How Recursive Language Models Generalize

Chenxiao Yang, Zhiyuan Li, David McAllester, and Nathan Srebro

Toyota Technological Institute at Chicago

Large language models are increasingly used as agents that generate long reasoning traces, decompose tasks into subproblems, and manage separate contexts while interacting with users, tools, or other model instances. A central design choice in such systems is what the model should be allowed to see at each step. In standard chain-of-thought (CoT) reasoning, every subproblem and every intermediate step is written into one growing sequence, and the next-token predictor reads the complete generated trace [1,2]. A recursive language model instead starts a new isolated sequence for each subproblem, runs the same predictor there, and passes only the answer back to the parent [3]. The immediate benefit is computational: the recursive model reads a shorter sequence at each step and can handle computations whose flattened CoT trace would exceed the context budget. This paper asks whether the same context-isolation principle also improves learning and reliability.

The answer depends on the kind of generalization. In distribution, limiting context gives little statistical advantage. CoT is the more general learner: it reads the complete sequence rather than enforcing the context isolation that the recursive model builds in. We show that this generality is cheap in distribution. A CoT Transformer can reconstruct the sequence the recursive model would have seen and apply the same next-token rule, using only constant overhead in depth and parameter count. Thus the standard IID description-length generalization guarantee changes only by a constant factor. This matches our experiments: on i.i.d. samples from the training generator, the recursive model may learn faster early on, but CoT catches up.

Out of domain, the conclusion changes. The same generality that lets CoT contain the recursive rule also lets it use tokens outside the current subproblem. These tokens can form easy-to-read patterns that are predictive in the training domain without being part of the subproblem's actual rule. For example, an intermediate value may remain visible in the flattened trace and happen to equal the final answer on the training domain, so CoT can learn to copy it rather than solve the subproblem. Recursive context isolation hides such cues by construction.

We formalize this danger using a minimum-description-length view of learning. The simplicity bias that makes generality cheap in distribution does not distinguish the intended mechanism from a shorter rule over the complete sequence that merely fits training. The desired counterproperty is invariance: if the correct answer for a subproblem depends only on the subproblem's own description, then changing the surrounding sequence should not change the prediction [4–7]. The recursive model satisfies this by construction. CoT can represent the same invariant rule, but it can also represent shorter non-invariant shortcut rules that fit the same training examples. We prove that when such a shortcut is shorter than the isolated-context rule, the CoT predictor must break this invariance.

Experiments on recursive expression-evaluation tasks support this mechanism. CoT performs well on held-out in-distribution examples, but degrades sharply on traces that are longer or deeper than those seen in training. In controlled shortcut experiments, its predictions often follow trace-level cues that were valid only in the training domain. The recursive model remains substantially more robust because those cues are not visible to the predictor.

These results connect context management in LLM agents with a basic learning-theoretic question: which hypotheses are made simple by the information exposed to the learner? For IID prediction, it is often sufficient that the model class contains a good rule. For out-of-domain reasoning, coverage is not enough: seeing more can make the wrong rule easier to learn. Reliable agentic reasoning may therefore require specificity that rules out bad explanations, not merely a general class that contains the right one.

References.

[1] M. Nye et al. Show your work: Scratchpads for intermediate computation with language models. 2021.

[2] J. Wei et al. Chain-of-thought prompting elicits reasoning in large language models. NeurIPS, 2022.

[3] C. Yang, N. Srebro, and Z. Li. Recursive models for long-horizon reasoning. ICML, 2026.

[4] J. Rissanen. Modeling by shortest data description. Automatica, 1978.

[5] P. Grunwald. The Minimum Description Length Principle. MIT Press, 2007.

[6] J. Peters, P. Buhlmann, and N. Meinshausen. Causal inference using invariant prediction. JRSS-B, 2016.

[7] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz. Invariant risk minimization. 2019.