Think Just Enough: Leveraging Self-Assessed Confidence for Adaptive Reasoning in Language Models

Junyeob Kim[1], **Sang-goo Lee**[1 2], **Taeuk Kim**[3 *]
[1]Seoul National University, [2]IntelliSys, Korea, [3]Hanyang University
{juny116, sglee}@europa.snu.ac.kr
kimtaeuk@hanyang.ac.kr

Abstract

Recent reinforcement learning (RL)-trained language models have demonstrated strong performance on complex reasoning tasks by producing long and detailed reasoning traces. However, despite these advancements, they often struggle with finding the right balance in reasoning length: some terminate prematurely before reaching a correct answer (underthinking), while others continue reasoning beyond necessity, leading to inefficiency or even degraded accuracy (overthinking). To address these challenges, we propose a method for optimizing reasoning length via self-assessed confidence. By prompting the model to evaluate its own confidence at intermediate reasoning steps, we enable dynamic stopping once sufficient reasoning is achieved. Experiments across multiple reasoning benchmarks show that our approach improves computational efficiency without compromising answer quality. Furthermore, we find that confidence estimates from RL-trained reasoning models are more reliable than those from standard LLMs, making it a valuable internal signal for controlling reasoning depth.

1 Introduction

Recent advances in large reasoning models (LRMs)—a class of large language models (LLMs) explicitly optimized for multi-step reasoning—have shown remarkable capabilities in complex tasks such as mathematical problem solving and code generation (Xu et al., 2025; Chen et al., 2025; Li et al., 2025). Notably, models trained with reinforcement learning (RL), such as OpenAI's O-series (Jaech et al., 2024) and DeepSeek-R1 (Guo et al., 2025), exemplify this trend by producing extended chains of thought (CoT) (Wei et al., 2022) that involve reflection, verification, and even backtracking. These long-form reasoning traces are widely considered essential for enabling System-2-style cognition in LLMs (Li et al., 2025). However, emerging research reveals that reasoning length does not always positively correlated with accuracy. Instead, models often suffer from underthinking, where they prematurely truncate reasoning, or overthinking, where they continue unnecessarily past the point of correct conclusions, resulting in wasted computation or even degraded answers (Wang et al., 2025; Su et al., 2025; Yang et al., 2025b).

To address these issues, various approaches have been proposed. Learning-based methods include additional preference optimization to encourage shorter answers (Su et al., 2025; Yang et al., 2025b), as well as frameworks that enable switching between fast and slow thinking (Zhang et al., 2025a,b). In addition, inference time strategies include enforcing fast thinking (Ma et al., 2025), reducing thought transitions to prevent underthinking (Wang et al., 2025), implementing early exit mechanisms based on answer consistency (Fu et al., 2025) or internal signals (Yang et al., 2025a; Yong et al., 2025), and selecting shorter responses from a set of batch-sampled candidates (Hassid et al., 2025). While these methods help mitigate the issue that the extent of reasoning in LRMs is often suboptimal, they often rely on internal signals that are difficult to access in closed-source models or require memory-intensive techniques such as batch decoding.

To address the limitations of prior approaches, we propose a self-assessment–based method for adaptively regulating the reasoning process. This framework is motivated by recent findings that suggest that LRMs can often estimate their own confidence with reasonable reliability (Yoon et al., 2025). Unlike previous methods that depend on internal model instrumentation or external sampling, our approach directly prompts the model to express its confidence at intermediate steps during reasoning, providing an accessible and inter-

*Corresponding author.

Findings of the Association for Computational Linguistics: EACL 2026, pages 5000–5006

pretable signal for deciding whether to continue or terminate reasoning. While conceptually related to confidence-based methods such as DEER—which estimate confidence from internal logit-derived signals after producing a provisional answer—our method instead sources the confidence signal from the model's self-generated assessment within its ongoing chain of thought. This design choice allows the method to operate without access to internal activations and offers a modest interpretability advantage, although both approaches share the broader goal of using confidence to adaptively regulate reasoning. If the model indicates a sufficiently high level of certainty, reasoning is stopped early; otherwise, it proceeds until confidence improves or a threshold is reached. Through this framework, we investigate whether explicit self-assessment alone is sufficient for effective reasoning length control.

We validate our method across multiple mathematical reasoning benchmarks and open-source RL-trained LRMs. We additionally include small-scale preliminary analyses in non-mathematical settings. Empirical results demonstrate that self-assessed confidence enables substantial reductions reasoning generation length without compromising answer accuracy. Our findings further support recent claims that reasoning-trained LRMs are capable of introspectively evaluating their own confidence. Through additional calibration analyses, we observe a correlation between self-assessed confidence and logit-based internal confidence signals. This alignment suggests that reasoning-oriented models not only assess their own certainty, but also encode it coherently within their internal activations. Such correlation between generated and internal confidence signals stands in contrast to standard instruction-tuned LLMs, highlighting a distinctive self-monitoring capability that emerges through reinforcement learning-based reasoning optimization.

2 Methodology

2.1 Preliminaries

Unlike conventional instruction-tuned LLMs that produce answers in a single pass, LRMs explicitly separate the reasoning phase from the answer generation phase. This is typically achieved by enclosing the intermediate thought process within special tokens, commonly <think> ... </think>, which prompt the model to perform "slow thinking" in a deliberate, step-by-step manner. After completing this reflective phase, the model proceeds to generate the final response based on its internal reasoning trace.

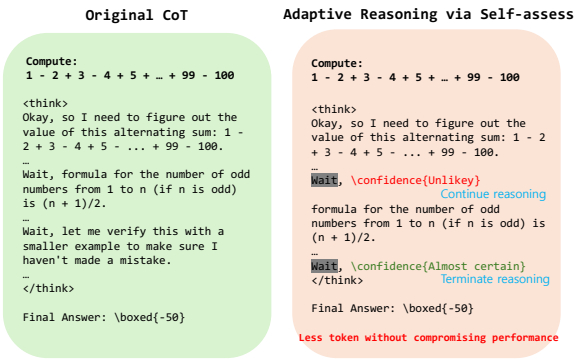

Figure 1: Overview of Self-Assessed Confidence–Based Adaptive Reasoning

```
During the thinking process, you must periodically assess the confidence of your
current reasoning. This assessment should not only reflect how confident you are in the
current reasoning steps, but also whether you have sufficiently explored the problem,
considered possible errors, and ensured there are no flaws or gaps in the logic. Assess
whether you have thought deeply enough to be confident in your final answer.

Use one of the following confidence levels:
- "Almost no chance" (0.0–0.1)
- "Highly unlikely" (0.1–0.2)
- "Chances are slight" (0.2–0.3)
- "Unlikely" (0.3–0.4)
- "Less than even" (0.4–0.5)
- "Better than even" (0.5–0.6)
- "Likely" (0.6–0.7)
- "Very good chance" (0.7–0.8)
- "Highly likely" (0.8–0.9)
- "Almost certain" (0.9–1.0)
Each category reflects the overall confidence that your answer is correct, **only if**
you have performed sufficient reasoning to justify your final answer. High confidence
should only be reported when you have thought through the problem thoroughly and
believe your reasoning supports the conclusion beyond reasonable doubt
Always denote this using \confidence{X}, where X is the confidence class name (only the
name, without the probability range). For example, if you are 95% sure of your
reasoning, you would write \confidence{Almost certain}. If you are 12% sure, you would
write \confidence{Highly unlikely}
```

Figure 2: System prompt template for self-assessed confidence estimation.

During the reasoning process, LRMs often exhibit behaviors such as self-checking, error correction, and alternative path exploration following tokens such as "Wait" and "Alternatively." In RL-trained LRMs, these reflective markers appear frequently and consistently, while in general-purpose LLMs they can be sparse or entirely absent—a distinction that later affects how reliably keyword-triggered confidence estimation can operate.

2.2 Self-Assessed Confidence Estimation

To dynamically control reasoning length, we implement a self-assessment mechanism that prompts the model to evaluate its own confidence at intermediate stages of reasoning. Rather than relying on internal metrics, the model explicitly outputs a confidence label on a 10-class scale ranging from *Almost no chance* (0.0–0.1) to *Almost certain* (0.9–1.0), provided through a system prompt to ensure consistent interpretation.

As shown in Figure 1, when reflective markers

such as “Wait” or the closing tag </think> appear, we treat these events as decision points: generation is paused and an additional prompt is injected to request the model’s confidence based on its reasoning so far. At this point, the token `confidence{` is included to enforce a structured confidence response and prevent further reasoning during estimation.

Although </think> typically marks the end of the reasoning phase, we perform one final confidence check at this point. If the estimated confidence is below the threshold, the token is replaced with a continuation cue (implemented as “Wait”), allowing a brief extension of reasoning. This extension is infrequent in practice but helps prevent premature termination.

When the model’s confidence reaches the threshold (set to *Almost certain*), we insert </think> to terminate reasoning and proceed to the final answer. Since this threshold directly affects termination behavior, we employ finer bins in the high-confidence region where thresholding is applied. The full system prompt is shown in Figure 2.

3 Experiments

3.1 Experimental Setup

3.1.1 Datasets

We evaluate our method on four mathematical reasoning benchmarks: **MATH-500** (Hendrycks et al., 2021), **AIME25**, **AIME24** (MAA Committees, 2024), and **AMC23** (AI-MO, 2024). To assess generality beyond math, we additionally include the **GPQA Diamond** split (Rein et al., 2024), a high-difficulty scientific reasoning dataset.

3.1.2 Models

We evaluate three RL-trained LRMs: **QwQ-32B** (Team, 2025b), **Qwen3-32B** (Team, 2025a), and **R1-Distill-Qwen-32B** (Guo et al., 2025).

3.1.3 Baselines

We compare inference-time strategies grouped by the signal used for length control: (i) **Vanilla** (no early stopping); (ii) **DEER** (Yang et al., 2025a), which induces a provisional answer and derives confidence from internal logits; (iii) **Ours**, which performs adaptive early stopping based on self-assessed confidence expressed during reasoning.

We restrict our evaluation to single-trajectory inference methods, excluding approaches that require multiple sampled candidates or voting, as these involve substantially different computational assumptions.

3.1.4 Implementation Details

All models are deployed using the `vLLM` backend, ensuring efficient inference. The experiments are conducted using the following decoding parameters: temperature 0.6, top-p 0.95, and top-k 20, which are commonly reported to yield strong overall performance. To ensure statistical reliability, we report the average results over three runs for each experiment.

3.2 Main results

As shown in Table 1, both DEER and Ours achieve reasoning tokens reductions in reasoning length compared to the Vanilla setting, with little or no loss in accuracy. In some cases, they even yield higher accuracy, indicating that appropriately guided early stopping does not necessarily harm correctness. Our method also shows stable performance on challenging datasets such as AIME25 and AIME24, suggesting that confidence-based early stopping remains reliable under increased task complexity.

Across models, distinct trends emerged. For Qwen3-32B, DEER yielded slightly lower accuracy and noticeably shorter reasoning, suggesting that its logit-derived confidence may be overconfident and prone to triggering early termination. In contrast, for R1-Distill-32B, DEER achieved marginally higher accuracy across datasets. This pattern indicates that the effectiveness of logit-based stopping can vary by model, potentially reflecting differences in how well each model’s internal probabilities are calibrated with respect to its actual reasoning progress.

Overall, these results support the hypothesis that RL-trained reasoning models possess well-aligned internal confidence, and that vanilla generation often produces redundant reasoning. Leveraging confidence as a stopping signal thus offers a promising path toward improving inference efficiency in reasoning-optimized language models.

3.3 Ablation results

3.3.1 Periodic Confidence Probing

As a keyword-free alternative, **Periodic-Conf(k)** pauses generation every k tokens to query self-assessed confidence. We evaluate two intervals, $k = 100$ and $k = 1000$, with results summarized in Table 2. Smaller intervals naturally produce more estimation points (AvgPoint), allowing more opportunities for early stopping. Consistent with

Dataset	MATH-500		AIME25		AIME24		AMC23		GPQA-diamond		Average	
Method	Accuracy↑	Length↓	Accuracy↑	Length↓	Accuracy↑	Length↓	Accuracy↑	Length↓	Accuracy↑	Length↓	Accuracy↑	Length↓
QwQ-32B												
Vanilla	93.80	3602	**66.67**	14857	66.67	13551	95.00	6782	61.62	7509	76.75	9260
DEER	93.80	2677	61.11	12412	73.33	10567	95.00	5678	**64.14**	6313	77.48	7529
Ours	**94.00**	2560	65.55	11985	**76.67**	11338	**97.50**	4944	62.12	6732	**79.17**	7512
Qwen3-32B												
Vanilla	94.20	3944	72.22	12724	72.22	11857	**97.50**	5411	65.48	6804	80.32	8148
DEER	93.80	2535	70.00	9605	71.11	7732	95.00	4154	63.97	4122	78.78	5630
Ours	**94.33**	2662	**75.56**	11321	**74.44**	10166	96.25	4182	65.66	4788	**81.25**	6624
R1-Distill-32B												
Vanilla	88.00	2632	41.11	12154	60.00	9731	82.50	5100	62.12	5670	66.75	7057
DEER	**88.47**	2545	**42.22**	8931	**63.33**	8445	83.33	4641	**63.84**	4326	**68.24**	5778
Ours	87.73	1855	41.11	8721	62.22	8529	**84.17**	3881	61.61	3947	67.37	5387

Table 1: Comparison of reasoning performance on four math benchmarks (MATH-500, AIME25, AIME24, AMC23). We report accuracy (↑) and reasoning length (↓) for five LRMs under three strategies: Vanilla, DEER, and Ours. Bold indicates the best accuracy within each model, and underlining denotes the shortest reasoning length.

Method	Acc↑	Length↓	AvgPoint
Periodic-Conf(100)	74.44	10881	91.29
Periodic-Conf(1000)	77.78	12182	12.30
Wait	76.67	11338	54.77
Alternatively	71.11	11297	30.67
Both	74.44	11038	59.90
Almost certain	76.67	11338	54.77
Highly	70.00	10954	36.30
Very	71.11	10762	36.10

Table 2: Comparison of confidence estimation strategies and threshold settings. Acc denotes accuracy; Length is number of generated tokens; AvgPoint indicates the average number of estimation points per trace.

this behavior, $k = 100$ yields noticeably shorter final reasoning lengths. However, the large number of queries also introduces substantial overhead, as confidence must be recomputed very frequently.

Conversely, $k = 1000$ greatly reduces the number of estimation points and the associated overhead, but may miss earlier stopping opportunities, resulting in longer reasoning traces. These observations indicate that the effectiveness of periodic probing depends heavily on how well the chosen interval matches the typical reasoning length of the task.

Since k must be selected in advance and cannot adapt to instance-specific reasoning needs, we treat periodic probing as an auxiliary keyword-free option, while using keyword-triggered estimation as our primary method. Its applicability to models with sparse reflective markers (e.g., instruction-tuned LLMs) is discussed further in Section 4.

3.3.2 Keyword-Based Estimation Strategies

We conduct an ablation study on which reflective markers are most suitable for triggering confidence estimation. After RL-based reasoning training, LRMs tend to use certain tokens—most notably "Wait" and "Alternatively"—to signal transitions in their thought process. We evaluate three variants: using only "Wait," only "Alternatively," and a combined setting that reacts to both.

Although keyword selection is inherently heuristic, it is less dependent on the dataset than periodic probing and instead more closely tied to each model's characteristic reasoning style. In our experiments, all three evaluated LRMs produce "Wait" more frequently than "Alternatively," and Table 2 shows that "Wait"-based triggering yields more stable accuracy and confidence estimates. In contrast, "Alternatively" appears less often and provides fewer, and in some cases less reliable, estimation points—possibly reflecting its use in more exploratory or tentative reasoning. The combined strategy falls between the two but does not outperform using "Wait" alone.

3.3.3 Impact of Confidence Threshold

To compare performance across different confidence thresholds, we evaluate three thresholds: *Very good chance (0.7–0.8)*, *Highly likely (0.8–0.9)*, and *Almost certain (0.9–1.0)*. As shown in Table 2, lowering the confidence threshold leads to a reduction in generation length and the number of estimation points, as expected. A notable observation is that there is little change when lowering the threshold from *Highly likely* to *Very good chance*, suggesting that most cases were already classified into the *Highly likely* or higher confidence classes. Although earlier results demonstrated the usefulness of self-assessed confidence, this reveals a potential limitation: the confidence values are not perfectly calibrated in a linear fashion, with many predictions clustered in higher confidence classes.

4 Discussions

4.1 Confidence alignment

We perform a tentative comparison between self-assessed confidence and a log-probability–based

proxy measured at matched points along the AIME24 reasoning traces. To place the two quantities on a comparable scale, self-assessment classes are mapped to approximate numeric midpoints. The resulting Pearson correlation of 0.33 reflects a modest but consistent positive association between the two measures.

Given that the log-probability–based proxy is itself only an indirect and potentially coarse indicator of confidence, the correlation should not be interpreted as evidence of precise calibration. Rather, the observation suggests that the two signals are not independent of one another and may be influenced by related aspects of the model's internal dynamics during reasoning. While preliminary, this trend offers a indirect support for the idea that LRMs encode some form of confidence-related signal, even if neither measure captures it in a fully calibrated way.

4.2 Length vs Correctness

To compare our adaptive approach with manually specified reasoning depth, we vary the length of the chain-of-thought using system-level instructions on MATH-500 and AIME24. Concretely, we prompt the model to produce *low*, *medium*, or *high* amounts of reasoning, and additionally include a *no-thinking* setting using the officially supported mode of Qwen3. These configurations follow common practice in prior work but rely on fixed instructions that cannot adjust to instance-level difficulty.

As shown in Figure 3, accuracy generally improves from low to moderate reasoning depth, after which further increases offer limited or even negative returns. This highlights the challenge of selecting a single, globally appropriate level of manual reasoning. Our self-assessment–based method reaches accuracy close to the best manual setting while using substantially fewer tokens, suggesting that adaptive stopping can provide a more efficient alternative to static, instruction-driven control.

4.3 Generalizability Beyond LRMs

We also examine whether our approach extends to instruction-tuned models by evaluating Qwen2.5-7B-Instruct on MATH-500 using periodic probing with $k = 300$. While reasoning length still decreases, the accuracy drops much more noticeably than in the LRM experiments. One contributing factor is that self-assessed confidence is generally more stable and better calibrated in LRMs, making confidence-based stopping less reliable in standard instruction-tuned models. Beyond calibration, structural differences also appear to matter: LRMs naturally separate their reasoning and answer phases and tend to accumulate sufficient intermediate thought before </think>, which allows them to recover a coherent final answer even when stopped early. In contrast, instruction-tuned models lack this mechanism. When </think> is replaced with a cue such as "So the answer is:", the model often shifts immediately to answering rather than consolidating prior reasoning, making early stopping more disruptive. Together, these observations suggest that confidence-based stopping is more dependable in models with an explicit reasoning phase and more stable confidence behavior.

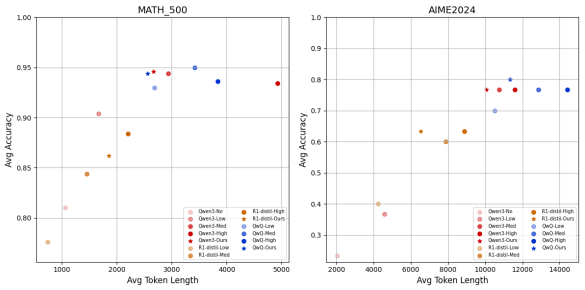

Figure 3: Accuracy vs. token length across reasoning settings for three models on MATH-500 and AIME24. Each point represents a model–setting pair: color denotes the model, transparency reflects reasoning depth (from No to High), and stars indicate Adaptive settings.

5 Conclusion

In this work, we propose a simple, training-free method for dynamically controlling reasoning length in RL-trained language models using self-assessed confidence. By prompting the model to evaluate its confidence during reasoning, our approach enables early stopping once sufficient certainty is reached, improving efficiency without accuracy loss. Our results show that self-assessed confidence aligns well with internal signals such as logit-based estimates, indicating meaningful representations of certainty. This work advances understanding of how confidence can guide efficient reasoning, and points toward future training of LRMs that inherently produce near-optimal reasoning traces.

Limitations

While self-assessed confidence provides a simple mechanism for controlling reasoning length, several limitations remain. Our evaluation focuses

primarily on mathematical reasoning tasks, and it is unclear how well the approach transfers to open-ended or multimodal settings where reasoning structure is less consistent. The method also depends on specific prompting formats and threshold choices, which may not generalize across models and can introduce additional inference overhead.

A further limitation is that self-assessed confidence appears more reliable in LRMs than in standard instruction-tuned LLMs. Models without an explicit reasoning–answer separation can be more sensitive to early stopping, and their self-generated confidence tends to be less stable, leading to greater accuracy degradation. This suggests that the effectiveness of confidence-based control may depend on properties that are more pronounced in RL-trained reasoning models than in general-purpose LLMs.

At the same time, recent model developments increasingly emphasize more explicit or reflective reasoning behaviors. As such trends continue, confidence-guided stopping may become more broadly applicable. Future work could explore fine-tuning strategies, prompting techniques, or lightweight architectural adjustments that encourage more stable intermediate reasoning and improve the calibration of self-assessed confidence in models beyond LRMs.

GenAI Usage Disclosure

During the preparation of this manuscript, Generative AI tools were used primarily for translation and refining the phrasing of English sentences. In some cases, they were also consulted for minor suggestions regarding experimental code revisions. All core ideas, experimental design, analysis, and conclusions were independently developed by the authors, and AI tools were used only for supportive purposes.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT). (No. RS-2025-00562784) The ICT at Seoul National University provides research facilities for this study.

References

AI-MO. 2024. Amc 2023, 2024. https://huggingface.co/datasets/AI-MO/aimo-validation-amc.

Qiguang Chen, Libo Qin, Jinhao Liu, Dengyun Peng, Jiannan Guan, Peng Wang, Mengkang Hu, Yuhang Zhou, Te Gao, and Wanxiang Che. 2025. Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567*.

Yichao Fu, Junda Chen, Yonghao Zhuang, Zheyu Fu, Ion Stoica, and Hao Zhang. 2025. Reasoning without self-doubt: More efficient chain-of-thought through certainty probing. In *ICLR 2025 Workshop on Foundation Models in the Wild*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Michael Hassid, Gabriel Synnaeve, Yossi Adi, and Roy Schwartz. 2025. Don't overthink it. preferring shorter thinking chains for improved llm reasoning. *arXiv preprint arXiv:2505.17813*.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, and 1 others. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Zhong-Zhi Li, Duzhen Zhang, Ming-Liang Zhang, Jiaxin Zhang, Zengyan Liu, Yuxuan Yao, Haotian Xu, Junhao Zheng, Pei-Jie Wang, Xiuyi Chen, and 1 others. 2025. From system 1 to system 2: A survey of reasoning large language models. *arXiv preprint arXiv:2502.17419*.

Wenjie Ma, Jingxuan He, Charlie Snell, Tyler Griggs, Sewon Min, and Matei Zaharia. 2025. Reasoning models can be effective without thinking. *arXiv preprint arXiv:2504.09858*.

MAA Committees. 2024. Aime problems and solutions. https://artofproblemsolving.com/wiki/index.php/AIME_Problems_and_Solutions.

David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. 2024. Gpqa: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*.

Jinyan Su, Jennifer Healey, Preslav Nakov, and Claire Cardie. 2025. Between underthinking and overthinking: An empirical study of reasoning length and correctness in llms. *arXiv preprint arXiv:2505.00127*.

Qwen Team. 2025a. Qwen3 technical report. *Preprint*, arXiv:2505.09388.

Qwen Team. 2025b. Qwq-32b: Embracing the power of reinforcement learning.

Yue Wang, Qiuzhi Liu, Jiahao Xu, Tian Liang, Xingyu Chen, Zhiwei He, Linfeng Song, Dian Yu, Juntao Li, Zhuosheng Zhang, and 1 others. 2025. Thoughts are all over the place: On the underthinking of o1-like llms. *arXiv preprint arXiv:2501.18585*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Fengli Xu, Qianyue Hao, Zefang Zong, Jingwei Wang, Yunke Zhang, Jingyi Wang, Xiaochong Lan, Jiahui Gong, Tianjian Ouyang, Fanjin Meng, and 1 others. 2025. Towards large reasoning models: A survey of reinforced reasoning with large language models. *arXiv preprint arXiv:2501.09686*.

Chenxu Yang, Qingyi Si, Yongjie Duan, Zheliang Zhu, Chenyu Zhu, Zheng Lin, Li Cao, and Weiping Wang. 2025a. Dynamic early exit in reasoning models. *arXiv preprint arXiv:2504.15895*.

Wenkai Yang, Shuming Ma, Yankai Lin, and Furu Wei. 2025b. Towards thinking-optimal scaling of test-time compute for llm reasoning. *arXiv preprint arXiv:2502.18080*.

Xixian Yong, Xiao Zhou, Yingying Zhang, Jinlin Li, Yefeng Zheng, and Xian Wu. 2025. Think or not? exploring thinking efficiency in large reasoning models via an information-theoretic lens. *arXiv preprint arXiv:2505.18237*.

Dongkeun Yoon, Seungone Kim, Sohee Yang, Sunkyoung Kim, Soyeon Kim, Yongil Kim, Eunbi Choi, Yireun Kim, and Minjoon Seo. 2025. Reasoning models better express their confidence. *arXiv preprint arXiv:2505.14489*.

Jiajie Zhang, Nianyi Lin, Lei Hou, Ling Feng, and Juanzi Li. 2025a. Adaptthink: Reasoning models can learn when to think. *arXiv preprint arXiv:2505.13417*.

Shengjia Zhang, Junjie Wu, Jiawei Chen, Changwang Zhang, Xingyu Lou, Wangchunshu Zhou, Sheng Zhou, Can Wang, and Jun Wang. 2025b. Othink-r1: Intrinsic fast/slow thinking mode switching for over-reasoning mitigation. *Preprint*, arXiv:2506.02397.