

The Evolution of Thought: Tracking LLM Overthinking via Reasoning Dynamics Analysis

Zihao Wei^{1,2}, Liang Pang^{1,†}, Jiahao Liu, Wenjie Shi, Jingcheng Deng^{1,2}, Shicheng Xu^{1,2}, Zenghao Duan^{1,2}, Jingang Wang, Fei Sun¹, Huawei Shen¹, Xueqi Cheng¹

¹State Key Laboratory of AI Safety, Institute of Computing Technology, CAS

²University of Chinese Academy of Sciences

{weizihao22z, pangliang}@ict.ac.cn

Abstract

Test-time scaling via explicit reasoning trajectories significantly boosts large language model (LLM) performance but often triggers overthinking. To explore this, we analyze reasoning through two lenses: **Reasoning Length Dynamics**, which reveals a compensatory trade-off between thinking and answer content length that eventually leads to thinking redundancy, and **Reasoning Semantic Dynamics**, which identifies semantic convergence and repetitive oscillations. These dynamics uncover an instance-specific Reasoning Completion Point (RCP), beyond which computation continues without further performance gain. Since the RCP varies across instances, we propose a Reasoning Completion Point Detector (RCPD), an inference-time early-exit method that identifies the RCP by monitoring the rank dynamics of termination tokens (e.g., `</think>`). Across AIME and GPQA benchmarks using Qwen3 and DeepSeek-R1, RCPD reduces token usage by up to 44% while preserving accuracy, offering a principled approach to efficient test-time scaling.

1 Introduction

Test-time scaling via explicit reasoning trajectories has significantly enhanced the problem-solving capabilities of Large Language Models (LLMs) (DeepSeek-AI et al., 2025; Yang et al., 2025a; Team et al., 2025). While longer trajectories are generally associated with improved performance, recent studies reveal that this relationship is not strictly linear. Instead, excessive reasoning often leads to diminishing returns or even performance degradation, a phenomenon widely referred to as overthinking (Ghosal et al., 2025; Han et al., 2025).

This non-monotonic scaling behavior (Wu et al., 2025b) motivates a fundamental question: how does each step of the reasoning process incrementally influence the final outcome? Despite its importance, such a fine-grained analysis of the in-

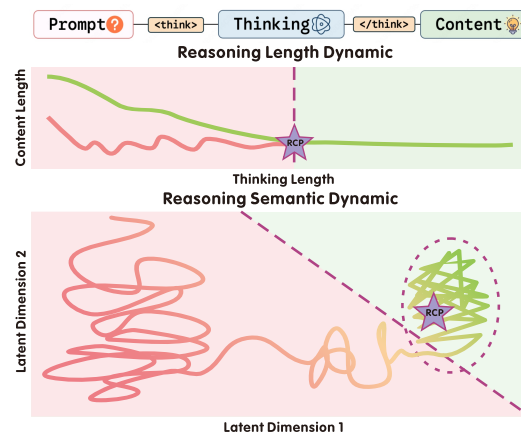


Figure 1: Overview of Reasoning Dynamics and RCP. The top panel summarizes Reasoning Length Dynamics, where content length decreases as thinking length grows under a thinking-content compensation regime until reaching the RCP. The bottom panel summarizes Reasoning Semantic Dynamics, where the latent semantic trajectory transitions from broad exploration to a stable neighborhood with repetitive oscillations, with the onset of convergence aligning with the RCP. The top and bottom panels are defined in §2 and §3, respectively.

termediate reasoning steps remains insufficiently explored. By quantifying the functional impact of each step on the final answer, we implement a study of reasoning dynamics to uncover the internal mechanisms of overthinking and identify the regime where further reasoning tokens become redundant. To explore this, we analyze the reasoning dynamic of LLM from two complementary perspectives.

First, we explicitly characterize the inference as the interaction between the thinking process and the content process. Motivated by recent findings on the non-linear relationship between reasoning length and accuracy (Su et al., 2025; Wu et al., 2025a), we investigate their **Reasoning Length Dynamics** and uncover a phenomenon we term *thinking-content compensation*, as shown in Figure 1. Initially, the thinking process and content

generation exhibit a compensatory relationship: as thinking trajectories expand, the subsequent content becomes increasingly concise. However, this dynamic eventually transitions into a saturation phase. In this regime, although the model has acquired sufficient information to formulate a correct answer, the thinking process continues to lengthen while the content remains unchanged, leading to redundancy.

Second, to uncover the underlying mechanism driving these surface-level length patterns, we further analyze the trajectory of high-dimensional representations in the semantic space (Li et al., 2025). In terms of **Reasoning Semantic Dynamics**, we identify *semantic path convergence*, as shown in Figure 1. We observe that these representations initially exhibit significant fluctuations, reflecting active exploration of the solution space. Subsequently, the inference path enters a convergence regime where the semantic representations converge toward a stable neighborhood, exhibiting repetitive oscillations within that localized space. This suggests that the solution hypothesis has stabilized.

Bridging these two dynamics suggests a structural explanation for overthinking: it arises when the model continues to generate reasoning tokens after the process has converged. This motivates a principled boundary between necessary reasoning and redundant computation after convergence. We call this boundary the Reasoning Completion Point (RCP). At the RCP, thinking-content compensation ends and the semantic trajectory begins to converge. This boundary is latent and instance-dependent, so mitigating overthinking requires detecting the RCP online during decoding rather than relying on a fixed reasoning length budget. Based on this formalization, we propose a Reasoning Completion Point Detector (RCPD), which monitors the rank of the special token designed to terminate the thinking process (e.g., `</think>`) to detect the RCP during decoding without additional computational overhead. Detecting the RCP enables early truncation of redundant reasoning, reducing token consumption while maintaining accuracy.

2 Reasoning Length Dynamics

Many reasoning-specialized LLMs expose an intermediate thinking process delimited by special tokens (e.g., `<think>` and `</think>`) and then transition to a content process that produces the final an-

swer (DeepSeek-AI et al., 2025; Yang et al., 2025a). This section asks a simple but diagnostic question: as we allocate more steps to the thinking process, how do the resulting content length and answer correctness change? We study this via an intervention that forces the model to stop thinking at controlled steps and immediately begin generating content.

2.1 Dynamics Setup: Stepwise Truncation

We formalize LLM generation as a sequential coupling of two distinct phases: a *thinking process* t and a *content process* y . Even when the raw trajectory t is withheld in proprietary models due to safety or competitive considerations (Jaech et al., 2024; Google DeepMind, 2025), it remains the indispensable structural precursor that conditions the final output. Formally, the generation follows the joint distribution $P(y, t|x) = P(t|x)P(y|x, t)$, where x denotes the input prompt.

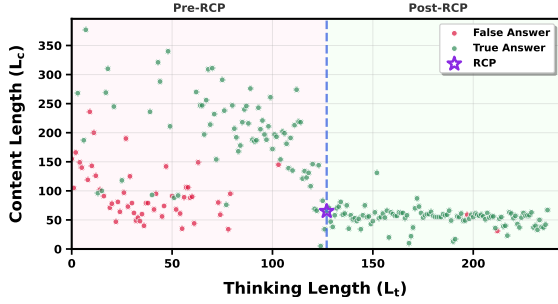
To analyze the reasoning process at a granular level, we decompose both the thinking trajectory t and the content trajectory y into discrete *reasoning steps* s . For a trajectory consisting of raw tokens, we segment it into steps, $t = (s_1, \dots, s_K)$, where each step s_k represents a complete sentence segmented via NLTK (Bird and Loper, 2004). This mapping allows us to intervene on the thinking budget by defining the thinking prefix up to step k as $t_{1:k} = (s_1, \dots, s_k)$.

To explore how each step incrementally influences the outcome, we implement a **stepwise truncation protocol**. For each $k \in \{1, \dots, K\}$, we truncate the thinking trajectory immediately after s_k and force-inject the end-of-thinking delimiter (e.g., `</think>`). This injected delimiter serves as an explicit control signal that terminates the internal thinking process and triggers the transition to content generation:

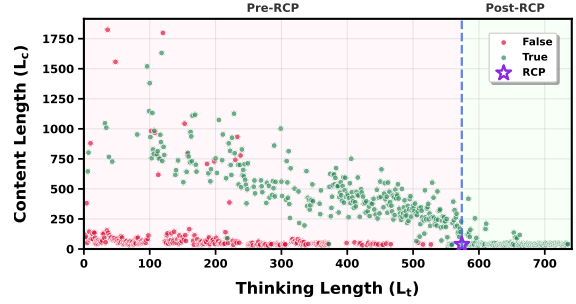
$$y_k \sim P(y | x, t_{1:k}). \quad (1)$$

Under this protocol, we monitor two core length measures quantified by reasoning steps: (1) **Thinking Length** (L_t): The cumulative number of allowed reasoning steps in the thinking phase, $L_t = k$; (2) **Content Length** (L_c): The number of steps in the content process $y^{(k)} = (s_1^y, \dots, s_{L_c}^y)$, where L_c is the count of steps generated in response to the truncated thinking prefix $t_{1:k}$.

By monitoring the evolution of L_t , L_c , and the corresponding answer correctness across k , we implement a dynamical analysis of the reasoning process and its transition toward redundancy.



(a) AIME24 Question 19 Length Dynamics



(b) AIME25 Question 17 Length Dynamics

Figure 2: Two-stage reasoning dynamics separated by RCP: the Pre-RCP Active Reasoning Stage and the Post-RCP Converged Reasoning Stage. The vertical dashed line indicates the RCP boundary. Additional examples are provided in Appendix Figure 6.

2.2 Observations

Figure 2 provides a representative visualization of thinking-content compensation and the eventual stabilization pattern.

Observation 1: Reasoning Compensation

In the early stage, shorter thinking force the model to offload more reasoning into the content to maintain performance, resulting in an increased content length.

In the early region of Figure 2, this effect appears directly as a separation between correct and incorrect samples. In Figure 2a around $k \approx 50$, many incorrect answers cluster at very short content, whereas correct answers at the same k often require substantially longer content. Figure 2b exhibits the same pattern: in the early region, incorrect answers concentrate near the bottom band, while correct answers extend to much larger content steps. These examples show that when the thinking budget is insufficient, maintaining correctness frequently requires expanding the content, and failures often manifest as terse outputs.

Observation 2: Thinking Redundancy

With sufficient thinking, content retracts to its conciseness and accuracy stabilizes; any further ineffective reasoning is manifested solely as an expansion of thinking.

In the late region of Figure 2, content length collapses into a narrow band and stabilizes after the dashed line, remaining constant even as k increases by over 100 steps (Figures 2a, 2b). This confirms that additional thinking budget in this stage no longer alters content volume but primarily accu-

mulates as redundancy.

3 Reasoning Semantic Dynamics

While Section 2 establishes a macroscopic trade-off between thinking and content length, these metrics are mere proxies that do not distinguish between active hypothesis refinement and futile repetition. To uncover the mechanistic trigger of overthinking, we shift from surface-level length statistics to fine-grained semantic-level evolution, namely Reasoning Semantic Dynamics.

Following the conceptualization of reasoning as a trajectory in latent semantic space (Cho et al., 2025), we probe the model’s internal evolution through its intermediate states. Since thinking tokens serve as intermediate computational steps, the most direct manifestation of the model’s current state is the answer it produces when forced to terminate. By tracking the semantic trajectory across truncation steps, we can trace the stabilization of the model’s emerging conclusion and map its search path.

Consequently, this section investigates a pivotal question: as the thinking process extends, how does the induced content evolve semantically, and when does it reach a state of terminal convergence?

3.1 Dynamics Setup: Semantic Lens

To explore the internal evolution of thoughts, we apply a semantic lens to the content distribution induced by stepwise truncation. Following Eq. 1, for each prefix $t_{1:k}$, we sample $N = 32$ independent content continuations $y_k^{(i)N}$ to robustly capture the model’s hypothesis at step k .

Each continuation is mapped to a latent space using the Qwen3 Embedding model (Zhang et al.,

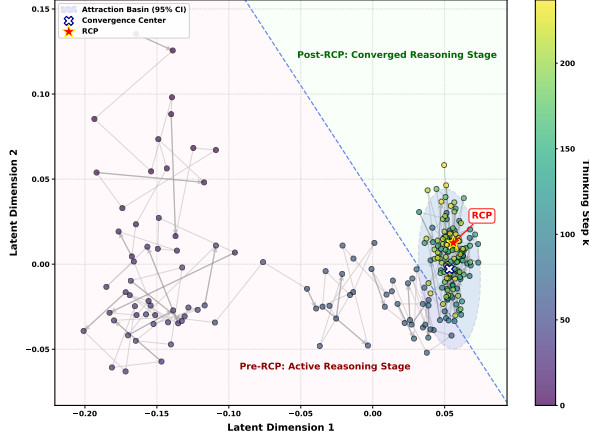


Figure 3: Semantic trajectory for AIME24 Question 19, the same instance as Figure 2a, showing the transition from Pre-RCP Active Exploration to Post-RCP Reasoning Convergence. The dashed line indicates the RCP boundary. Additional examples are provided in Appendix Figure 7.

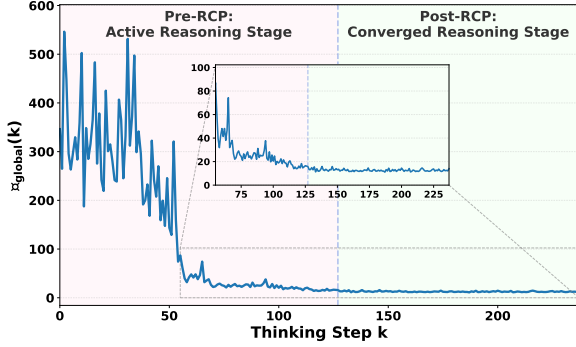


Figure 4: Semantic convergence residual over thinking steps. $\mathcal{D}_{\text{global}}(k)$ declines and then approaches a low plateau. The vertical dashed line indicates the RCP boundary; the inset zooms into the late-step region for readability.

2025b), yielding vectors $v_k^{(i)} = \mathcal{E}(y_k^{(i)}) \in \mathbb{R}^d$. To visualize these dynamics in a consistent coordinate system, we perform Principal Component Analysis (PCA) on the pooled set of embeddings over all truncation steps k and samples i . We project each embedding to a 2D point $\tilde{v}_k^{(i)} \in \mathbb{R}^2$, and the resulting semantic trajectory is traced by the per-step centroid:

$$\tilde{\mu}_k = \frac{1}{N} \sum_{i=1}^N \tilde{v}_k^{(i)}. \quad (2)$$

The evolution of $\tilde{\mu}_k$ captures how the *content semantics* mature as L_t increases. We use the 2D view for interpretability.

High-Dimensional Convergence Metric. The 2D PCA projection is only for visualization. For

quantitative analysis, we operationalize convergence via a semantic proxy distribution in a higher-dimensional PCA space that retains substantially more information than the 2D view. We project embeddings to an r -dimensional PCA space ($r \gg 2$), obtain $z_k^{(i)} \in \mathbb{R}^r$, and approximate the induced distribution by a Gaussian $Q_k = \mathcal{N}(\mu_k, \Sigma_k)$:

$$\begin{aligned} \mu_k &= \frac{1}{N} \sum_{i=1}^N z_k^{(i)}, \\ \Sigma_k &= \text{Cov}\left(\{z_k^{(i)}\}_{i=1}^N\right) + \lambda I, \end{aligned} \quad (3)$$

where λI is a small ridge term for numerical stability. We estimate a terminal reference Q_∞ by pooling samples from a short tail window of the final truncation steps and computing its mean and covariance. We then measure the distance to the converged semantic regime by the global convergence residual

$$\mathcal{D}_{\text{global}}(k) \triangleq D_{\text{KL}}(Q_k \parallel Q_\infty), \quad (4)$$

which will later serve as a key ingredient for formalizing convergence.

3.2 Observations

We use AIME24 Question 19 as a running example because it clearly exhibits the same RCP transition under both the length and semantic views; Figure 3 shows the semantic trajectory of the same instance as Figure 2a, while additional examples are provided in the appendix. We refer to the resulting transition as *semantic path convergence*.

Observation 3: Semantic Exploration

During the initial phase, the semantic within the reasoning process undergo rapid shifts, indicating that the model is actively exploring feasible solution paths.

In the left region of Figure 3, the projected induced samples are sparsely scattered, with large gaps between neighboring points, and the per-step mean $\tilde{\mu}_k$ makes large jumps across truncation steps, suggesting that the model is still actively revising its semantic hypothesis. Figure 4 then quantifies this instability: $\mathcal{D}_{\text{global}}(k)$ varies widely in the early stage, indicating that the induced content distribution can shift markedly between adjacent truncation steps. This semantic instability also explains the length-level outcome bifurcation in Figure 2: at the same moderate k , some runs compensate by

generating longer content and reach the correct hypothesis, while others collapse to short, incorrect outputs.

Observation 4: Reasoning Convergence

As reasoning progresses, the semantics gradually converge toward a specific neighborhood, exhibiting repetitive oscillations within that localized space.

The right region in Figure 3 makes the convergence picture explicit: once the trajectory crosses the dashed line, the induced samples concentrate inside the 95% attraction basin around the convergence center, and $\tilde{\mu}_k$ largely stays within that localized area. To make this stable-neighborhood claim testable, Appendix D fits a 95% confidence ellipse in the 2D PCA plane and finds that 97.3% of post-transition points remain inside the basin.

Figure 4 provides the corresponding high-dimensional signal: after the turning point marked by the dashed line, $\mathcal{D}_{\text{global}}(k)$ approaches a low plateau, indicating that the induced content distribution is already close to its terminal state. Notably, this non-zero floor reflects an intrinsic entropy gap between the instantaneous state Q_k and the aggregate convergence basin Q_∞ , suggesting a dynamic equilibrium within the basin rather than collapse to a single singularity. This also aligns with the right region in Figure 2, where thinking continues to grow while the content length stays almost unchanged. In this regime, extra steps mainly accumulate as redundant thinking and rarely alter the final answer.

These two observations together explain why more thinking can stop helping: once the content semantics have converged, additional thinking steps occur after the induced content distribution has effectively stabilized, making overthinking largely redundant.

Together with the length lens in §2, the semantic trajectory suggests a shared transition from an early stage of active semantic change to a late regime of semantic stationarity.

4 Reasoning Completion Point

The empirical observations in previous sections suggest that LLM reasoning is not a monolithic process but a bipartite evolution. By synthesizing thinking-content compensation and semantic path convergence, we identify a fundamental two-stage

transition in the model’s reasoning: **Pre-RCP: Active Reasoning** and **Post-RCP: Converged Reasoning**.

To formalize these stages, we first establish two quantitative metrics to capture the stability of the reasoning output. Consistent with the observations in § 2, as thinking length increases, the compensation effect eventually saturates. Content length L_c ceases to change materially, and the induced content stabilizes. We formalize this step-to-step content change as:

$$\Delta_{\text{content}}(k) \triangleq |L_c(k) - L_c(k-1)|, \quad (5)$$

where $L_c(k)$ is measured at truncation step k following § 2.1. Content stabilization occurs when $\Delta_{\text{content}}(k)$ drops below a small threshold ϵ_c and remains near zero thereafter.

Simultaneously, building upon the semantic path analysis in § 3, the induced semantics transition from high-drift exploration to a stable neighborhood. We quantify this through the global convergence residual $\mathcal{D}_{\text{global}}(k)$, which stabilizes at a low plateau as thinking progresses. We operationalize the entry into this convergence region when $\mathcal{D}_{\text{global}}(k) \leq \epsilon_D$, where ϵ_D is set by the empirical tail fluctuations used to estimate Q_∞ .

Based on these metrics, we categorize the reasoning process into two distinct stages:

Pre-RCP: Active Reasoning Stage.

$$\mathcal{D}_{\text{global}}(k) > \epsilon_D \quad \vee \quad \Delta_{\text{content}}(k) > \epsilon_c. \quad (6)$$

The model is in a state of high-drift exploration; the semantic path continues to evolve, and the induced content remains volatile, frequently expanding to compensate for insufficient thinking. Computation in this stage is typically essential for answer maturation.

Post-RCP: Converged Reasoning Stage.

$$\mathcal{D}_{\text{global}}(k) \leq \epsilon_D \quad \wedge \quad \Delta_{\text{content}}(k) \leq \epsilon_c. \quad (7)$$

The semantic state stabilizes within a converged neighborhood, and the content length reaches a plateau. Additional compute beyond this point yields diminishing returns, often manifesting as redundant rationalization or "overthinking."

The critical boundary separating these two regimes is the **Reasoning Completion Point (RCP)**. Formally, we define the RCP as the earliest

truncation step k_{RCP} where both length stabilization and semantic convergence are simultaneously achieved:

$$k_{\text{RCP}} \triangleq \min \left\{ k : \begin{array}{l} \Delta_{\text{content}}(k) \leq \epsilon_c \\ \wedge \mathcal{D}_{\text{global}}(k) \leq \epsilon_D \end{array} \right\}, \quad (8)$$

The thresholds ϵ_c and ϵ_D are empirically determined by the tail fluctuations in the late-step window. Under this unified formalization, the RCP serves as an instance-specific "stopping criterion" that respects the intrinsic complexity of the prompt. Beyond k_{RCP} , additional thinking is largely redundant as it no longer materially alters the induced content's length or conceptual essence. Consequently, mitigating overthinking reduces to an online detection problem: identifying the transition k_{RCP} during decoding to terminate the thinking phase at the precise moment of reasoning maturity.

5 Online Overthinking Elimination

The pivot to eliminating overthinking lies in the precise detection of the RCP. The logical progression of this section follows a "distill-to-detect" pipeline: we first leverage our reasoning dynamic findings (§ 2 and § 3) to extract gold RCP labels from full reasoning trajectories; however, to eliminate overthinking in practice, these insights must be translated into an online detector capable of real-time monitoring during inference. Specifically, we use the extracted offline data to train a tree-based model that identifies the latent transition into convergence. The resulting rules enable rapid RCP detection during standard decoding, effectively truncating redundant computation without sacrificing accuracy.

5.1 Offline RCP Extraction

In the "distill" stage of our pipeline, we start from the formal RCP criterion in Eq. 8 and extract instance-level annotations from full thinking trajectories. For each instance, we generate a complete reasoning trajectory, perform controlled truncation over candidate sentence boundaries to identify the earliest boundary at which reasoning has effectively converged, denoted k_{RCP} . This procedure is inherently offline, as it depends on access to complete reasoning trajectories and auxiliary semantic probes that are unavailable during standard decoding. Accordingly, we treat the resulting k_{RCP} annotations as gold supervision.

A key empirical observation from these labeled traces is that k_{RCP} typically coincides with the *first emergence* of the final answer in the reasoning trajectories. While such answer emergence marks the onset of the Post-RCP convergence regime, it is not a reliable target for online detection. Answer surface forms vary widely across instances, including numbers, units, natural language, and multiple-choice letters, and may be rephrased, rendering string-based criteria brittle. To replace this unobservable event with an intrinsic proxy, we examine the model's next-token distributions around k_{RCP} and find a consistent transition in the rank R_k of the thinking-termination delimiter `</think>`. Specifically, R_k drops sharply as the answer first emerges, reflecting an increasing preference to terminate thinking. This delimiter-rank dynamics is internal to the model and largely invariant to answer format, making it the key distilled feature used for online RCP detection.

5.2 Online RCP Detection

Offline extraction reveals an online-accessible signature of convergence. When the final answer first emerges, the rank R_k of the thinking-termination token `</think>` drops sharply (Figure 5). This motivates a practical reformulation of online RCP detection: given the rank observations available up to the current sentence boundary, decide whether decoding has reached the Reasoning Completion Point and can stop.

We instantiate this idea as a lightweight Reasoning Completion Point Detector (RCPD). At each sentence boundary, RCPD queries the next-token distribution and extracts the current R_k together with a short history of preceding rank values. Using offline-extracted k_{RCP} as supervision on Math-500 (Lightman et al., 2023), we label boundaries and fit an interpretable decision tree whose splits capture the implicit stopping criterion in terms of rank dynamics. We then distill the tree into deterministic stopping rules (exact thresholds and window sizes in Appendix C).

During inference, we apply the same rules online; once triggered, we force `</think>` to terminate the thinking phase and immediately transition to answer generation. This completes the distill-to-detect pipeline and yields a practical stopping strategy for eliminating Post-RCP overthinking.

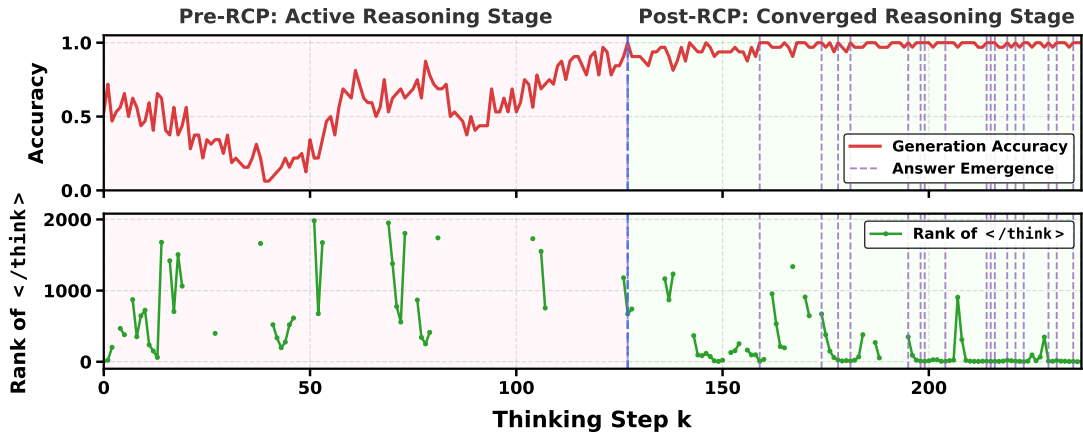


Figure 5: Top panel: Accuracy stabilizes around answer emergence. Bottom panel: The rank of $\langle /think \rangle$ (R_k) drops precipitously at answer emergence; this drop serves as a signature of convergence.

5.3 Experimental Evaluation

We evaluate on AIME24, AIME25 (MAA, 2025), and GPQA-D (Rein et al., 2023), using Qwen3 (8B/14B/30B-A3B) and DeepSeek-R1-Distill-8B. We report efficiency measured by the average number of generated tokens (Tok), Accuracy (Acc), and the relative performance-cost ratio (RPCR), which measures the accuracy gain over the No-Think baseline per generated token. For readability, we report RPCR with two decimal places. For ease of comparison, we show the Compression Rate (CR; relative to full decoding) as an integer percentage in parentheses next to Tok. Metric definitions and token accounting details are provided in Appendix B. For autoregressive generation, Tok directly measures generation length. We compare against Full Reasoning, Budget Force (BF), No-Think, and DEER.

As shown in Table 1, results align with *Reasoning Dynamics*, indicating that extending the thinking process into the Post-RCP convergence regime offers no accuracy gains. RCPD exploits this by substantially reducing token generation compared to full decoding while maintaining competitive performance across inference-time baselines. This demonstrates that full decoding often involves redundant computation well beyond the necessary RCP.

To ensure a rigorous evaluation, we compare with BF as a fixed thinking-token budget baseline. For each (model, benchmark), we retrospectively tune the token budget so that BF matches the average token usage of RCPD, yielding an oracle iso-compute comparison that is favorable to BF. BF nevertheless exhibits significant degradation,

underscoring the limitation of ignoring Reasoning Dynamics. By enforcing a static budget, BF often truncates the process during Pre-RCP, before the answer has converged. In contrast, RCPD adapts to instance-specific complexity, allowing reasoning to proceed until it enters Post-RCP convergence.

Finally, qualitative analyses further support the Post-RCP interpretation. Appendix F distinguishes two failure modes of early stopping: late detections are typically conservative and mainly leave compression gains unrealized, whereas rare premature exits can reduce accuracy by forcing an early structural transition and causing *alignment failure*. Appendix G further shows that once the model enters Post-RCP convergence, it often exhibits *cognitive inertia*: subsequent tokens repeatedly rationalize the first-emerged answer rather than correcting it, and may even drift into redundant loops. RCPD truncates this regime by design, explaining why early stopping can improve both efficiency and robustness.

6 Related Work

Based on the efficient reasoning taxonomy established in (Sui et al., 2025; Wang et al., 2025), we categorize related work into three classes, following and extending previous research: Post-training Based Methods, Prompt-based Methods, and Early Exit Methods.

Post-training Based Methods. Recent post-training studies show that training-time intervention can improve the quality and efficiency of reasoning trajectories. These methods leverage supervised fine-tuning with variable-length chain-of-thought (CoT) data (Yu et al., 2024; Kang

Method	AIME24			AIME25			GPQA-D			Average		
	Tok↓ (CR%↓)	Acc↑	RPCR↑	Tok↓ (CR%↓)	Acc↑	RPCR↑	Tok↓ (CR%↓)	Acc↑	RPCR↑	Tok↓ (CR%↓)	Acc↑	RPCR↑
Based on Qwen3-8B												
Full	15435 (100%)	72.22	33.84	17828 (100%)	63.33	23.68	9514 (100%)	60.10	23.09	14259 (100%)	65.22	27.22
No-Think	7271 (47%)	19.99	0.00	5036 (28%)	21.11	0.00	2687 (28%)	38.13	0.00	4998 (35%)	26.40	0.00
BF	10373 (67%)	58.88	37.49	11772 (66%)	55.56	29.26	3962 (42%)	55.56	43.99	8702 (61%)	56.66	34.77
DEER	13952 (90%)	72.22	37.44	16628 (93%)	67.78	28.07	9085 (95%)	59.60	23.63	13222 (93%)	66.53	30.35
RCPD (Ours)	9958 (65%)	72.22	52.45	10067 (56%)	63.33	41.94	4130 (43%)	64.65	64.21	8052 (56%)	66.73	50.09
Based on Qwen3-14B												
Full	13350 (100%)	75.56	35.79	16711 (100%)	70.00	27.26	7711 (100%)	64.31	13.81	12591 (100%)	69.96	27.54
No-Think	6294 (47%)	27.78	0.00	3533 (21%)	24.44	0.00	2733 (35%)	53.66	0.00	4186 (33%)	35.29	0.00
BF	8974 (67%)	54.44	29.71	10546 (63%)	56.67	30.56	3932 (51%)	62.29	21.95	7817 (62%)	57.80	28.80
DEER	12265 (92%)	72.22	36.23	14387 (86%)	70.00	31.67	7274 (94%)	64.48	14.87	11309 (90%)	68.90	29.72
RCPD (Ours)	8799 (66%)	73.33	51.77	9987 (60%)	68.89	44.51	3709 (48%)	67.17	36.42	7498 (60%)	69.80	46.03
Based on Qwen3-30B-A3B												
Full	13449 (100%)	82.22	41.30	16457 (100%)	74.44	31.06	7510 (100%)	66.50	25.17	12472 (100%)	74.38	33.56
No-Think	5667 (42%)	26.67	0.00	3995 (24%)	23.33	0.00	2637 (35%)	47.60	0.00	4099 (33%)	32.53	0.00
BF	11147 (83%)	74.44	42.85	12708 (77%)	65.56	33.23	4617 (61%)	64.81	37.28	9490 (76%)	68.27	37.66
DEER	12326 (92%)	80.00	43.27	15891 (97%)	67.78	27.97	7361 (98%)	67.85	27.51	11860 (95%)	71.87	33.17
RCPD (Ours)	11229 (83%)	82.22	49.47	12670 (77%)	74.44	40.34	4469 (60%)	68.69	47.19	9457 (76%)	75.11	45.02
Based on DeepSeek-R1-Distill-8B												
Full	21339 (100%)	76.67	26.56	23237 (100%)	70.00	21.04	11537 (100%)	58.73	17.86	18704 (100%)	68.47	22.49
BF	16202 (76%)	72.22	32.24	19740 (85%)	62.22	20.83	8909 (77%)	56.22	20.31	14950 (80%)	63.55	24.85
DEER	21105 (99%)	72.22	24.75	23129 (100%)	67.77	20.17	11207 (97%)	58.25	17.95	18480 (99%)	66.08	21.47
RCPD (Ours)	16981 (80%)	76.67	33.38	19602 (84%)	67.77	23.80	9814 (85%)	58.25	20.50	15465 (83%)	67.56	26.61

Table 1: Experimental results across various reasoning models. CR (relative to full decoding) is shown as an integer percentage in parentheses next to Tok, and we report RPCR (accuracy gain over the No-Think baseline per generated token) with two decimal places. DeepSeek-R1-Distill-8B lacks a No-Think mode; its RPCR is computed against the Qwen3-8B No-Think baseline.

et al., 2025; Xia et al., 2025; Ma et al., 2025), process supervision or step-level preference optimization (Chen et al., 2024a,b; Nie et al., 2026), integrate length rewards in reinforcement learning (Kimi Team et al., 2025; Luo et al., 2025; Aggarwal and Welleck, 2025), or employ latent space reasoning methods (Hao et al., 2024; Cheng and Durme, 2024; Shen et al., 2025). While existing approaches reduce reasoning tokens through various training strategies, they require extensive data and retraining. In contrast, our method requires no training.

Prompt-based Methods. Approaches in this category (Han et al., 2024; Muennighoff et al., 2025; Lee et al., 2025; OpenAI et al., 2025) use different prompts to enforce the model to generate concise CoT, thus reducing unnecessary reasoning steps (Kang et al., 2025; Peng et al., 2025; Huang et al., 2025). Although this approach has shown some success in generating more concise reasoning chains, it has significant limitations. Specifically, the quality and flexibility of the prompt design directly affect the model’s performance. In the face of diverse and complex reasoning tasks, static prompting methods often fail to adapt effectively to the

variability of inputs, leading to reasoning chains that lack sufficient depth and precision.

Early Exit Methods. These methods intervene in the reasoning process by prematurely terminating generation to mitigate redundancy. Existing approaches primarily fall into two categories. The first relies on external monitoring mechanisms or additional model training. For instance, methods utilizing auxiliary classifiers (Liu and Wang, 2025), bandit controllers (Sun et al., 2025), or difficulty estimators (Pu et al., 2025) require extra components, while **S-GRPO** (Dai et al., 2025) necessitates extra training of the LLM to learn early termination strategies. These approaches inevitably introduce additional training costs or inference latency. In contrast, a more recent line of work focuses on an overhead-free paradigm that leverages the model’s intrinsic states without extra components. **DEER** (Yang et al., 2025b) exemplifies this by enabling early exiting without computational overhead. Our study aligns with this efficient paradigm.

7 Conclusion

In this work, we reveal the Reasoning Dynamics of LLMs, characterized by Thinking-Content Compensation and Semantic Path Convergence. These dynamics shape the reasoning process into an active reasoning phase and a converged reasoning phase separated by the RCP. We demonstrate that a simple method can effectively detect this point to mitigate overthinking. Our findings identify post-RCP redundancy as the primary cause of overthinking, highlighting reliable self-termination as a cornerstone for efficient test-time scaling.

Acknowledgments

This work was supported in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDB0680302, by the National Natural Science Foundation of China under Grant No. 62276248, by the Key Research and Development Program of Xinjiang Uyghur Autonomous Region under Grant No. 2024B03026, by the Beijing Nova Program under Grant No. 20250484765, and by the Youth Innovation Promotion Association of the Chinese Academy of Sciences under Grant No. 2023111.

8 Limitations

Our analysis and evaluation focus on reasoning models and are conducted primarily on the Qwen3 and DeepSeek-R1 families, spanning multiple model sizes and benchmarks. We prioritize these open-weight systems because they expose explicit thinking delimiters and the decoding signals needed for controlled interventions and RCPD. While these models exhibit consistent Reasoning Dynamics, we did not run the full suite of RCP analyses across the broader ecosystem of models, such as closed-source APIs, tool-augmented agents and multimodal models. Establishing how universal RCP-like phase transitions are beyond this scope remains an important direction.

9 Potential Risks

The proposed RCPD method primarily focuses on enhancing the efficiency of reasoning processes in LLMs. As such, there are no inherent risks associated with its implementation or deployment. The method does not introduce any new computational mechanisms that could adversely affect the model’s operation. Additionally, RCPD aims to mitigate

overthinking in LLMs, which, in turn, reduces unnecessary resource consumption and computational overhead.

References

- Pranjal Aggarwal and Sean Welleck. 2025. [L1: controlling how long A reasoning model thinks with reinforcement learning](#). *CoRR*, abs/2503.04697.
- Steven Bird and Edward Loper. 2004. [NLTK: the natural language toolkit](#). In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, Spain, July 21-26, 2004 - Poster and Demonstration*. ACL.
- Guoxin Chen, Minpeng Liao, Chengxi Li, and Kai Fan. 2024a. [Alphamath almost zero: Process supervision without process](#). In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*.
- Guoxin Chen, Minpeng Liao, Chengxi Li, and Kai Fan. 2024b. [Step-level value preference optimization for mathematical reasoning](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7889–7903, Miami, Florida, USA. Association for Computational Linguistics.
- Jeffrey Cheng and Benjamin Van Durme. 2024. [Compressed chain of thought: Efficient reasoning through dense representations](#). *CoRR*, abs/2412.13171.
- Dongkyu Cho, Amy B. Z. Zhang, Bilel Fehri, Sheng Wang, Rumi Chunara, Rui Song, and Hengrui Cai. 2025. [Correct reasoning paths visit shared decision pivots](#). *CoRR*, abs/2509.21549.
- Muzhi Dai, Chenxu Yang, and Qingyi Si. 2025. [S-GRPO: early exit via reinforcement learning in reasoning models](#). *CoRR*, abs/2505.07686.
- DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, and 1 others. 2025. [Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning](#). *CoRR*, abs/2501.12948.
- Soumya Suvra Ghosal, Souradip Chakraborty, Avinash Reddy, Yifu Lu, Mengdi Wang, Dinesh Manocha, Furong Huang, Mohammad Ghavamzadeh, and Amrit Singh Bedi. 2025. [Does thinking more always help? mirage of test-time scaling in reasoning models](#). *Preprint*, arXiv:2506.04210.
- Google DeepMind. 2025. [Gemini 3 pro model card](#). Model card, Google DeepMind.
- Jinyi Han, Ying Huang, Ying Liao, Zishang Jiang, Xikun Lu, Haiquan Zhao, Xinyi Wang, Guanghao Zhou, Sihang Jiang, Jiaqing Liang, Weikang Zhou, Zeye Sun, Fei Yu, and Yanghua Xiao. 2025. [Your models have thought enough: Training large reasoning models to stop overthinking](#). *CoRR*, abs/2509.23392.

- Tingxu Han, Zhenting Wang, Chunrong Fang, Shiyu Zhao, Shiqing Ma, and Zhenyu Chen. 2024. [Token-budget-aware LLM reasoning](#). *CoRR*, abs/2412.18547.
- Shibo Hao, Sainbayar Sukhbaatar, DiJia Su, Xian Li, Zhiting Hu, Jason Weston, and Yuandong Tian. 2024. [Training large language models to reason in a continuous latent space](#). *CoRR*, abs/2412.06769.
- Jiameng Huang, Baijiong Lin, Guhao Feng, Jierun Chen, Di He, and Lu Hou. 2025. [Efficient reasoning for large reasoning language models via certainty-guided reflection suppression](#). *CoRR*, abs/2508.05337.
- Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, and 1 others. 2024. [Openai o1 system card](#). *CoRR*, abs/2412.16720.
- Yu Kang, Xianghui Sun, Liangyu Chen, and Wei Zou. 2025. [C3ot: Generating shorter chain-of-thought without compromising effectiveness](#). In *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, pages 24312–24320. AAAI Press.
- Kimi Team, Angang Du, Bofei Gao, Bowei Xing, and 1 others. 2025. [Kimi k1.5: Scaling reinforcement learning with llms](#). *CoRR*, abs/2501.12599.
- Ayeong Lee, Ethan Che, and Tianyi Peng. 2025. [How well do llms compress their own chain-of-thought? A token complexity approach](#). *CoRR*, abs/2503.01141.
- Melody Zixuan Li, Kumar Krishna Agrawal, Arna Ghosh, Komal Kumar Teru, Adam Santoro, Guillaume Lajoie, and Blake A. Richards. 2025. [Tracing the representation geometry of language models from pretraining to post-training](#). *CoRR*, abs/2509.23024.
- Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. 2023. [Let’s verify step by step](#). *arXiv preprint arXiv:2305.20050*.
- Xin Liu and Lu Wang. 2025. [Answer convergence as a signal for early stopping in reasoning](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 17907–17918, Suzhou, China. Association for Computational Linguistics.
- Haotian Luo, Li Shen, Haiying He, Yibo Wang, Shiwei Liu, Wei Li, Naiqiang Tan, Xiaochun Cao, and Dacheng Tao. 2025. [O1-pruner: Length-harmonizing fine-tuning for o1-like reasoning pruning](#). *CoRR*, abs/2501.12570.
- Xinyin Ma, Guangnian Wan, Runpeng Yu, Gongfan Fang, and Xinchao Wang. 2025. [Cot-valve: Length-compressible chain-of-thought tuning](#). *CoRR*, abs/2502.09601.
- MAA. 2025. [Aime problems and solutions](#).
- Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel J. Candès, and Tatsunori Hashimoto. 2025. [s1: Simple test-time scaling](#). *CoRR*, abs/2501.19393.
- Shuaiyi Nie, Siyu Ding, Wenyuan Zhang, Linhao Yu, Tianmeng Yang, Yao Chen, Tingwen Liu, Weichong Yin, Yu Sun, and Hua Wu. 2026. [ATTNPO: attention-guided process supervision for efficient reasoning](#). *CoRR*, abs/2602.09953.
- OpenAI, Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, and 1 others. 2025. [gpt-oss-120b & gpt-oss-20b model card](#). *Preprint*, arXiv:2508.10925.
- Keqin Peng, Liang Ding, Yuanxin Ouyang, Meng Fang, and Dacheng Tao. 2025. [Revisiting overthinking in long chain-of-thought from the perspective of self-doubt](#). *CoRR*, abs/2505.23480.
- Xiao Pu, Michael Saxon, Wenyue Hua, and William Yang Wang. 2025. [THOUGHTTERMINATOR: benchmarking, calibrating, and mitigating overthinking in reasoning models](#). *CoRR*, abs/2504.13367.
- David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R. Bowman. 2023. [Gpqa: A graduate-level google-proof q&a benchmark](#). *Preprint*, arXiv:2311.12022.
- Zhenyi Shen, Hanqi Yan, Linhai Zhang, Zhanghao Hu, Yali Du, and Yulan He. 2025. [CODI: compressing chain-of-thought into continuous space via self-distillation](#). *CoRR*, abs/2502.21074.
- Jinyan Su, Jennifer Healey, Preslav Nakov, and Claire Cardie. 2025. [Between underthinking and overthinking: An empirical study of reasoning length and correctness in llms](#). *CoRR*, abs/2505.00127.
- Yang Sui, Yu-Neng Chuang, Guanchu Wang, Jiamu Zhang, Tianyi Zhang, Jiayi Yuan, Hongyi Liu, Andrew Wen, Shaochen Zhong, Hanjie Chen, and Xia Ben Hu. 2025. [Stop overthinking: A survey on efficient reasoning for large language models](#). *CoRR*, abs/2503.16419.
- Renliang Sun, Wei Cheng, Dawei Li, Haifeng Chen, and Wei Wang. 2025. [Stop when enough: Adaptive early-stopping for chain-of-thought reasoning](#). *CoRR*, abs/2510.10103.
- Meituan LongCat Team, Anchun Gui, Bei Li, Bingyang Tao, Bole Zhou, Borun Chen, Chao Zhang, Chengcheng Han, Chenhui Yang, Chi Zhang, Chong Peng, Chuyu Zhang, Cong Chen, Fengcun Li, Gang Xu, Guoyuan Lin, Hao Jiang, Hao Liang, Haomin Fu, and 80 others. 2025. [Longcat-flash-thinking technical report](#). *CoRR*, abs/2509.18883.

Rui Wang, Hongru Wang, Boyang Xue, Jianhui Pang, Shudong Liu, Yi Chen, Jiahao Qiu, Derek Fai Wong, Heng Ji, and Kam-Fai Wong. 2025. [Harnessing the reasoning economy: A survey of efficient reasoning for large language models](#). *CoRR*, abs/2503.24377.

Yuyang Wu, Yifei Wang, Tianqi Du, Stefanie Jegelka, and Yisen Wang. 2025a. [When more is less: Understanding chain-of-thought length in llms](#). *CoRR*, abs/2502.07266.

Yuyang Wu, Yifei Wang, Ziyu Ye, Tianqi Du, Stefanie Jegelka, and Yisen Wang. 2025b. [When more is less: Understanding chain-of-thought length in llms](#). *Preprint*, arXiv:2502.07266.

Heming Xia, Yongqi Li, Chak Tou Leong, Wenjie Wang, and Wenjie Li. 2025. [Tokenskip: Controllable chain-of-thought compression in llms](#). *CoRR*, abs/2502.12067.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, and 1 others. 2025a. [Qwen3 technical report](#). *Preprint*, arXiv:2505.09388.

Chenxu Yang, Qingyi Si, Yongjie Duan, Zheliang Zhu, Chenyu Zhu, Zheng Lin, Li Cao, and Weiping Wang. 2025b. [Dynamic early exit in reasoning models](#). *CoRR*, abs/2504.15895.

Ping Yu, Jing Xu, Jason Weston, and Ilya Kulikov. 2024. [Distilling system 2 into system 1](#). *CoRR*, abs/2407.06023.

Wenyuan Zhang, Shuaiyi Nie, Xinghua Zhang, Zefeng Zhang, and Tingwen Liu. 2025a. [S1-bench: A simple benchmark for evaluating system 1 thinking capability of large reasoning models](#). *CoRR*, abs/2504.10368.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, Fei Huang, and Jingren Zhou. 2025b. [Qwen3 embedding: Advancing text embedding and reranking through foundation models](#). *CoRR*, abs/2506.05176.

A Additional Experiment

Due to space constraints, we report the Qwen3-32B results in Table 4 in the Appendix.

B Additional Evaluation Details

This section complements the main experimental description by specifying the evaluation protocol and baseline implementations used for early-exit comparisons. Specifically, our evaluation utilizes AIME24 and AIME25, which each contain 30 problems, and GPQA-D, which comprises 198 problems.

Decoding Configuration and Token Accounting.

We use each model’s official chat template and its delimiter-based reasoning format (e.g., `<think>...</think>`). Unless otherwise stated, we follow the recommended nucleus-sampling settings (Temperature = 0.6, Top- p = 0.95) and set the maximum generation length to 32,768 tokens so that behaviors are not driven by hard token caps.

We report **Tok** as the average number of generated tokens per instance (including both thinking and content). To measure the efficiency trade-off, we report the relative performance–cost ratio (**RPCR**), defined as the accuracy gain over the No-Think baseline per generated token:

$$\text{RPCR} = \frac{\text{Acc} - \text{Acc}_{\text{No-Think}}}{\text{Tok}} \times 10^4$$

For readability, we report **RPCR** with two decimal places. Additionally, the Compression Rate (**CR**), defined as the token ratio relative to full decoding (i.e., $\text{Tok}/\text{Tok}_{\text{Full}}$), is shown as an integer percentage in parentheses next to **Tok**.

Baselines. We compare RCPD against representative inference-time strategies:

- **Full:** standard decoding without early stopping.
- **No-Think:** prompting the model to skip explicit reasoning and answer directly, corresponding to a System-1-style response setting (Zhang et al., 2025a).
- **BF (Budget Force):** static truncation that forces `</think>` after a fixed thinking-token budget.
- **DEER:** a trigger-based early-exit method that terminates upon reflective cue words and a confidence threshold on the final answer format.

Iso-Compute Calibration for BF. To make BF a strong baseline, we sweep fixed thinking-token budgets for each (model, benchmark) pair and retrospectively select the budget whose average generated tokens are closest to RCPD. This gives BF oracle access to a globally tuned budget chosen under the same compute target, making the comparison iso-compute and favorable to BF.

Impact of Temporal Consistency on Stopping Robustness. Table 2 compares RCPD with a simplified baseline that triggers stopping as soon as the `</think>` token enters the top-5 of the next-token distribution at a sentence boundary (denoted as `</think>-5`). While this single-threshold rule confirms that delimiter rank is a meaningful convergence proxy, RCPD performs better by incorporating short-horizon temporal patterns (Appendix C) that reduce sensitivity to noise.

C Implementation Details of RCPD Stopping Criteria

As discussed in Section §3.3, to balance computational efficiency with detection accuracy, we distill the decision boundaries mined from the probe dataset into a set of deterministic, symbolic rules. The generation process terminates immediately if the rank trajectory of the `</think>` token, denoted as R_t , satisfies any of the following logical conditions.

To facilitate reproducibility and future research, our code is publicly available at <https://github.com/Hi-archers/reasoning-dynamics>.

1. Immediate Saturation. This condition captures scenarios where the model exhibits overwhelming certainty. A rank of $R_t \leq 5$ indicates that the reasoning process has effectively collapsed into a definitive conclusion, rendering further generation redundant.

$$\mathcal{C}_{\text{sat}} : R_t \leq 5 \quad (9)$$

2. Progressive Convergence. This rule corresponds to the “ladder pattern” observed in successful deductions. It requires a monotonic ascent in the token’s rank probability (i.e., descending rank values) over a fixed window, representing a directed convergence toward the solution.

$$\mathcal{C}_{\text{conv}} : \bigwedge_{k=0}^3 (R_{t-k} \leq \theta_k), \quad (10)$$

where $\theta = [10, 50, 100, 1000]$

3. Sustained Plateau. To distinguish between true convergence and stochastic noise, this condition detects whether the model has reached an “entropic floor.” Even if the rank does not hit the absolute peak ($R_t \approx 1$), a sustained presence within a low-rank region signals that no new information is being generated. We define this via two sub-conditions:

- **Short-term Stability (Noise Filtering):** Ensures high confidence is maintained over a narrow window.

$$\mathcal{C}_{\text{short}} : \forall k \in \{0, 1, 2\}, R_{t-k} \leq 20 \quad (11)$$

- **Long-term Persistence (Sampling Inertia):** Detects a prolonged plateau where the model effectively loops or stalls in a high-probability region.

$$\mathcal{C}_{\text{long}} : \forall k \in \{0, \dots, 5\}, R_{t-k} \leq 50 \quad (12)$$

The final stop signal S_t is triggered by the disjunction of these conditions:

$$S_t = \mathcal{C}_{\text{sat}} \vee \mathcal{C}_{\text{conv}} \vee (\mathcal{C}_{\text{short}} \vee \mathcal{C}_{\text{long}}) \quad (13)$$

D Quantifying Semantic Convergence

We quantify the stability of the semantic trajectory after the RCP by fitting a 95% confidence ellipse in the 2D PCA space (Figure 3). Let \hat{k} denote the RCP step, and let $\tilde{\mathbf{v}}_k^{(i)}$ denote the PCA-projected embedding of the i -th sampled continuation at step k . We collect all post-RCP points into $\mathcal{C} = \{\tilde{\mathbf{v}}_k^{(i)} : k \geq \hat{k}\}$ and estimate their sample mean $\boldsymbol{\mu}$ and covariance Σ .

For any point \mathbf{x} in this PCA plane, we measure its semantic deviation from the converged region using the squared Mahalanobis distance:

$$D_M(\mathbf{x})^2 = (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad (14)$$

We then define the 95% confidence ellipse (semantic basin) as the level set:

$$\mathcal{B}_{0.95} = \{\mathbf{x} \mid D_M(\mathbf{x})^2 \leq \chi_2^2(0.95)\}, \quad (15)$$

where $\chi_2^2(0.95) \approx 5.991$ is the 95th percentile of the χ^2 distribution with 2 degrees of freedom. Under a locally elliptical (approximately Gaussian) stationary regime in the PCA plane, $D_M(\mathbf{x})^2$ follows χ_2^2 , so $\mathcal{B}_{0.95}$ contains 95% probability mass of the converged distribution.

Model	Method	Acc
Qwen3-8B	</think>-5	61.11
Qwen3-8B	RCPD	75.11
Qwen3-14B	</think>-5	65.72
Qwen3-14B	RCPD	69.79
Qwen3-30B-A3B	</think>-5	73.84
Qwen3-30B-A3B	RCPD	75.11
Qwen3-32B	</think>-5	73.26
Qwen3-32B	RCPD	73.37

Table 2: Static stopping by a single rank threshold (</think>-5) is competitive but consistently worse than RCPD, indicating that simple temporal patterns improve robustness.

Finally, we compute the fraction of post-RCP points that remain within this basin. In Figure 3, 97.3% of subsequent trajectory points lie inside $\mathcal{B}_{0.95}$, indicating that after the RCP the semantic path no longer drifts but fluctuates within a compact, steady neighborhood.

E Additional Visualizations

To demonstrate that the same phenomena consistently appear across datasets, we provide additional examples that could not fit in the main text.

Using the same 95% confidence ellipse protocol in Section D, we quantify post-RCP semantic stability for these additional cases. For AIME24 Question 1 (Figure 7a), 94.3% of Post-RCP points lie inside the 95% basin, and for AIME24 Question 30 (Figure 7b), the ratio is 96.5%. The consistently high in-basin fractions support that semantic convergence after the RCP is not instance-specific but recurs across different problems.

F Failure Modes of Early Stopping

Early-stopping errors mainly fall into two categories. Late stopping is typically conservative: on AIME24 with Qwen3-32B, the ideal RCP occurs at about 63% of the full generation trajectory, whereas RCPD triggers at about 84% on average. Its main cost is therefore unrealized compression rather than a clear loss in accuracy. By contrast, premature stopping can occasionally hurt accuracy when forcing </think> transitions the model into answer generation before the latent reasoning state is fully stabilized. We refer to this failure mode as *alignment failure*.

G Examples of Overthinking

This section describes three typical cases of overthinking phenomena in model reasoning. The first case involves repeated verification of a correct answer, while the other two cases involve repeated derivation of speculative incorrect answers. These examples illustrate how overthinking can lead to unnecessary computational waste and inefficient problem-solving processes.

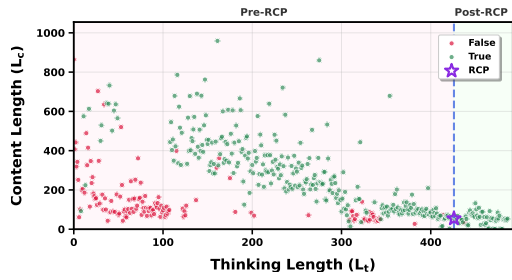
To further assess the effectiveness of the model’s post-RCP validation mechanism, we conduct a counterfactual modification experiment. We manually edit the initial tentative answer in the reasoning trace of Table 3. Concretely, we replace “...perhaps the number of intersections is 12...” with “...perhaps the number of intersections is 16...”. After injecting this incorrect intermediate conclusion, we run 64 independent reasoning trials. As shown in Table 3, the model exhibits a pronounced **confirmation bias**. Only 11 trials recover the correct answer (12), whereas 49 out of 64 converge to the injected value (16). We observe the same pattern under the symmetric setting that flips the injection from 16 to 12, where the model again predominantly adopts the newly specified answer. Overall, these results indicate that Post-RCP generation seldom performs genuine re-verification and instead tends to rationalize earlier tokens.

H Declaration of AI Use

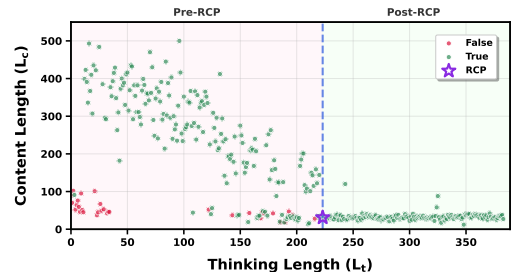
In accordance with the ACL Policy on AI Writing Assistance, we openly disclose the use of AI tools in the preparation of this manuscript.

Writing Assistance We utilized Gemini 3-Pro solely as a language editing tool. Its application was strictly limited to paraphrasing, checking grammar, and polishing the original text written by the authors to enhance readability and clarity. The AI tool was not used to generate new scientific concepts, formulate hypotheses, or interpret the results. The authors have reviewed all AI-assisted text and retain full responsibility for the accuracy, originality, and integrity of the content.

Experimental Models It is important to distinguish the writing assistant from the models studied in this work. The methodologies, experiments, and results presented in this paper are based on the **Qwen3** and **DeepSeek-R1** large language models. These models served exclusively as the subjects

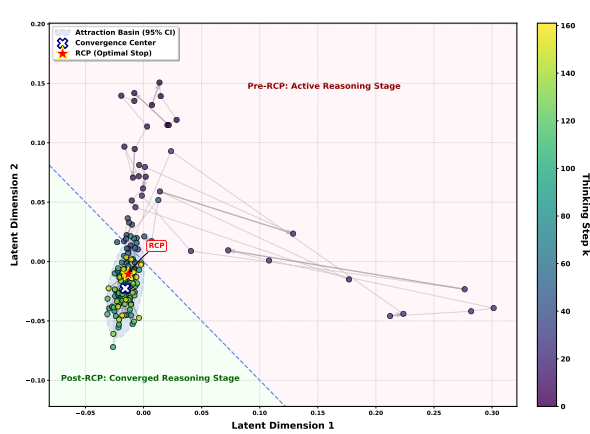


(a) AIME24 Question 2 Length Dynamics

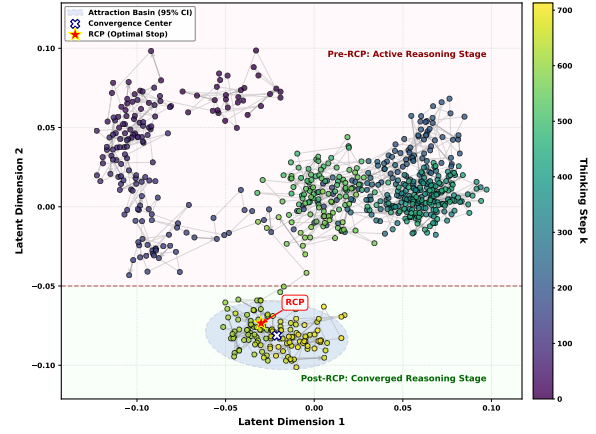


(b) AIME24 Question 9 Length Dynamics

Figure 6: Supplementary examples illustrating the transition from the Pre-RCP Active Reasoning Stage to the Post-RCP Converged Reasoning Stage (complementing Figure 2).



(a) AIME24 Question 1 Semantic Trajectory



(b) AIME24 Question 30 Semantic Trajectory

Figure 7: Additional semantic trajectory examples (supplementary to Figure 3).

of our research and the foundation for our experimental framework; they were not involved in the writing or drafting of this manuscript.

Experimental Setting	Final Answer Distribution (out of 64)		
	Recovered Original	Adopted Injection	Other
Injection: 16 (Original: 12)	11	49	4
Injection: 12 (Original: 16)	11	51	2

Table 3: Impact of Modified Intermediate Answer on Final Reasoning Outcome. The results demonstrate a strong confirmation bias: in both settings, the model overwhelmingly adopts the injected intermediate answer rather than recovering the correct original answer.

Method	AIME24			AIME25			GPQA-D			Average		
	Tok↓ (CR%↓)	Acc↑	RPCR↑	Tok↓ (CR%↓)	Acc↑	RPCR↑	Tok↓ (CR%↓)	Acc↑	RPCR↑	Tok↓ (CR%↓)	Acc↑	RPCR↑
Based on Qwen3-32B												
Full	11955 (100%)	82.22	40.90	16878 (100%)	65.56	28.31	7357 (100%)	69.53	25.35	12063 (100%)	72.44	31.87
No-Think	5731 (48%)	33.33	0.00	5472 (32%)	17.78	0.00	2277 (31%)	50.88	0.00	4493 (37%)	33.99	0.00
BF	10071 (84%)	78.89	45.24	11772 (70%)	55.56	32.09	4137 (56%)	64.98	34.08	8660 (72%)	66.48	37.52
DEER	12002 (100%)	81.11	39.81	14905 (88%)	66.67	32.80	6841 (93%)	69.70	27.51	11249 (93%)	72.49	34.23
RCPD	10062 (84%)	82.22	48.59	10917 (65%)	66.67	44.78	4094 (56%)	71.21	49.66	8358 (69%)	73.37	47.12

Table 4: Additional results on Qwen3-32B, reporting Tok (CR%), Acc, and $10^4 \times \text{RPCR}$ (two decimals).

Representative premature-stop case (complex algebra problem).
Thinking state at stop: "... Therefore, the product $P = 8321$. The remainder when divided by 1000 is $8321 - 8000 = 321$. Hence, the answer is 321."
Forced content after early stop: "... Using the binomial theorem or a known identity" followed by a trivial sign error, which derails the remaining computation and leads to the final answer 65.
Observation: The reasoning prefix has already reached the correct conclusion, but forcing the structural transition from thinking to answer generation breaks the model's cognitive inertia and corrupts the final surface realization. The resulting output is incorrect despite an already-mature internal solution state, illustrating an *alignment failure*. For brevity, we show only the decisive fragments of the trace.

Table 5: A representative premature-stop failure. When the transition to answer generation is forced too early, the model may fail to faithfully externalize an already-mature reasoning state, causing an alignment failure.

User: Please reason step by step, and put your final answer within boxed. Find the number of rectangles that can be formed inside a fixed regular dodecagon (12-gon) where each side of the rectangle lies on either a side or a diagonal of the dodecagon. The diagram below shows three of those rectangles.
Assistant: ... Therefore, returning to the dodecagon, with 6 diameters, the number of rectangles would be $C(6,2)=15$. But the problem says that the diagram shows three of those rectangles, and asks for the total number. Maybe 15 is the answer?
... Therefore, our formula seems valid. Therefore, returning to the dodecagon. If there are 6 diameters, then number of rectangles is $C(6,2) = 15$. But the problem says "three of those rectangles" are shown in the diagram. Maybe the diagram is just an example, and the answer is indeed 15?
... For example, choosing 0-6 and 1-7 gives a rectangle with vertices 0,1,6,7. Choosing 0-6 and 2-8 gives rectangle 0,2,6,8. Etc. Each of these should be unique. Therefore, total number is 15.
... Therefore, the rectangle must be inscribed in the same circle. Therefore, must have diagonals as diameters. Therefore, 15.
...
Therefore, I think the answer is 15.
****Final Answer****
15

Table 6: As shown in the table, the answer 15 was repeatedly derived through a speculative combinatorial formula during the problem-solving process. The assistant multiple times applied the same formula $\binom{6}{2} = 15$ based on the assumption of 6 diameters, without rigorous validation or alternative approaches. For conciseness, only representative instances of this repetitive reasoning are included. This circular derivation process illustrates a clear case of overthinking, resulting in unnecessary computational effort.

User: Please reason step by step, and put your final answer within \square . Every morning Aya goes for a 9-kilometer-long walk and stops at a coffee shop afterwards. When she walks at a constant speed of s kilometers per hour, the walk takes her 4 hours, including t minutes spent in the coffee shop. When she walks $s + 2$ kilometers per hour, the walk takes her 2 hours and 24 minutes, including t minutes spent in the coffee shop. Suppose Aya walks at $s + \frac{1}{2}$ kilometers per hour. Find the number of minutes the walk takes her, including the t minutes spent in the coffee shop.

Assistant: ... The walking time is 3 hours, coffee shop time is 24 minutes. So total time is 3 hours + 24 minutes. Let me convert 3 hours to minutes: 180 minutes + 24 minutes = 204 minutes.
... So 204 minutes.

Wait, but let me check again. Wait, 3 hours is 180 minutes. 180 + 24 = 204.
So 204 minutes. Is that the answer?

...

Then adding 24 minutes gives 204 minutes. Therefore, the answer is 204.
Alternatively, convert 24 minutes to 0.4 hours, total time 3.4 hours, which is 3 hours 24 minutes.
Either way, the problem asks for the number of minutes. So 204 minutes.

****Final Answer****
204

Table 7: As shown in the table, the answer 204 was repeatedly verified multiple times during the problem-solving process. The reasoning process involved 20 instances of repeated verification, and for the sake of brevity, only the most representative instances are included in the table. This repetitive verification process led to a significant waste of computational resources, which is a typical case of overthinking.

User: Please reason step by step, and put your final answer within \square . Define $f(x) = ||x| - \frac{1}{2}|$ and $g(x) = ||x| - \frac{1}{4}|$. Find the number of intersections of the graphs of $[y = 4g(f(\sin(2\pi x)))]$ and $x = 4g(f(\cos(3\pi y)))$.

Assistant: ... This is a complex equation, but perhaps we can analyze it by considering the possible values. Let me note that both x and y are in $[0,1]$, and the functions are continuous. Therefore, the number of solutions is finite, and we can try to find them by considering the number of times the curves cross. Alternatively, since both functions are symmetric and periodic, perhaps the number of intersections is 12, or some other number.

... let me consider the following: each time the function $F(x)$ completes an oscillation, it crosses the function $G(y)$ multiple times. Since $F(x)$ has two oscillations and $G(y)$ has three, the total number of intersections could be $2 * 3 * 2 = 12$, but this is just a guess.

... Given the complexity, perhaps the number of intersections is 12, as each oscillation of $F(x)$ (2 oscillations) intersects each oscillation of $G(y)$ (3 oscillations) twice, giving $2 * 3 * 2 = 12$.

... Given the complexity, it's hard to say, but if each oscillation intersects twice, then $2 * 3 * 2 = 12$ might be the answer.

...

Given that, and considering that (1,1) is one of them, I'll go with 12 as the answer.

****Final Answer****
12

Table 8: As shown in the table, the answer 12 was repeatedly speculated multiple times during the problem-solving process. The reasoning involved several instances of similar speculation based on oscillatory patterns and multiplicative guesses, and for brevity, only representative excerpts are included. This repetitive speculation process led to a waste of computational resources, exemplifying overthinking.