Intrinsic Selection and Particle Resampling for Inference-Time Scaling Beyond Domain Verifiability

iS | *iPF* | *dPF* | NeurIPS 2026

Giorgio Giannone, Mustafa Eyceoz, Shabana Baig,
Shivchander Sudalairaj, Anna C. Doris, Faez Ahmed, Akash Srivastava, Kai Xu







Inference-Time Scaling Needs a Verifier

Inference-time scaling (ITS): sample many candidates, then select or steer toward the best.

It works where checking is **cheap**: symbolic parsing, unit tests, output matching.

Elsewhere, checking is **slow, subjective, or unavailable**: expert rubrics and LLM judges, or geometric solvers that take up to **40%** of the CAD generation loop.

Domain	Verification	Datasets
Mathematics	Symbolic	AIME, HMMT
Reasoning	Output	GPQA-Diamond
Coding	Execution	LiveCodeBench-v6
Healthcare	Semantic	HealthBench-Hard
Engineering	Solver	DeepCAD, Fusion360

Verification gets more expensive down the table.

Question

Can we guide inference-time scaling using only the **intrinsic statistics** of the generated sample set, with no verifier, reward model, or ground truth?

Three Regimes, Three Methods

With n correct samples among N:

Solved-easy

n/N sizable.

The answer is in the pool.

Select it.

→ **Intrinsic Selection** (*iS*)

Solved-hard

$n/N \ll 1$.

Correct paths are rare.

Reallocate compute to them.

→ **Intrinsic Particle Filtering** (*iPF*)

Systematic failure

$n/N \approx 0$.

No selection rule can help.

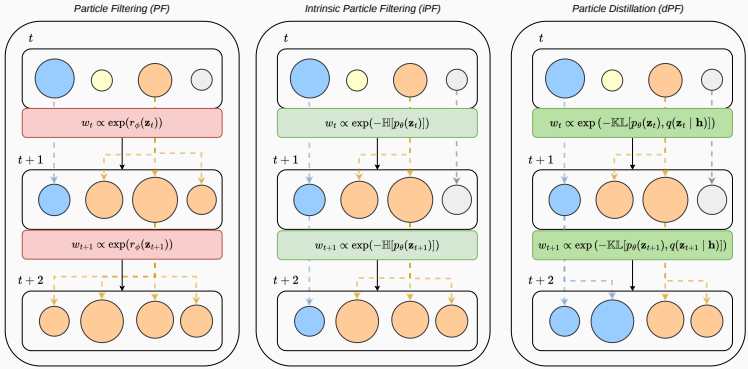
Change the sampling distribution.

→ **Particle Distillation** (*dPF*)

Difficulty gate

n is unknown at deployment. The **mean tail entropy of the sample set** estimates the regime: low-entropy problems go to *iS*, high-entropy ones to *iPF* or *dPF*.

Framework: Particle Filtering Without a Reward Model



PF: reweight by a process reward model. **iPF:** reweight by the model's own token entropy. **dPF:** reweight by agreement with a guide that sees a hint or rubric **h**.

Intrinsic Selection (iS): Adaptive Tail Entropy

Entropy is high during **exploration** for right and wrong trajectories alike. The **answer region** at the end is where a confident model settles, so ***iS*** scores only an adaptive tail.

Score the last t_c tokens, pick the lowest

$$t_c = \frac{\sqrt{N^{-1}\sum_n l_n}}{N^{-1}\sum_n \bar{\mathbb{H}}_n}, \qquad \mathcal{A}^{(n)} = \{t \mid l_n - t_c < t \le l_n\},$$

$$s_n = \underset{t \in \mathcal{A}^{(n)}}{\text{median}}\, \mathbb{H}\big[\mathbf{z}_t^{(n)} \mid \mathbf{z}_{<t}^{(n)}\big], \qquad \hat{n} = \arg\min_n s_n.$$

- Tail grows with $\sqrt{\text{mean length}}$
- Tail shrinks when the set is uncertain
- Set means over the 5th–95th length percentile

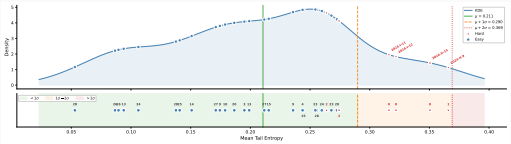
- Median is robust to formatting spikes
- **No** parsing, canonical forms, or reward model
- Works on open-ended outputs

Is Tail Entropy a Useful Signal?

Within a problem: lower tail entropy ranks a correct candidate first on **68–92%** of eligible problems.

When correct answers are rare ($0 < n/N \leq 0.25$):

Model	*iS*@1	SC	Random
Qwen3-4B	**25.0**	14.3	11.2
Gemma4-E2B	**23.1**	3.8	8.2



Set-level tail entropy flags hard problems: AUROC > 0.79, $> 90\%$ precision on AIME-2026 and HMMT-2026.

Confidence is not correctness

The signal ranks well but is not calibrated: a frequent, fluent misconception can be confidently wrong.

Results: Intrinsic Selection vs. Self-Consistency

Method	A24	A25	A26	H25	H26	GPQA	LCB	avg	med	wt avg
pass@1	56.5	45.3	50.0	30.2	28.5	42.3	19.0	38.8	42.3	37.0
SC@32	**72.5**	**60.8**	**65.8**	30.8	29.5	44.1	18.2	**46.0**	44.1	40.8
iS@1 (ours)	67.5	49.9	54.2	**35.0**	**34.8**	**47.7**	**21.5**	44.4	**47.7**	**42.0**
pass@32	87.5	75.8	83.3	56.7	52.3	79.3	35.0	67.1	75.8	66.6

Accuracy (%), Qwen3-4B-Instruct-2507, $N = 32$, four seeds. SC and *iS* select from the same pool; pass@1 is a random pick from it; pass@32 is the oracle.

Where SC wins

AIME: integer answers are easy to vote on, and the correct answer is often the most frequent one.

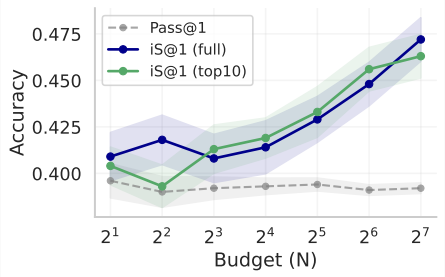
Where iS wins

HMMT, GPQA, LiveCodeBench: the correct answer is often not the majority. *iS* also applies to open-ended outputs, where voting is undefined.

Engineering Design: Image-to-CAD

A fine-tuned VLM writes **CadQuery** code from a rendered part. Scoring a candidate (IoU) requires running a **geometric kernel**.

	DeepCAD (ID)		Fusion360 (OOD)	
Method	Mean	Median	Mean	Median
pass@1	0.62	0.75	0.41	0.39
iS@1	**0.66**	**0.76**	**0.45**	**0.47**
pass@128	0.89	0.98	0.83	0.88



Fusion360 median IoU as the pool grows. *iS* improves with *N*; a random pick does not.

Takeaway

iS picks better designs **without executing a single candidate**.

Intrinsic Particle Filtering (iPF)

Every $B = 128$ tokens, reweight particles by the entropy of their latest step, standardized across the population:

$$w_t \propto w_{t-1} \exp\big(-\mathbb{H}[\mathbf{z}_t]\big), \qquad \tilde{w}_t = \frac{\mathbb{H}[\mathbf{z}_t] - m(\mathbb{H}[\mathbf{z}_t])}{s(\mathbb{H}[\mathbf{z}_t]) + \epsilon}, \qquad w_t = \mathrm{softmax}(-\tilde{w}_t).$$

- 16 particles advance in lockstep
- Systematic resampling when the ESS drops
- Stochastic, not top-k: uncertain particles can survive
- Uses logits the sampler already computes

A heuristic, stated plainly

iPF targets **confidence**, not correctness. Batch standardization acts as adaptive tempering, so the usual SMC unbiasedness guarantees do not apply.

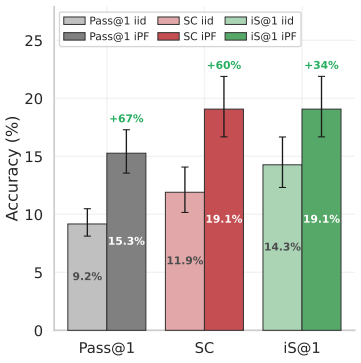
iPF on the Hardest 25% of AIME

i.i.d. → *iPF*	AIME 24	AIME 25	AIME 26
pass@1	17.0 → 22.8	4.9 → 8.3	5.6 → 14.7
SC@16	21.4 → 28.6	10.7 → 10.7	3.6 → 17.9
iS@1	25.0 → 25.0	7.1 → 17.9	10.7 → 14.3

Same budget of 16 particles; average over 10 seeds. pass@1 is the expected accuracy of a random final particle.

Result

+6.1 points in pass@1 on average at the same budget. Particles concentrate on good paths, so SC benefits too (3.6 → 17.9 on AIME 26).



Average over the three AIME years.

Particle Distillation (dPF)

A guide $q(\cdot \mid \mathbf{c}, \mathbf{h})$: the same model conditioned in context on a hint or rubric $\mathbf{h}$.

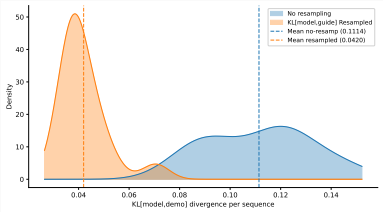
1. **Early logit blending** (α_t annealed toward 0):

$$\text{logits}(\mathbf{z}_t^h) = (1 - \alpha_t)\,\text{logits}(\mathbf{z}_t, \mathbf{c}) + \alpha_t\,\text{logits}(\mathbf{z}_t, \mathbf{c}, \mathbf{h})$$

2. **KL-guided resampling:**

$$w_t \propto w_{t-1} \exp\big(-\mathrm{KL}\big[p_\theta(\cdot \mid \mathbf{z}_{<t}, \mathbf{c}) \,\|\, q(\cdot \mid \mathbf{z}_{<t}, \mathbf{c}, \mathbf{h})\big]\big)$$

Inference-time only: no parameter updates, no access to the ground-truth answer.



Unsolved AIME 2024-II-15: guided resampling shifts the mean KL from 0.111 to 0.042.

dPF: Math Hints and Clinical Rubrics

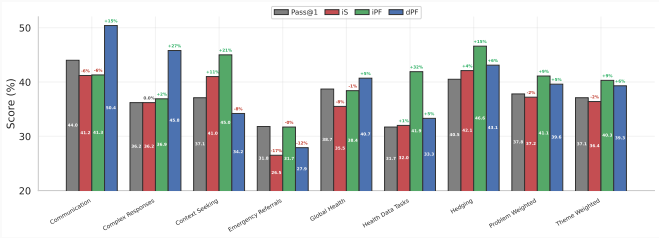
Math (AIME, HMMT)

Hints from Qwen3.5-27B, with final answers removed.

Weighted pass@1:
32 i.i.d. samples: 41.6
dPF, 8 particles: 46.4

HealthBench-Hard

iPF: best overall (+8.7%)
dPF: communication +14.5%, complex responses **+26.5%**
iS alone: −1.5%



HealthBench-Hard by theme (MedGemma-4B; judge MedGemma-27B).

Confidence alone does not ensure rubric coverage; steering does.

Summary

Contributions

1. **iS**: verifier-free selection by adaptive tail entropy, competitive with self-consistency
2. **iPF**: entropy-guided step-level resampling, +6.1 points on the hardest AIME problems
3. **dPF**: hints or rubrics via logit blending and KL-guided resampling, +26.5% on complex clinical responses
4. A **difficulty gate** from set-level entropy that routes each problem to the right method

Limitations

Tail entropy measures **confidence, not correctness**. The gate costs N samples up front, and *dPF* needs privileged information.

Thank you!







ggiorgio@mit.edu