ShorterBetter: Guiding Reasoning Models to Find Optimal Inference Length for Efficient Reasoning

Jingyang Yi[1*], Jiazheng Wang[1*], Sida Li[1]

[1] Data Science Institute, The University of Chicago *equal contribution {jingyang22, jiazhengw, listar2000}@uchicago.edu

Background & Contributions

✅ Large reasoning models (**LRMs**, e.g. OpenAI O1 models) have shown impressive performance in complicated benchmarks in math, programming, logic, etc.

❌ However, LRMs exhibit **overthinking** phenomenon that both (1) is computationally **inefficient**, and (2) hurts **performance** (see below analysis).

💡 We propose **ShorterBetter**, a novel **reinforcement learning** method that guides reasoning model to autonomously learn its **optimal reasoning lengths**.

📊 We introduce novel **reasoning trace analysis** methods to demonstrate how **ShorterBetter** enhances reasoning efficiency.

Optimal Length & Sample Optimal Lengh

A Theoretical Framework to Model **Optimal Inference Length**

- For a prompt x_i, let $y = (y^1, ..., y^t)$ be sampled auto-regressively from LLM, i.e., $y \sim p_\theta(\cdot \mid x_i)$, with $\ell(y) := t$ be its length.
- Let y_i^* be the **reference response** for x_i and $\mathscr{I}$ be a verifier that assigns a **correctness score** $s_{\mathscr{I}}(y, y_i^*) \in [0, 1]$ (higher better).
- Let $c \geq 0$ be a **correctness threshold** and $\varepsilon \geq 0$ be **probability threshold**, the **optimal reasoning length (OL)** is defined as

$$\ell^*_{c,\varepsilon}(x_i; \theta) := \min\{\ell(y) : y \in \mathscr{Y}_{c,\varepsilon}(x_i; \theta)\},$$

where

$$\mathscr{Y}_{c,\varepsilon}(x_i; \theta) := \{y : p_\theta(y \mid x_i) \geq \varepsilon \wedge s_{\mathscr{I}}(y, y_i^*) \geq c\}.$$

- **Interpretation:** $\ell^*_{c,\varepsilon}(x_i; \theta)$ represents the **minimal length** at which the model **reliably** (probability above ε) produces a **sufficiently correct** (score above c) response.

A Practical Implementation of **Sample-based Optimal Length**

- For a prompt x_i, the LLM samples m **rollouts** of candidate responses $G(x_i) = \{y_1, y_2, \ldots, y_n\}$.
- We then define the **sample optimal length (SOL)** for $G(x_i)$ as:

$$\ell^{\text{SOL}}(G(x_i)) = \begin{cases} \min_{y_j \in G(x_i): \mathbb{I}(y_j = y_i^*) = 1} \ell(y_j) & \text{if at least one correct response exists} \\ \frac{1}{n}\sum_{j=1}^{n} \ell(y_j) & \text{otherwise} \end{cases}$$

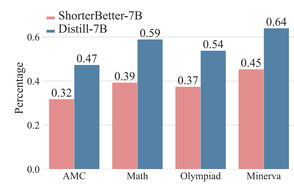
- The corresponding **reward function** for y_i is:

$$r(y_j) = \alpha \cdot \mathbb{I}(y_j = y_i^*) - \beta \cdot \left| \ell(y_j) - \ell^{\text{SOL}}(G(x_i)) \right|,$$

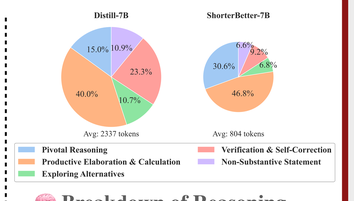
where α, β are hyperparameters balancing correctness and length.

Correctness reward — Length reward

Case Study



🔍 **Output Length after First Correct Answer:**
ShorterBetter generates **15–20% fewer tokens** after the first correct answer, indicating **less redundant self-verification** compared to the DeepSeek-Distill model.

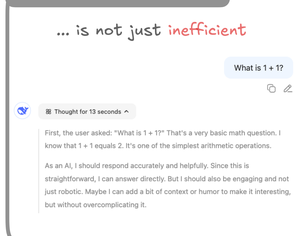


🧠 **Breakdown of Reasoning Traces:**
ShorterBetter produces **more concise, solution-driven reasoning** by increasing **pivotal steps** and **productive calculations**, while reducing **filler**, **self-corrections**, and **exploratory detours**.

Efficient Reasoning

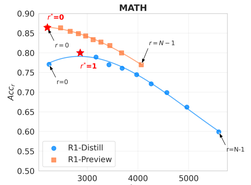
Overthinking

... is not just inefficient



Source: [Su et al. 2025]

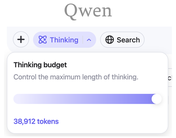
but also hurts performance

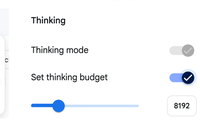


Existing Methods

For a comprehensive **review**, see : [Qu et al. 2025]

1: explicit **thinking budgets**





😐 Requires **manual** control

☹️ Does not **adapt** to the prompt/problem **difficulty**

2: use **reference models**

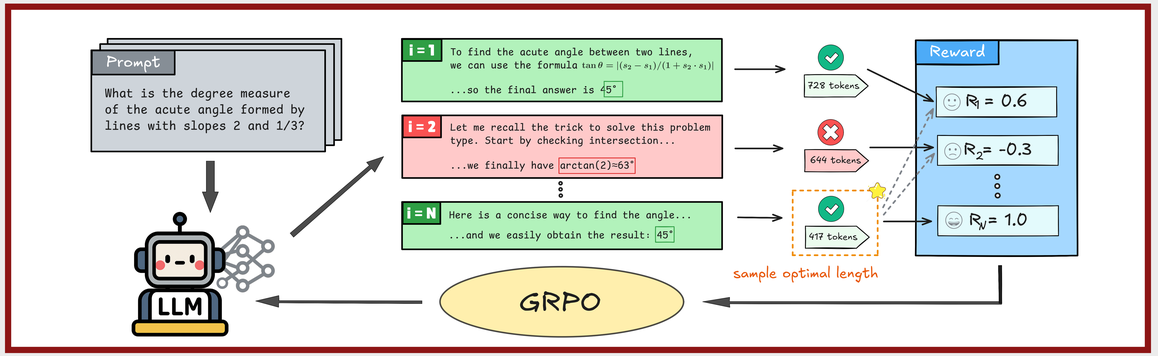
Reward Function O1-Pruner: [Luo et al. 2025]

$$\frac{L_{ref}}{L(y)} - 1 + \lambda(S(y) - S(y_{ref}))$$

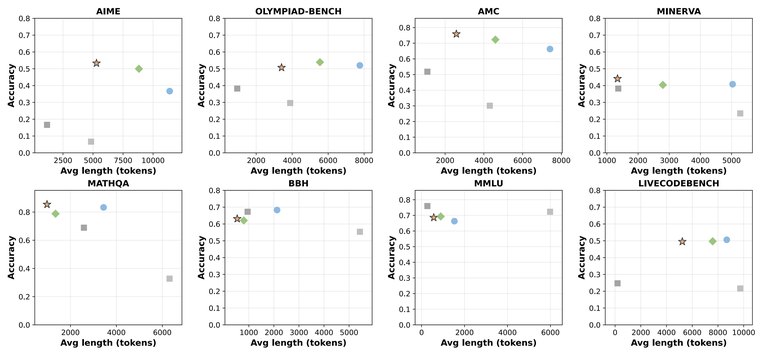
Length reward — Correctness reward

☹️ Requires **external** reference model

🤔 The target LLM it-self should **know its own "budget"**the best



Results



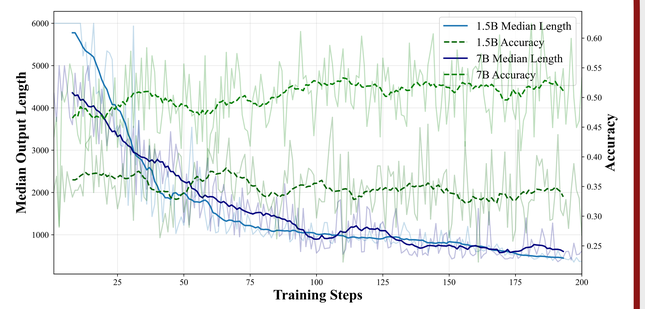
Compare performance of 7B-scale models across diverse benchmarks:

Metrics: Report both **accuracy** and **output length**.

Models:

- DeepSeek-R1-Distill-Qwen-7B
- Qwen2.5-7B-Instruct (non-reasoning baseline)
- ShorterBetter-7B with α = 2
- Training Efficient-7B (Arora & Zanette, 2025)
- O1-Pruner (Luo et al., 2025a)

Ablation



⚠️ **Naively adopting shortest response length destabilizes training**

🎯 **Sample Optimal Length (SOL) balances correctness and conciseness**

Resource

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

Daman Arora and Andrea Zanette. Training language models to reason efficiently. arXiv preprint arXiv:2502.04463, 2025.

Haotian Luo, Li Shen, Haiying He, Yibo Wang, Shiwei Liu, Wei Li, Naiqiang Tan, Xiaochun Cao, and Dacheng Tao. O1-pruner: Length-harmonizing fine-tuning for o1-like reasoning pruning. arXiv preprint arXiv:2501.12570, 2025a