Unveiling the Entropy Dynamics of Chain-of-Thought Reasoning

Ting Xu, Xu He, Yupu Lu, Jiankai Sun, Dong Li, Wai Lam, Jianye Hao

Presenter: Ting Xu

ICML 2026

CoT Reasoning: Background & Research Gap

CoT: A General Reasoning Paradigm

Mathematics, coding, science, logic — CoT unlocks complex reasoning
Enables constant-depth transformers to simulate size-T boolean circuits (Liu et al.)

Existing Work: Local Properties Only

Treat CoT as discrete segments; study individual steps/layers:

→ Token-level confidence for early exit (DEER)

→ Consecutive answer agreement for stopping (Dynasor)

→ Mechanistic analysis of single-step representations (CoT Vectors)

→ Markov chain modeling of individual step roles (Yu et al.)

→ All focus on LOCAL features — no global dynamics understanding

Research Gap

Problem: Reasoning is not isolated steps — it is a continuous process evolving over time

Gap: No systematic understanding of GLOBAL temporal dynamics of CoT reasoning

Question: How does reasoning EVOLVE from exploration to convergence? Is it gradual or abrupt?

★ Our Approach

First systematic analysis of GLOBAL entropy dynamics throughout the full CoT generation process — pinpointing how reasoning evolves from exploration to convergence.

Key Discovery: Two-Phase Structure of CoT Reasoning

What is Predictive Entropy?

At every k=128 tokens, we probe the model by inserting an answer-inducing prompt and let it generate an intermediate answer A_i. The predictive entropy is:

$$H(A_i|X,T_{1:i}) = -(1/m_i)\,\Sigma_j\,\Sigma_u\,P_j(u|\cdot)\,\log P_j(u|\cdot)$$

High entropy → model explores multiple competing hypotheses
Low entropy → model has converged to a confident, stable answer

Figure 1. **Dynamics of Entropy and Accuracy on Qwen3-4B-Thinking-2507**: CoT reasoning exhibits a two-phase structure: (1) an Uncertainty Region, where high entropy and diverse answers reflect exploration of multiple logical paths, and (2) a Confidence Region, where entropy collapse and answer stabilization signal convergence toward a reliable solution.

Uncertainty Region

High entropy: stochastic exploration of multiple hypotheses
Low accuracy: < 20%, diverse answers (5+ unique answers)

Confidence Region

Low entropy: abrupt collapse, deterministic convergence
High Reliability: accuracy surges to > 60%
High Redundancy: model generates > 30% extra tokens beyond the correct answer

Why Detect the Confidence Region?

The two properties of the Confidence Region unlock two powerful inference strategies

High Reliability

Once the model enters the Confidence Region, accuracy surges from < 20% to > 60% and stabilizes. Answers become trustworthy — safe to act on.

→ We can safely STOP generation here

High Redundancy

Models continue generating > 30% extra tokens after reaching the correct answer. These tokens add computation cost but NO accuracy gain.

→ We can PRUNE these redundant tokens

Early Exit — Efficient & Reliable Inference

Exploit High Reliability to safely terminate computation the moment the model enters the Confidence Region. Exploit High Redundancy to prune wasteful tokens — continued generation yields near-zero accuracy gains.

★ 11.1% token reduction | accuracy preserved

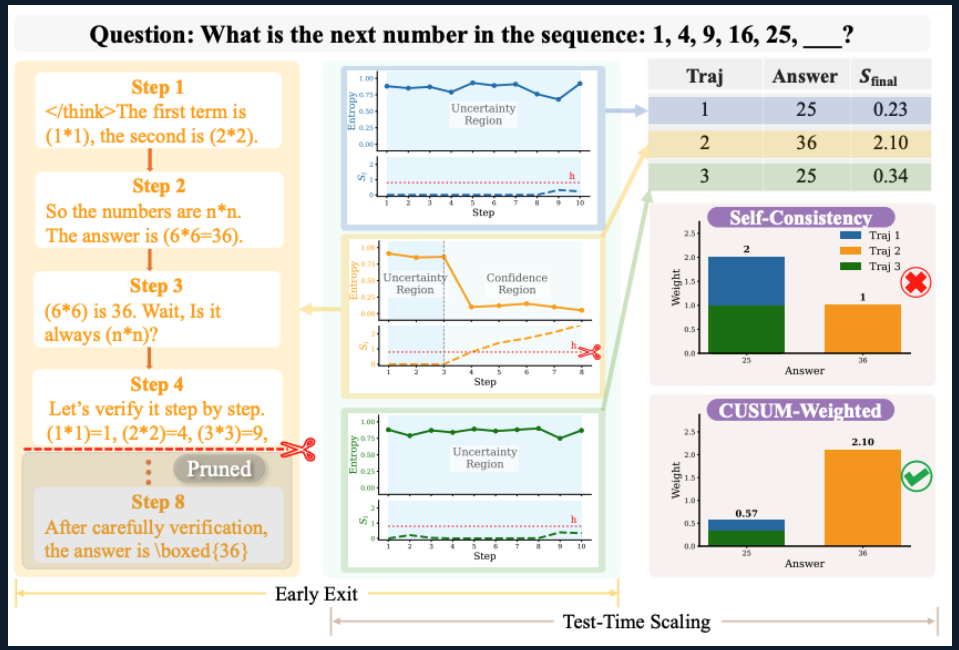
Test-Time Scaling — Reliable & Reliable Inference

Use the Confidence Region as a trajectory quality signal. Trajectories that converge (finite v) have stronger reasoning than those stuck in Uncertainty ($v=\infty$). Weight answers by CUSUM score instead of uniform voting.

★ Consistently outperforms self-consistency at scale

Confidence Region Detection via CUSUM

CUSUM detects the Uncertainty→Confidence transition for Early Exit and Test-Time Scaling



The CUSUM Algorithm

Formulation

Change-point detection on entropy sequence $H_1, H_2, \ldots$
Two regimes: f_0 (Uncertainty) and f_1 (Confidence)

Statistic

$S_i = \max(0, S_{i-1} + \log f_1(H_i)/f_0(H_i))$
Stop when $S_i \geq$ threshold h

Guarantees

Minimax optimal detection delay (Lorden 1971)
Controllable false-alarm rate: $E_\infty[\tau] \geq \gamma$

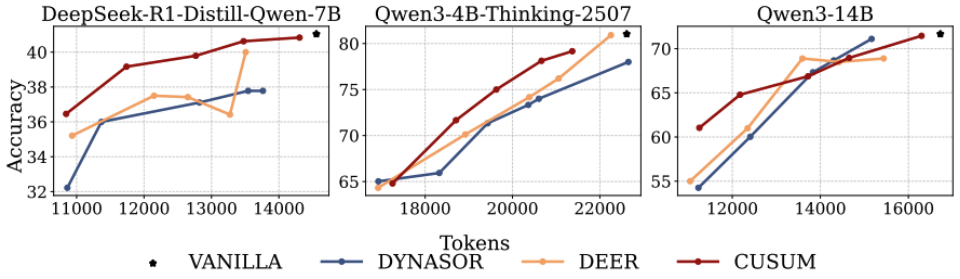
Training-free

Calibrated on 100 trajectories via histogram density estimation — no model fine-tuning required

Early Exit Results: Superior Efficiency–Accuracy Trade-off

Table 1. Early Exit Results on AIME25, AIME24, and GPQA-Diamond across three models, showing accuracy (%) and token count with reduction percentage.

Methods	AIME25		AIME24		GPQA		Average	
	Acc ↑	Tokens ↓	Acc ↑	Tokens ↓	Acc ↑	Tokens ↓	Acc ↑	Tokens ↓
DeepSeek-R1-Distill-Qwen-7B								
Vanilla	41.04	14556	55.63	13313	38.38	8637	45.02	12169
DEER	37.5	12148 ↓16.5%	46.67	**10756** ↓19.2%	35.73	8227 ↓4.8%	39.97	10377 ↓14.7%
Dynasor	37.11	12822 ↓11.9%	52.67	11215 ↓15.8%	19.32	8342 ↓3.4%	36.37	10793 ↓11.3%
Ours	**39.17**	**11740** ↓19.3%	**53.75**	10912 ↓18.0%	**40.40**	**8191** ↓5.2%	**44.44**	**10281** ↓15.5%
Qwen3-4B-Thinking-2507								
Vanilla	81.04	22613	77.29	19178	64.65	9442	74.32	17078
DEER	76.2	21056 ↓6.9%	76.0	18925 ↓1.3%	64.65	9082 ↓3.8%	72.28	16354 ↓4.2%
Dynasor	74	20702 ↓8.4%	75.56	**18709** ↓2.4%	63.14	**8968** ↓5.0%	70.9	16127 ↓5.6%
Ours	**78.13**	**20663** ↓8.6%	**76.88**	18714 ↓2.4%	**64.9**	8980 ↓4.9%	**73.3**	**16119** ↓5.6%
Qwen3-14B								
Vanilla	71.67	16727	81.46	13872	67.93	5988	73.69	12196
DEER	68.89	15438 ↓7.7%	70.42	11454 ↓17.4%	62.0	5942 ↓0.8%	67.1	10944 ↓10.3%
Dynasor	68.27	**14305** ↓14.5%	76.22	12508 ↓9.8%	61.99	5902 ↓1.4%	68.83	10905 ↓10.6%
Ours	**68.96**	14653 ↓12.4%	**77.71**	**12141** ↓12.5%	**67.68**	**5720** ↓4.5%	**71.45**	**10838** ↓11.1%



Average Performance
(3 models × 3 benchmarks)

63.06%

CUSUM Accuracy

59.78%

DEER Accuracy

58.70%

Dynasor Accuracy

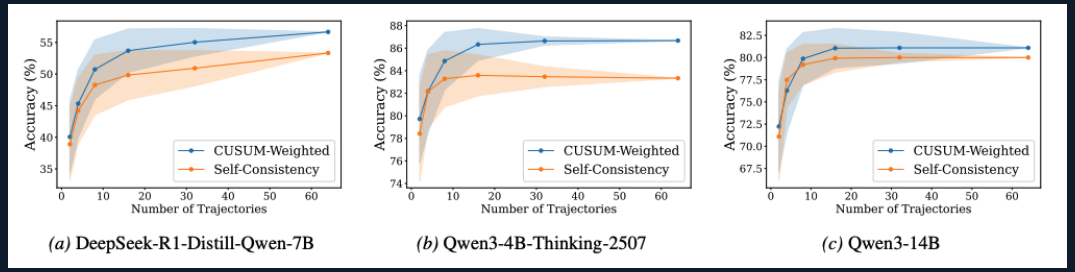
11.1%

Token Reduction
(CUSUM)

+3.28% vs. DEER | +4.36% vs. Dynasor in accuracy

Test-Time Scaling: CUSUM-Weighted Voting

CUSUM-Weighted consistently outperforms Self-Consistency on AIME25 across 3 models (N=2,4,8,16,32,64)



(a) DeepSeek-R1-Distill-Qwen-7B *(b)* Qwen3-4B-Thinking-2507 *(c)* Qwen3-14B

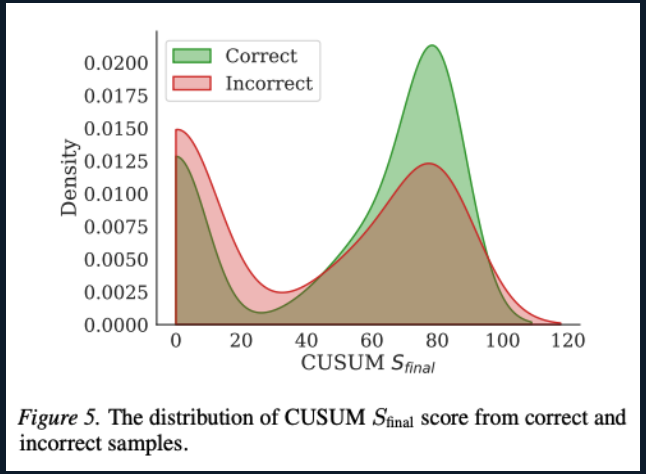


Figure 5. The distribution of CUSUM S_{final} score from correct and incorrect samples.

Key Insight: Sfinal as Quality Indicator

Final CUSUM score Sfinal measures cumulative evidence that the trajectory converged to the Confidence Region. Higher Sfinal correlates strongly with correct answers — the distributions of correct vs. incorrect trajectories are clearly separable.

Gap Widens with N

For Qwen3-4B, improvement over self-consistency grows from slim margin at N=2 to +3.33% at N=64. As more samples available, prioritizing converged trajectories over stalled-in-uncertainty ones becomes increasingly critical.

Why Self-Consistency Fails

Equal weighting dilutes the correct signal with noise from hallucinated or unconverged trajectories. CUSUM score acts as a convergence filter — more effective use of compute at scale.

Summary

ICML 2026 | Seoul, South Korea

Entropy Dynamics of CoT

✓ First systematic analysis of entropy dynamics in full CoT trajectories

✓ CUSUM: 63.06% acc. + 11.1% token reduction; outperforms DEER/Dynasor

✓ CUSUM-Weighted voting consistently beats self-consistency at scale