System 2 AI for Language Models: Formal Pathways via Chains, Votes, Trees, and Meta-Reasoning

Amir Mohammad Sharafaddini

July 5, 2025

Scope and Position

This article develops a formal, practical account of *System 2* reasoning for large language models (LLMs) using four mechanisms that have concrete empirical support in the literature: (i) chain-of-thought prompting (CoT), (ii) self-consistency aggregation (SC), (iii) tree-of-thoughts search (ToT), and (iv) meta chain-of-thought (Meta-CoT) with process supervision. Each mechanism is expressed as an optimization over intermediate structures (thought sequences, votes, trees, latent traces) with provable properties or principled surrogates, and each is connected to reported measurements in the source papers.

1 Problem Setting

Let $x = (x_1, \dots, x_n)$ denote an input sequence and y a target. An autoregressive LLM with parameters θ induces $p_\theta(t_k \mid t_{<k}, x)$ over tokens t_k. A *task* is an unknown conditional $p^\star(y \mid x)$; success is judged by a verifier $v(y, \hat{y} \mid x) \in \{0, 1\}$.

Definition 1.1 (Reasoning Program). *A reasoning program* Π *is any stochastic mapping that produces an answer by allocating an inference-time* compute budget C *over proposal, expansion, evaluation, and verification phases:*

$$\Pi(x; C) \;\rightsquigarrow\; \hat{y} \in \mathcal{Y}, \qquad C = \sum_j \mathrm{tok}(\mathit{phase}_j), \tag{1}$$

where $\mathrm{tok}(\cdot)$ *counts generated tokens. A procedure is* System 2 *if* C *can be adaptively increased to branch, backtrack, and verify prior to commitment.*

We analyze four instances of Π that are widely used in practice and studied in the provided papers.

2 Chain-of-Thought Prompting

Factorization. Chain-of-thought (CoT) augments inference with an explicit sequence of intermediate *thoughts* $z_{1:n}$ before emitting the answer:

$$p_\theta(y \mid x) \;=\; \sum_{z_{1:n}} p_\theta(y \mid x, z_{1:n})\, p_\theta(z_{1:n} \mid x). \tag{2}$$

Few-shot demonstrations of $z_{1:n}$ steer generation toward multi-step decompositions, producing large improvements on arithmetic and symbolic tasks when models are sufficiently large.

Emergence with scale. Empirical evidence shows that providing a handful of CoT exemplars to a 540B-parameter model achieves state-of-the-art accuracy on GSM8K, surpassing specialized finetuning baselines.

Limits. CoT narratives can be post-hoc or incorrect even when answers are correct; annotation cost and faithfulness are noted limitations.

3 Self-Consistency Aggregation

Voting over sampled chains. Given m independently sampled reasoning paths $\{(z^{(i)}_{1:n_i}, a^{(i)})\}_{i=1}^{m}$, self-consistency outputs the majority label

$$\hat{a}_{\mathrm{SC}} = \arg\max_{a} \sum_{i=1}^{m} \mathbb{1}\{a^{(i)} = a\}. \tag{3}$$

Length- or confidence-weighted variants replace $\mathbb{1}$ by $w_i\,\mathbb{1}\{\cdot\}$ where

$$w_i = \exp\left(\frac{1}{K_i}\sum_{k=1}^{K_i} \log p_\theta(t^{(i)}_k \mid t^{(i)}_{<k}, x)\right). \tag{4}$$

Risk reduction under independence. If a single path is correct with probability $p > 1/2$ and samples are i.i.d., the probability that the majority vote is correct is

$$\mathbb{P}[\text{correct}] = \sum_{k=\lceil (m+1)/2\rceil}^{m} \binom{m}{k} p^k (1-p)^{m-k} \;\geq\; 1 - \exp\Big(-2m\,\big(p-\tfrac{1}{2}\big)^2\Big), \tag{5}$$

by a Chernoff bound, revealing exponential decay of error with m.

Empirical profile. Across arithmetic and commonsense benchmarks, self-consistency yields striking gains (e.g., GSM8K +17.9 points with PaLM-540B; robust to sampling strategy and model scale) and outperforms beam search and prompt-ensembles at fixed sample counts. It also improves zero-shot CoT and equation-style intermediate traces; vote consistency correlates with accuracy, providing a usable uncertainty proxy.

Compute accounting. Let c_{tok} be the per-token cost and ℓ_{cot} the average length of one CoT decode. Then

$$\mathbb{E}[\text{cost}_{\text{SC}}] = c_{\text{tok}}(m\,\mathbb{E}[\ell_{\text{cot}}] + \ell_{\text{vote}}), \qquad \text{gain}(m) \approx 1 - \exp\Big(-2m\,\big(p-\tfrac{1}{2}\big)^2\Big). \tag{6}$$

Empirical scaling curves for accuracy vs. #paths support the monotone improvement predicted by (5).

4 Tree-of-Thoughts Search

State, expansion, heuristic. ToT upgrades chain sampling to stateful search over thoughts. Let $s_t = (x, z_{1:t})$ be a node; successors are candidate next thoughts $z^{(j)}_{t+1} \sim p_\theta(\cdot \mid x, z_{1:t})$. A heuristic $h(s_t) \in \mathbb{R}$ estimates promise, and a policy (BFS/DFS/beam) expands the frontier $\mathcal{F}$ until a verified solution appears:

$$\mathcal{F} \leftarrow \text{TopB}\big(\mathcal{F} \cup \{(x, z_{1:t}, z^{(j)}_{t+1})\}_{j=1}^{k}\ ;\ h\big), \qquad \text{stop if } v(x, z_{1:t}, y) = 1. \tag{7}$$

Empirical evidence. On tasks requiring lookahead (e.g., Game of 24), ToT with GPT-4 attains $\sim 74\%$ vs. $\sim 4\%$ for plain CoT; it also lifts creative writing and crosswords with appropriate thought granularity and heuristics.

Algorithm 1 Tree-of-Thoughts(x, B, k, h)

```
$\mathcal{F} \leftarrow \{(x, \varnothing)\}$
**while** budget not exhausted **do**
    select $s = (x, z_{1:t}) \in \mathcal{F}$ with highest $h(s)$
    sample $k$ successors $\{z_{t+1}^{(j)}\}_{j=1}^{k} \sim p_\theta(\cdot \mid x, z_{1:t})$
    **for** $j = 1, \dots, k$ **do**
        $s' \leftarrow (x, z_{1:t}, z_{t+1}^{(j)})$; compute $h(s')$
        **if** $s'$ yields verified $y$ **then return** $y$
        **end if**
        $\mathcal{F} \leftarrow \mathcal{F} \cup \{s'\}$
    **end for**
    $\mathcal{F} \leftarrow \mathrm{TopB}(\mathcal{F}; h)$
**end while**
**return** best verified candidate (if any)
```

Complexity profile. Let B_t be the beam width at depth t and ℓ_{step} the expected token cost per expansion. The token cost is

$$\mathbb{E}[\text{cost}_{\text{ToT}}] \approx c_{\text{tok}} \sum_{t=1}^{d} \mathbb{E}[B_t]\, \mathbb{E}[\ell_{\text{step}}], \quad \text{with } \mathbb{E}[B_t] \text{ controlled by pruning via } h. \tag{8}$$

Reported ablations show accuracy-cost trade-offs by varying k, B, and the evaluation prompt used to compute h.

5 Meta Chain-of-Thought

Latent-process model. Meta-CoT posits an *unseen* deliberation trace $Z = (z_{1:K}^{\text{latent}})$ that causally precedes both the visible chain $S = (s_{1:n})$ and the answer a:

$$p(a, S \mid q) \;=\; \int p(a, S \mid Z, q)\, p(Z \mid q)\, dZ. \tag{9}$$

This legitimizes training on linearized *search traces* and process rewards to internalize System 2 behavior within a single model.

Learning objectives. Two practical objectives arise. *Latent-variable ELBO*:

$$\max_{\theta, q} \; \mathbb{E}_{Z \sim q(\cdot \mid q)}\big[\log p_\theta(S \mid Z, q)\big] - \beta\, \mathrm{KL}\big(q(Z \mid q) \,\|\, p_\theta(Z \mid q)\big). \tag{10}$$

Process reward maximization with a process reward model $r_{\text{proc}}(q, S_{1:t}) \in [0, 1]$:

$$\max_{\theta} \; \mathbb{E}\Big[\sum_{t=1}^{n} \gamma^{t-1}\, r_{\text{proc}}(q, S_{1:t})\Big]. \tag{11}$$

Both support amortizing search into a single autoregressive policy.

Meta-STaR objective. When verified search traces $\widehat{Z}, \widehat{S}$ are available, a simple objective is

$$\mathcal{L}_{\text{Meta-STaR}}(\pi_\phi) \;=\; -\,\mathbb{E}_{(q, \widehat{Z}, \widehat{S})}\big[\log \pi_\phi(\widehat{S}, \widehat{Z} \mid q)\big], \tag{12}$$

i.e., teach the model to generate both the hidden search and the visible plan.

Empirical analyses. Evidence is presented that frontier models exhibit in-context search-like scaling with inference budget and that search traces can be internalized in small controlled domains (A* in mazes; Countdown arithmetic), with performance improving as training and inference compute increase.

6 Generator–Verifier Factorization

Definition 6.1 (Generator and Process Verifier). *A generator G_θ induces $p_\theta(S \mid q)$; a process verifier V_ϕ scores partial traces $S_{1:t}$:*

$$V_\phi(q, S_{1:t}) \in [0,1], \qquad \textit{and} \qquad \max_\theta \; \mathbb{E}_{S \sim p_\theta(\cdot|q)}\big[V_\phi(q,S)\big]. \tag{13}$$

This decomposition enables guided search (ToT) and process reward training (Meta-CoT), and mirrors classic generator–value architectures in search and games.

7 Statistical Analyses

7.1 Majority Aggregation and Diversity

Let $\mathcal{A}$ be a finite answer set and $\pi(a) = \mathbb{P}(a^{(i)} = a)$ for a single path. The Bayes-optimal 0–1 decision selects $a^\star = \arg\max_a \pi(a)$. The m-sample majority error is

$$\mathcal{R}_m = 1 - \sum_{k=\lceil (m+1)/2 \rceil}^{m} \binom{m}{k} \pi(a^\star)^k \big(1 - \pi(a^\star)\big)^{m-k}, \tag{14}$$

decreasing in m. Correlated samples reduce the effective m; diverse decoding (temperature, top-k, nucleus) mitigates correlation—a design principle supported by ablations.

7.2 Faithfulness and Surrogate Rewards

Let $R(x, S, a)$ be the ground-truth process reward available only on a subset of tasks (e.g., verifiable math). We optimize a surrogate $\widehat{R} = V_\phi(q, S_{1:t})$ trained from outcomes; mismatch $|R - \widehat{R}|$ contributes to bias in learned search policies. Meta-CoT proposes to reduce this gap by modeling latent Z and training on linearized traces.

8 Compute and Complexity

8.1 Test-Time Budgets

Combining (6) and (8) yields a unified notion of *test-time compute* C:

$$C = m\,\ell_{\text{cot}} + \ell_{\text{vote}} \quad \text{(self-consistency)}, \qquad C = \sum_{t=1}^{d} B_t\,\ell_{\text{step}} \quad \text{(tree-of-thoughts)}. \tag{15}$$

Empirical curves show monotone accuracy increase with C across models and tasks.

8.2 Search Depth and Branching

For branching factor b and effective depth d_{eff}, the open-list size in BFS scales as $O(b^{d_{\text{eff}}})$; beam search constrains this to $O(Bd_{\text{eff}})$. Heuristics h reduce d_{eff} by pruning subtrees whose expected verification probability is low, a language-native instantiation emphasized in ToT.

9 Formal Procedures

9.1 Self-Consistency as Approximate Marginalization

Define the CoT-induced posterior over answers

$$\pi_\theta(a \mid x) = \sum_{z_{1:n}} p_\theta(a \mid x, z_{1:n})\, p_\theta(z_{1:n} \mid x). \tag{16}$$

Sampling m chains and voting approximates $\arg\max_a \pi_\theta(a \mid x)$, with Monte Carlo error decaying in m. Beam search concentrates on a narrow mode and under-explores diverse correct chains, consistent with head-to-head comparisons.

9.2 ToT as Stochastic Planning

Let the transition kernel be $\mathcal{T}((x, z_{1:t}) \to (x, z_{1:t}, z_{t+1})) = p_\theta(z_{t+1} \mid x, z_{1:t})$ and define a value surrogate $h(x, z_{1:t}) \approx \mathbb{P}(\text{reachable \& verifiable} \mid x, z_{1:t})$. Then

$$\text{policy:} \quad \pi(s) = \mathrm{TopB}\big(\mathrm{Succ}(s); h\big), \qquad \text{stop when } v = 1. \tag{17}$$

Language-native heuristics h provide a practical bridge between LLMs and classical search.

9.3 Meta-CoT as Latent-Variable Inference

Combining (9)–(10), training jointly over S and Z aligns surface plans and hidden deliberation with process supervision; Section 4 of the source delineates empirical setups (A*, MCTS traces; Countdown; analyze budget vs. success).

10 Selected Empirical Facts Tied to the Formalism

- CoT with PaLM-540B achieves SOTA on GSM8K with few-shot exemplars; CoT emerges at large scale.
- Self-consistency delivers large absolute gains across arithmetic and commonsense tasks, robust across sampling regimes and model scales; it beats beam search and prompt ensembles at matched budgets.
- ToT reframes reasoning as tree search over thoughts and substantially improves success on lookahead-heavy tasks (e.g., $\sim$74% on Game of 24).
- Meta-CoT formalizes latent search and provides training pipelines to internalize search traces and process rewards; analyses show budget–performance scaling and small-domain internalization.

11 Limitations and Assumptions

- **Faithfulness.** Visible chains can be rationalizations; Meta-CoT narrows but does not eliminate the gap between story and computation.
- **Verifier bottlenecks.** Many domains lack cheap verifiers; ToT requires useful h or PRMs to avoid verbosity without accuracy.
- **Data quality.** Advanced reasoning datasets often contain label noise or insufficient difficulty diversity, complicating process-supervised training.

References

[1] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, D. Zhou. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. NeurIPS 2022.

[2] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. H. Chi, S. Narang, A. Chowdhery, D. Zhou. *Self-Consistency Improves Chain-of-Thought Reasoning in Language Models*. ICLR 2023.

[3] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, K. Narasimhan. *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*. 2023.

[4] V. Xiang, C. Snell, K. Gandhi, A. Albalak, A. Singh, *et al. Towards System 2 Reasoning in LLMs: Learning How to Think With Meta Chain-of-Thought*. arXiv:2501.04682, 2025.