Idan Mehalel
The NSF-Simons collaboration on deep learning theory

A Theory of Online Learning with Autoregressive Chain-of-Thought Reasoning

NeurIPS 2026

Joint work with

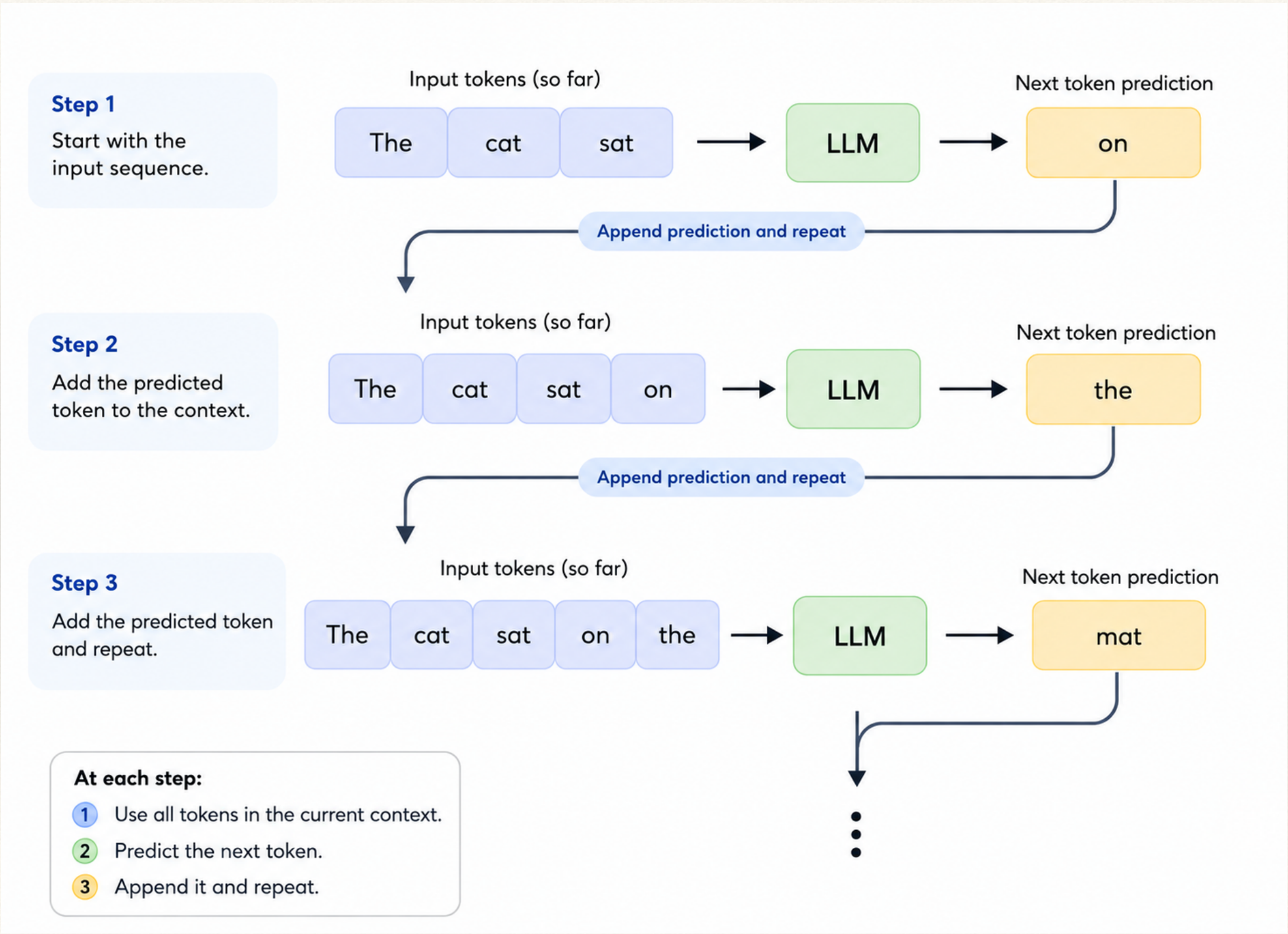


Ilan Doron-Arad
MIT



Elchanan Mossel
MIT

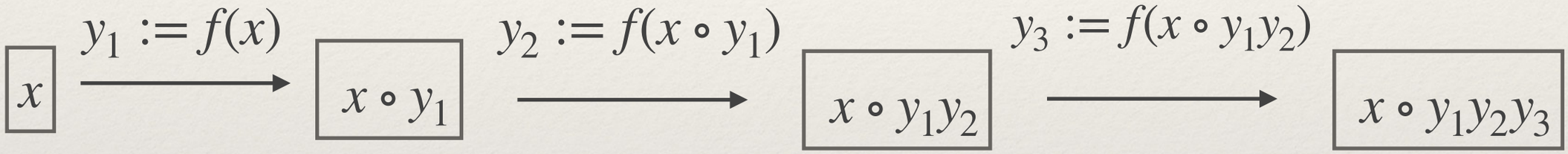
LLMs are autoregressive next-token generators



The Autoregressive Chain-of-Thought model

[Joshi, Vardi, Block, Goel, Li, Misiakiewicz, Srebro '25]

Next-token-generator $f : \{0,1\}^\star \to \{0,1\}$



$M = 3$: Three chain-of-thought steps

$$f^{\mathsf{e2e}-M}(x) = y_3 \qquad f^{\mathsf{CoT}-M}(x) = y_1 y_2 y_3$$

Goal: Learn a target $f_\star^{\mathsf{e2e}-M} \in \mathscr{F}^{\mathsf{e2e}-M}$

$$\mathscr{F}^{\mathsf{e2e}-M} := \{f^{\mathsf{e2e}-M} : f \in \mathscr{F}\}$$

Supervision models

End-to-End (e2e)

Chain-of-Thought (CoT)

Goal: Learn a target $f_\star^{\mathsf{e2e}-M} \in \mathscr{F}^{\mathsf{e2e}-M}$

Input examples: $(x, f_\star^{\mathsf{e2e}-M}(x))$

$f^{\mathsf{e2e}-3}(x) = y_3$

Standard online learning of $\mathscr{F}^{\mathsf{e2e}-M}$

Input examples: $(x, f_\star^{\mathsf{CoT}-M}(x))$

$f^{\mathsf{CoT}-3}(x) = y_1 y_2 y_3$

Online learning of $\mathscr{F}^{\mathsf{e2e}-M}$ with feedback that exposes the Chain-of-Thought

The online learning game

An adversary picks $x_1, \ldots, x_T \in \{0,1\}^\star$ and $f_\star \in \mathscr{F}$.

In each round $t \in [T]$, x_t is revealed to the learner, then the learner predicts $\hat{y}_T \in \{0,1\}$ and receives feedback:

- $f_\star^{\mathsf{CoT}-M}(x_t)$ if CoT-supervision
- $f_\star^{\mathsf{e2e}-M}(x_t)$ if e2e-supervision

Finally, the learner suffers a *mistake* if $\hat{y}_t \neq f_\star^{\mathsf{e2e}-M}(x_t)$.

We study the *optimal mistake bounds*:

$\mathtt{MB}^{\mathsf{e2e}}(\mathscr{F}^{\mathsf{e2e}-T})$, with e2e-supervision

$\mathtt{MB}^{\mathsf{CoT}}(\mathscr{F}^{\mathsf{e2e}-T})$ with CoT-supervision

Selected main results

Theorem (e2e worst-case bounds) [this work]:

$$\mathtt{MB}^{\mathsf{e2e}}(\mathscr{F}^{\mathsf{e2e}-M}) = O(\mathtt{MB}(\mathscr{F}) \cdot \log M)$$

for all $\mathscr{F}$ and $M \geq \mathtt{MB}(\mathscr{F})$. Furthermore, for some classes and all large enough M:

$$\mathtt{MB}^{\mathsf{e2e}}(\mathscr{F}^{\mathsf{e2e}-M}) \overset{\text{Known}}{\geq} \mathtt{VC}(\mathscr{F}^{\mathsf{e2e}-M}) = \Omega(\mathtt{MB}(\mathscr{F}) \cdot \log M).$$

The lower bound on $\mathtt{VC}(\mathscr{F}^{\mathsf{e2e}-M})$ resolves an open question of [Joshi et al. '25].

Theorem (CoT worst-case bound) [this work]:

$$\mathtt{MB}^{\mathsf{CoT}}(\mathscr{F}^{\mathsf{e2e}-M}) \leq \mathtt{MB}(\mathscr{F})$$

for any learnable class $\mathscr{F}$ and generation length M. **No dependence on M.**