Mitigating Premature Exploitation in Particle-based Monte Carlo for Inference-Time Scaling

Entropic Particle Filtering (ePF) | ICML 2026

Giorgio Giannone, Guangxuan Xu, Nikhil Nayak, Rohan Awhad,
Shivchander Sudalairaj, Kai Xu, Akash Srivastava

AI Innovation Team, Red Hat & Core AI, IBM

Inference-Time Scaling with Particle Filtering

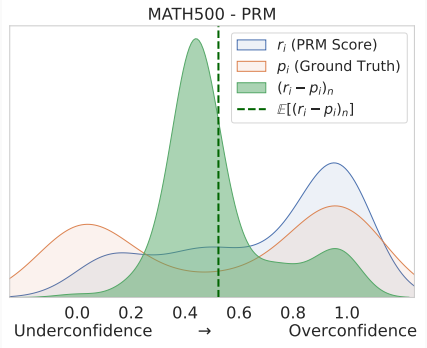
Inference-Time Scaling (ITS): allocate more compute at generation time to improve LLM reasoning.

Particle Filtering (PF) is a principled ITS method:

1. **Propagate** N candidate trajectories
2. **Weight** each via a Process Reward Model
3. **Resample** proportionally to weights

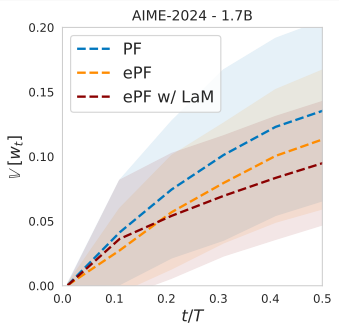
The Problem

PRM-guided PF suffers from **premature exploitation**: overconfident early scores cause particle collapse.

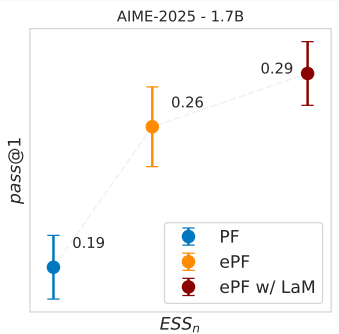


PRM rewards vs. ground-truth correctness on MATH500.

PRM Overconfidence Causes Particle Impoverishment



High $\mathbb{V}[w_t]$ on AIME 2024 leads to greedy-like search.



Task success correlates with high ESS.

The Problem

PRM Overconfidence $\rightarrow$ High $\mathbb{V}[w_t]$ $\rightarrow$ Low ESS $\rightarrow$ **Particle Impoverishment**

Our Approach: Entropic Particle Filtering (ePF)

Core Hypothesis

Dynamically preserving search diversity + **forward-looking guidance** $\Rightarrow$ resilience to reward overconfidence.

ePF integrates two complementary mechanisms:

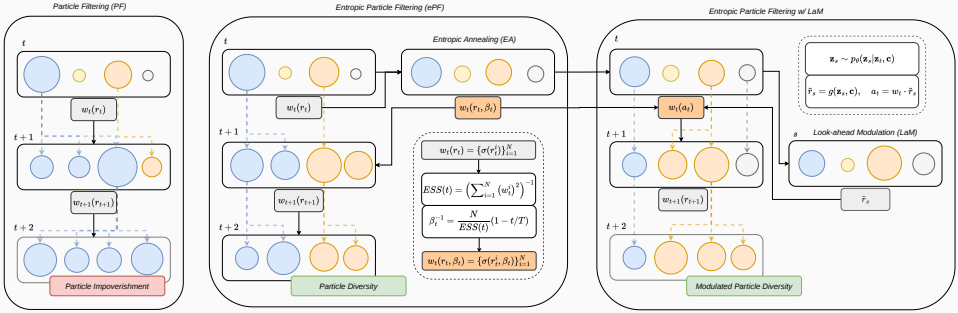
1. Entropic Annealing (EA)

Monitors diversity via ESS. When diversity drops, **raises temperature** to flatten resampling and preserve exploration.

2. Look-ahead Modulation (LaM)

One-step **predictive guide**: re-weights particles by predicted successor quality, making resampling less myopic.

The ePF Pipeline



Left: Standard PF (particle collapse). **Center:** ePF + EA (diversity preserved). **Right:** ePF + LaM.

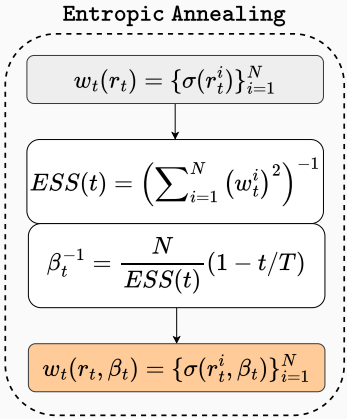
Entropic Annealing (EA)

Adaptive temperature modulated by diversity:

$$\beta_t^{-1} = \frac{N}{ESS(t)}\,(1 - t/T)$$

$$w_t^i = \frac{\exp(r_t^i \cdot \beta_t)}{\sum_j \exp(r_t^j \cdot \beta_t)}$$

- Low ESS $\Rightarrow$ high temp $\Rightarrow$ **explore**
- $t \to T$: anneals to 1 $\Rightarrow$ **exploit**
- Systematic resampling (low-variance)



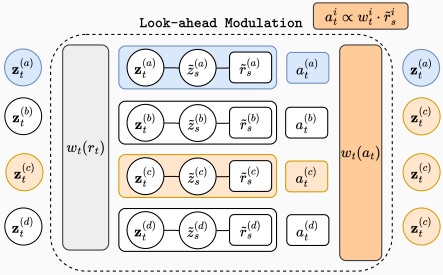
EA flattens the resampling distribution when diversity is low.

Look-ahead Modulation (LaM)

PF is **myopic**. LaM evaluates future potential:

$$\mathbf{z}_s^i \sim p_\theta(\mathbf{z}_s^i \mid \mathbf{z}_t^i, \mathbf{c})$$
$$a_t^i = w_t^i \cdot \tilde{r}_s^i$$
$$w_t^i = a_t^i / \textstyle\sum_j a_t^j$$

- Look-ahead states **discarded** after use
- Triggers in only **10–12%** of steps

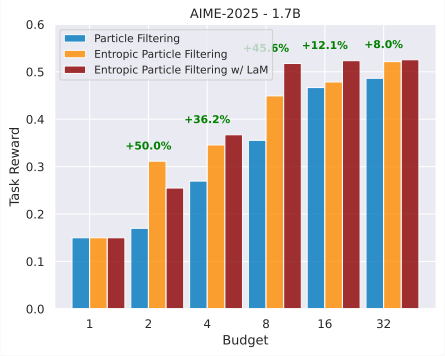


LaM biases resampling toward high-potential trajectories.

Benchmark Results

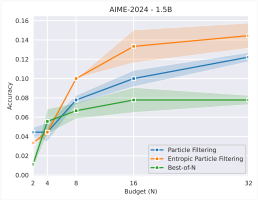
	Qwen2.5-7B			
	GSM	M500	DM	OM
Base	93.4	60.9	23.4	8.6
Self-Cons.	94.8	65.6	30.5	9.4
BoN	96.0	68.0	32.0	9.4
Beam	**96.2**	66.4	32.0	**10.9**
PF	**96.2**	70.3	34.4	10.2
ePF	95.8	**71.1**	**35.9**	**10.9**

Top-1 accuracy (%). M500=MATH500, DM=DEEPMATH, OM=OMNIMATH.

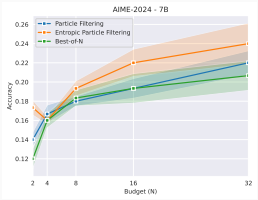


ePF and ePF w/ LaM on AIME-2025 (Qwen3-1.7B).

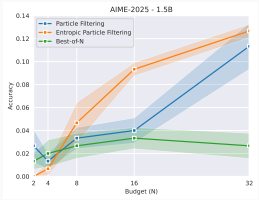
Scaling with Budget



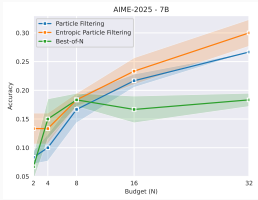
AIME 2024 (Qwen2.5-1.5B)



AIME 2024 (Qwen2.5-7B)



AIME 2025 (Qwen2.5-1.5B)

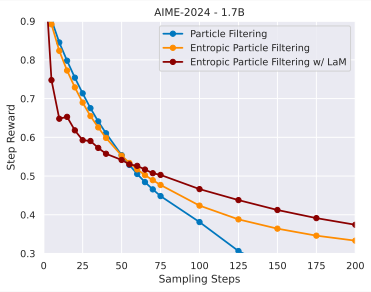


AIME 2025 (Qwen2.5-7B)

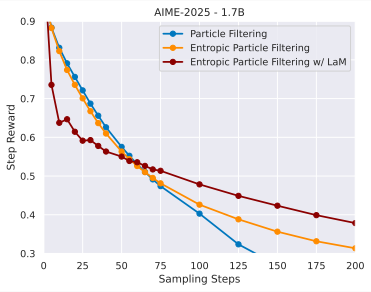
Key Finding

ePF scales better than PF and BoN increasing the particle budget.

Exploration Analysis



AIME 2024 (Qwen2.5-1.5B)

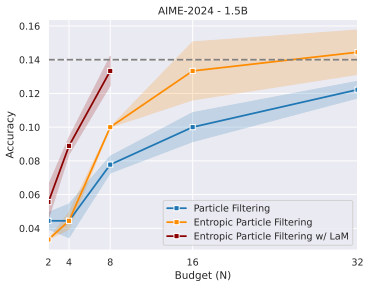


AIME 2025 (Qwen2.5-1.5B)

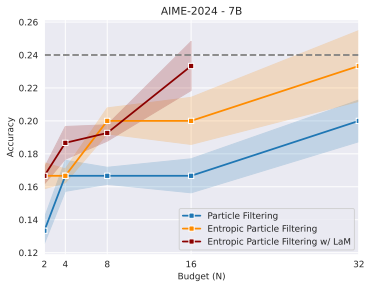
Insight

ePF's step rewards **initially dip** as EA forces exploration, but this investment discovers superior trajectories, leading to higher overall task rewards.

Particle Efficiency



AIME 2024 (Qwen2.5-1.5B)



AIME 2024 (Qwen2.5-7B)

Insight

ePF w/ LaM reaches comparable performance to standard ePF ($N = 32$) using significantly fewer particles (8 particles for the 1.5B model, and 16 for the 7B model), demonstrating **superior computational efficiency**.

Summary

Contributions

1. ePF: entropy-aware adaptive temperature prevents particle collapse
2. LaM: one-step predictive guidance overcomes myopia
3. Diagnostic: premature exploitation correlates with poor performance
4. Results on AIME, MATH500, DEEPMATH, OMNIMATH; up to **50% relative gain**

Key Takeaway

Better early exploration under constrained budgets yields better performance and particle efficiency.

Thank you!

ggiorgio@mit.edu