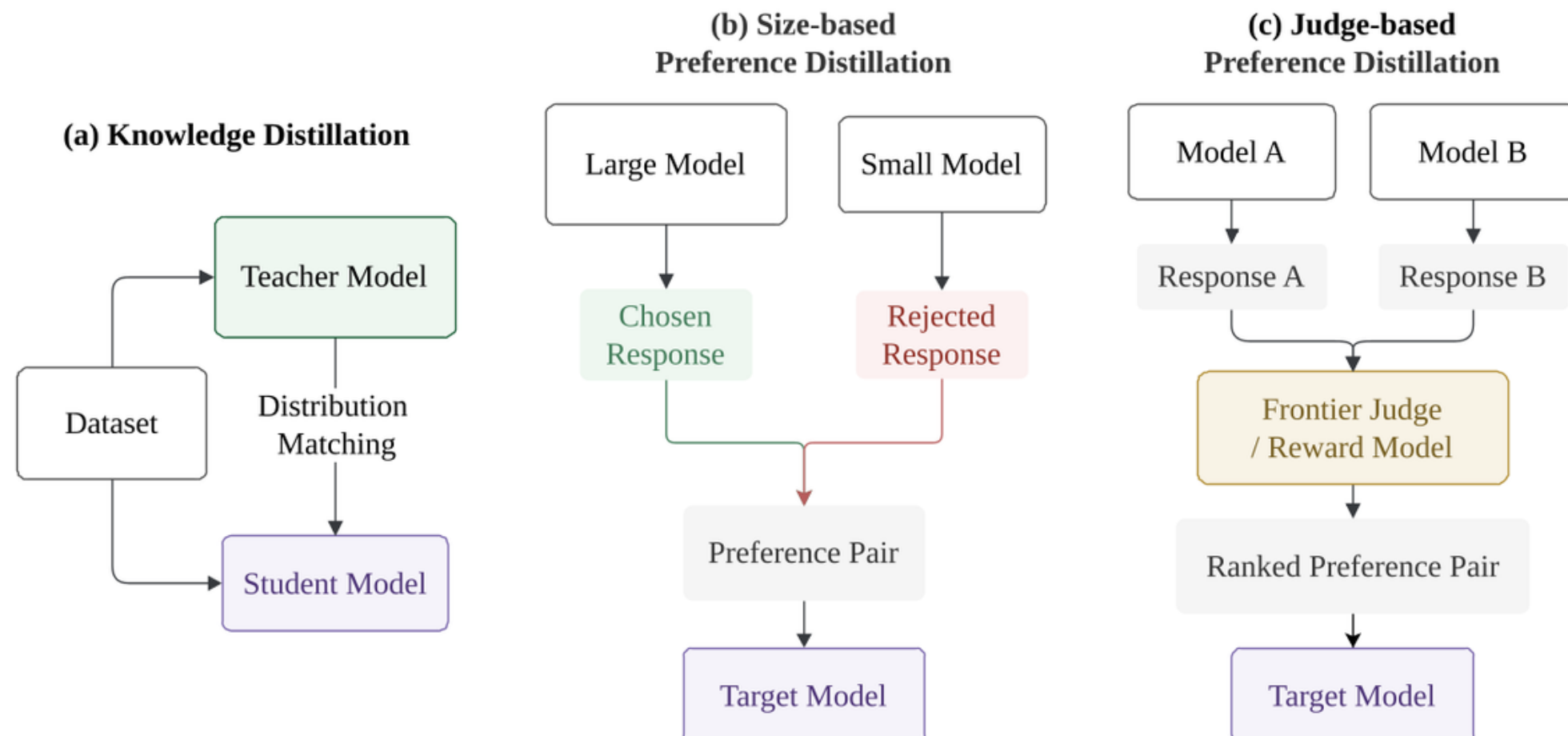
Chi-Ping Su^{1,*} Jing-Hong Hu^{1,*} Ching-Hsun Tseng² Shin-Jye Lee¹

¹National Yang Ming Chiao Tung University ²Delta Research Center



[Paper Link Here!](#)

Key Insights



Existing knowledge transfer or data construction pipelines typically assume that supervision either originate from a globally stronger model, or evaluated by a reward model.

However, different models exhibit distinct response level preference patterns even within the same model family, which may induce complementary learning signals for the target model.

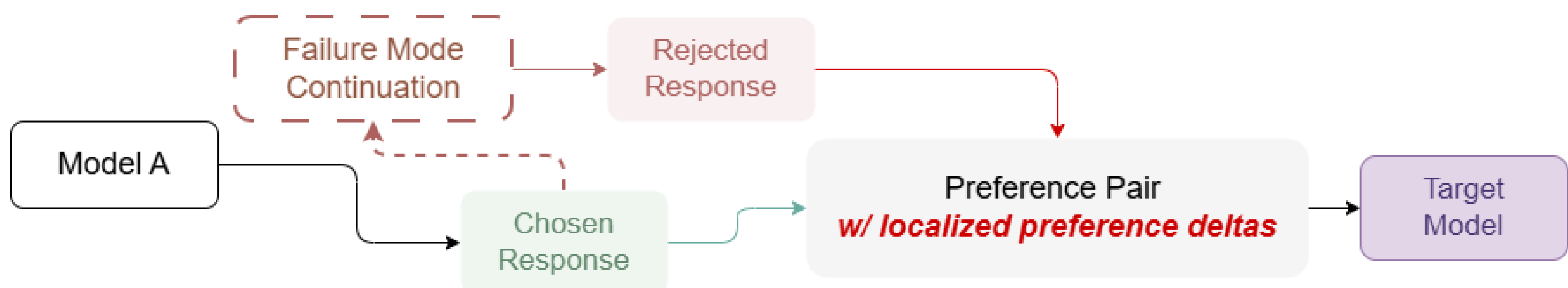
Methodology

Delta Preference Transfer (DPT) Constructs localized preference deltas within a single generator rollout through controlled failure-mode generation.

DPT preference pair construction

$$y^+ \sim \pi_{gen}(\cdot | x)$$

$$y^- \sim \pi_{gen}(\cdot | x, y^+, f)$$



Designed Failure Modes (f)

Logical Inconsistency.

Initially follow a correct reasoning trajectory before introducing a contradictory inference rule.

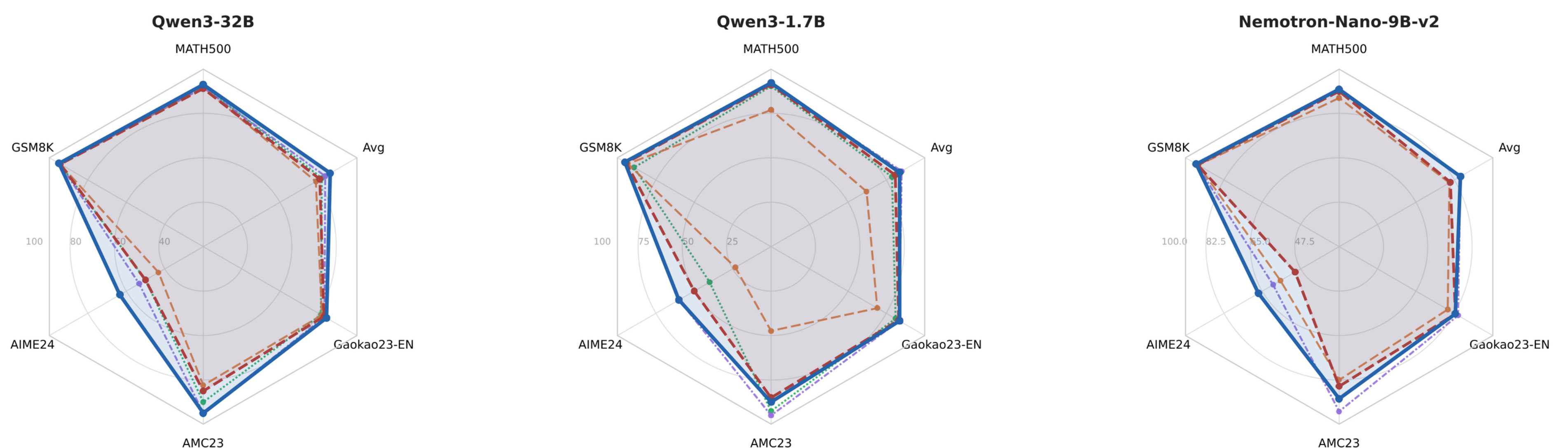
Premature Conclusion.

Terminating the reasoning process too early rather than introducing incorrect intermediate reasoning

Subtle Hallucination.

The examples preserve mathematically correct reasoning while introducing fabricated external justification.

Experiments



Results on math benchmarks across different generators with Qwen3-4B as target model

— Base — KD (SKL) — Size-based — Judge-based — Ours

Analysis and Conclusion

Subset	Base	Ours	Δ
Generator pref.	7.07	3.09	−56.3%
Target pref.	13.61	12.58	−7.6%

Conducted on the reward margin shown above (Qwen3-4B vs. Nemotron-Nano-9B-v2), the finetuned target model shows a distribution closer to that preferred by the generator, while largely preserving its original distribution on prompts where it was already preferred.

This suggests that DPT transfers behaviors associated with complementary strengths without relying on an external reward model, achieved entirely through localized preference deltas.