CoT Information: Improved Sample Complexity under Chain-of-Thought Supervision

Awni Altabaa, Omar Montasser, John Lafferty

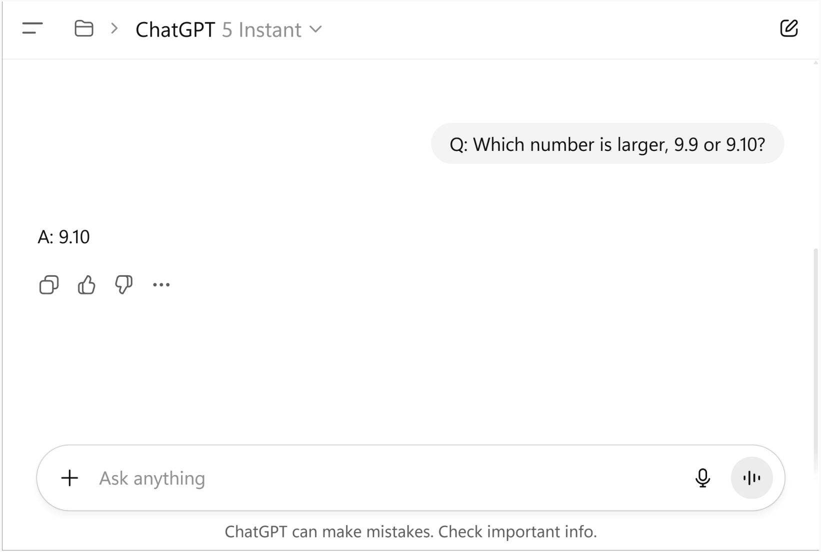
Yale University

`arXiv:2505.15927`, NeurIPS ’25

A Question for ChatGPT...

Question: *Which number is larger, 9.9 or 9.10?*

ChatGPT 5 Instant ✗



ChatGPT 5 Thinking ✓

ChatGPT 5 Thinking

Which number is larger, 9.9 or 9.10?

Thought for 7s

- The difference between 9.9 and 9.10 is subtle but important. In decimal notation, trailing zeros don't change the value, so 9.10 is the same as 9.1. But in versioning, 9.10 could be considered larger than 9.9. For a basic comparison, 9.90 is greater than 9.1, so 9.9 is larger. I'll keep it clear and concise.

Done

Q: Which is larger, 9.9 or 9.10?

A: 9.9

Ask anything

Thinking

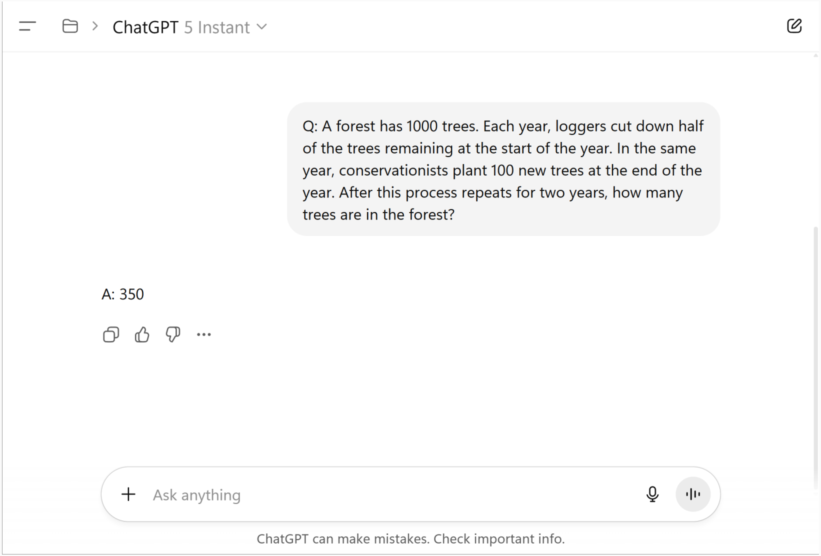
ChatGPT can make mistakes. Check important info.

Another example

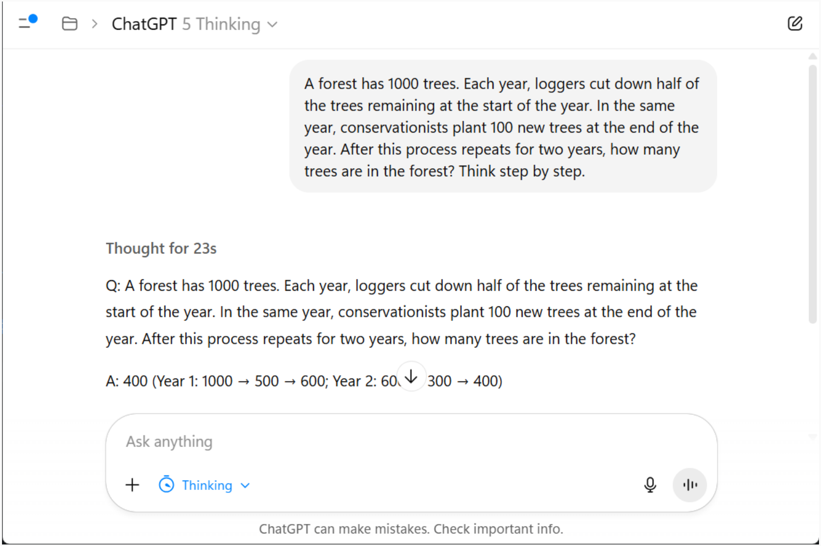
hmm... Let's try another question.

Question: *A forest has 1000 trees. Each year, loggers cut down half of the trees remaining at the start of the year. In the same year, conservationists plant 100 new trees at the end of the year. After this process repeats for two years, how many trees are in the forest?*

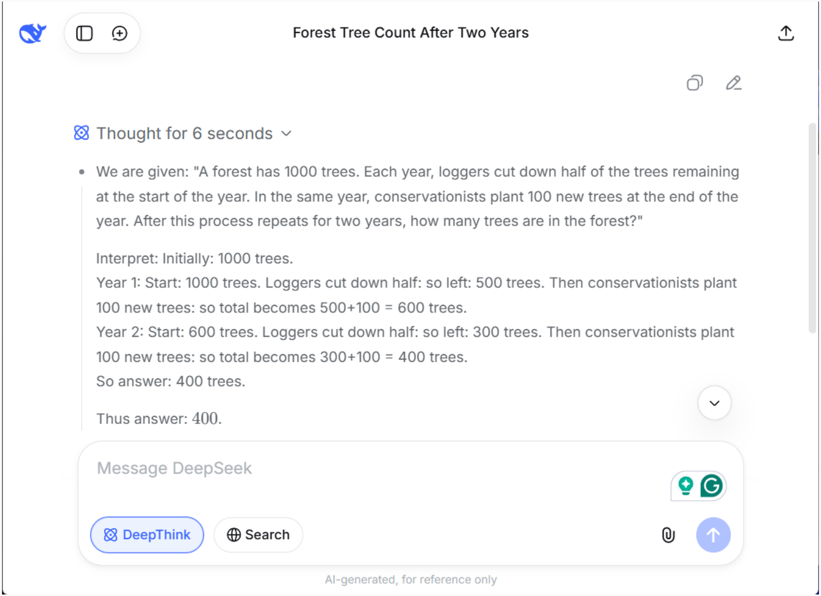
ChatGPT 5 Instant ✗



ChatGPT 5 Thinking ✓



What the CoT trace looks like (Courtesy of DeepSeek)



One last example

Okay... One more...

Question: *Suppose* $x + 2y - z = 4$*,* $-3x - y + z = 10$*,* $4x + y - 2z = 15$*. Solve for* x, y, z*.*

ChatGPT 5 Instant ✗

ChatGPT 5 Instant

Q: Suppose x +2y - z = 4, -3x - y +z = 10, 4x +y -2z = 15. Solve for x, y, z

A: $x = 3,\ y = 2,\ z = 4$

Ask anything

ChatGPT can make mistakes. Check important info.

ChatGPT 5 Thinking ✓

ChatGPT 5 Thinking

Now use (3): $4x + y - 2z = 15$.

$$4x + (2x + 14) - 2(5x + 24) = 15 \Rightarrow 6x + 14 - 10x - 48 = 15 \Rightarrow -4x - 34 = 15 \Rightarrow -4x$$

Then

$$y = 2x + 14 = 2\left(-\frac{49}{4}\right) + 14 = -\frac{98}{4} + \frac{56}{4} = -\frac{42}{4} = -\frac{21}{2},$$

$$z = 5x + 24 = 5\left(-\frac{49}{4}\right) + 24 = -\frac{245}{4} + \frac{96}{4} = -\frac{149}{4}.$$

$$\boxed{x = -\frac{49}{4},\ y = -\frac{21}{2},\ z = -\frac{149}{4}}$$

Ask anything

Thinking

ChatGPT can make mistakes. Check important info.

What's going on here?

- ChatGPT Instant is attempting to produce the answer immediately in one step.
- But… robustly computing the answers to these questions involves a multi-step process.
- CoT training provides a form of *intermediate supervision*: learning the "process" not just the "answer"
- ChatGPT Thinking carries out the computation *step-by-step*, emulating a learned problem-solving *procedure*, enabling it to arrive at the correct answer.

What's Going on Here?

- **Observation:** CoT supervision improves accuracy on reasoning tasks.
- **Question:** How can we formalize and measure the *statistical advantage* from CoT?
- **Next:** A learning-theoretic framework capturing this advantage.

Central questions

1. When does CoT reduce sample complexity vs. end-to-end labels?
2. How do we *quantify* the added information in CoT traces?
3. What upper bounds/lower bounds govern CoT learning?

Answers to these questions in...

CoT Information: Improved Sample Complexity under Chain-of-Thought Supervision

Awni Altabaa[1] **Omar Montasser**[2] **John Lafferty**[3]

June 10, 2025

Abstract: Learning complex functions that involve multi-step reasoning poses a significant challenge for standard supervised learning from input-output examples. Chain-of-thought (CoT) supervision, which provides intermediate reasoning steps together with the final output, has emerged as a powerful empirical technique, underpinning much of the recent progress in the reasoning capabilities of large language models. This paper develops a statistical theory of learning under CoT supervision. A key characteristic of the CoT setting, in contrast to standard supervision, is the mismatch between the training objective (CoT risk) and the test objective (end-to-end risk). A central part of our analysis, distinguished from prior work, is explicitly linking those two types of risk to achieve sharper sample complexity bounds. This is achieved via the *CoT information measure* $\mathcal{I}^{\mathrm{CoT}}_{\mathcal{D},h_\star}(\varepsilon;\mathcal{H})$, which quantifies the additional discriminative power gained from observing the reasoning process. The main theoretical results demonstrate how CoT supervision can yield significantly faster learning rates compared to standard E2E supervision. Specifically, it is shown that the sample complexity required to achieve a target E2E error ε scales as $d/\mathcal{I}^{\mathrm{CoT}}_{\mathcal{D},h_\star}(\varepsilon;\mathcal{H})$, where d is a measure of hypothesis class complexity, which can be much faster than standard d/ε rates. Information-theoretic lower bounds in terms of the CoT information are also obtained. Together, these results suggest that CoT information is a fundamental measure of statistical complexity for learning under chain-of-thought supervision.

To be presented at NeurIPS '25

Collaborators

Omar Montasser

Yale

John Lafferty

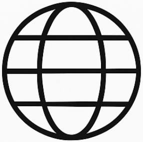
Yale

Plan

- **Part I:** Chain-of-Thought in LLMs, in practice
- **Part II:** Formalizing CoT-Supervised learning
- **Part III:** Upper Bounds + "CoT Information"
- **Part IV:** Information-theoretic Lower Bounds
- **Part V:** Simulations

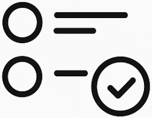
Chain-of-Thought in Large Language Models

A Hitchhiker's Guide to Building an LLM



Pre-training

Supervised Fine-Tuning



Post-Training

Step 1: Pre-training (foundation modeling).

Goal: Learn broad world knowledge and language ability

Data + Objective: Next-token prediction objective on internet-scale text.

Step 2: Supervised Fine-Tuning.

Goal: Teach model to *follow instructions* and perform tasks step-by-step.

Data + Objective: Smaller scale but high-quality human-written demonstrations in Q/A format; includes chain-of-thought style rationales*

*sometimes hidden at inference, depending on model

Step 3: Post-Training (Preference Optimization & Alignment)

Goal: align outputs to human preferences, enforce safety policies, enable tool-use and function-calling

Data + Objective: varies; includes human preferences (RLHF/DPO), SFT on tool-calling traces, synthetic data, ...

Our Focus

We will focus on the CoT supervision during SFT

what statistical advantage does it provide?

How to obtain CoT traces to train on?

[redacted] • 1:33 PM

Incredible opportunity with Amazon!

Hi Awni ,

Amazon is working on a number of exciting machine learning projects aimed at building out and improving our AI models, and we are looking for qualified experts like yourself to support us in our work! By combining human intelligence with sophisticated machine learning algorithms, we aim for our model's data in **Mathematics** to be meticulously annotated and labeled, capturing nuances that machines alone might miss. This blend of human insight and technological precision helps to ensure a responsible approach to training our AI models that truly reflects real-world complexities and subtleties.

What you can expect:

- **Flexibility:** You can opt to complete as many, or as few tasks as you prefer. Complete tasks remotely from anywhere!
- **Compensation:** Earn $10-$25 per task, with each task taking 10-20 minutes—equivalent to $40-$75 per hour. Choose your payment schedule: daily, weekly, or bi-weekly!
- **Cutting-Edge Projects:** You will play a part in shaping the future of AI!

Qualifications:

- Bachelor's Degree
- Fluent in the English language
- Must be US Citizen or Green Card to participate

[redacted]

AI | Live Events | Strategy | Writer

Sponsored Oct 9

[redacted]

Apply your Stats expertise— $85+/hr flexible, remote work

Hi Awni!

I'm Alex with the Handshake AI team. I'm reaching out because I think your Statistics expertise would make you a great candidate for our Stats expert position with the Handshake AI Fellowship program.

We partner with leading AI labs to advance how AI understands specialized fields. As a Stats expert, you'll help AI interpret complex data, improve predictive models, and strengthen analytical reasoning.

What to know:

- Compensation begins at **$85+/hour,** with potential for increases
- You work remotely and **entirely on your own schedule**
- **OPT/CPT Eligible**, see program details for conditions.
- The **application takes about 10 minutes,** followed by training and practice exercises
- Once approved, you'll join a priority pool considered for **ongoing and upcoming AI lab projects**

This is paid, flexible work that applies your academic training while building your professional profile.
If you're curious, you can **apply here**.

Best regards,
[redacted]
Handshake AI

[redacted]

[redacted] (He/Him) • 4:28 AM

Mercor: Partnering with Elite STEM Talent

Hey Awni

We are currently running a project with one of the top AI Labs where we would love to have candidates with a PhD in STEM working with us. Your LinkedIn profile is incredible so I wanted to send an email to see if you would are open to new opportunities. You can learn more and apply on our platform here

Screening Process: After submitting your resume, you will complete a brief AI interview (that is the only step in the process).

Pay for the role is: **55-90 USD/hr.** t is **part time (10-20 hrs/ week), completely remote** and **asynchronous.** We have already started on the project and expanding so the timeline is pretty immediate

About Mercor: We're a staffing company based in San Francisco, CA valued at $2B and backed by investors like Benchmark, General Catalyst, Peter Thiel etc. We've paid out 12 million dollars to over 3k people and specialize in recruiting talented experts for leading tech companies.

Thanks

[redacted]
Technical Project Manager @ Mercor

Ask FT

FINANCIAL TIMES

My Account

Artificial intelligence + Add to myFT

AI groups spend to replace low-cost 'data labellers' with high-paid experts

Industry shifts from paying gig economy workers in Africa and Asia in push to build 'smarter' models

Share Save

Melissa Heikkilä in London

Published JUL 20 2025

113

Top artificial intelligence groups are replacing low-cost "data labellers" in Africa and Asia with highly paid industry specialists, in the latest push to build "smarter" and more powerful models.

Companies such as Scale AI, Turing and Toloka are hiring experts in fields such as biology and finance to help AI groups create more sophisticated training data that is crucial for developing the next generation of AI systems.

The rise of so-called "reasoning" models, such as OpenAI's o3 and Google's

How to get a dataset of CoT Traces?

- Human-authored : e.g., domain experts or trained annotators; or mined from educational content (e.g., step-by-step solutions in textbooks)
- Model-generated : traces generated by prompting, filtered via self-consistency or self-verification
- Hybrid : model generates, humans filter/rate/edit
- Programmatic synthesis : for certain tasks, it may be possible to generate traces programmatically

Cobbe et al. [arXiv:2110.14168], Wang et al. [arXiv:2203.11171], Zelikman et al. [arXiv:2203.14465], Lightman et al. [arXiv:2305.20050]

Our Model of CoT

For us: we will abstract away from the specific method of obtaining CoT traces or their format (e.g., natural language or otherwise)

We assume the CoT dataset exists, and ask:

What statistical advantage does training with CoT confer compared to traditional supervised learning from input-output examples?

Formalizing Learning with CoT Supervision

Two Aspects of Chain-of-Thought Theory

1. Function Approximation — Expanded representational capacity
    - lots of work here; pretty well-understood*
2. Statistical — More rapid & sample-efficient learning
    - less understood; *our focus for this work*

* e.g., Pérez et al. [arXiv:1901.03429], Merrill & Sabharwal [arXiv:2310.07923], Li et al. [arXiv:2402.12875]

Traditional Supervised Learning

Want to learn $h_\star : \mathcal{X} \to \mathcal{Y}$ in some function class $\mathcal{H} \subset \mathcal{Y}^\mathcal{X}$

Observe labeled examples (x_i, y_i) from random inputs $x_i \overset{\text{i.i.d.}}{\sim} \mathcal{D}$, with label $y_i = h_\star(x_i)$

Learning algorithm maps training sample to predictor:

$$\mathcal{A} : S = \{(x_i, y_i)\}_{i=1}^m \mapsto \hat{h}$$

Goal: achieve small prediction error

$$\mathcal{R}(\hat{h}) := \mathbb{P}_{x \sim \mathcal{D}} \left[\hat{h}(x) \neq y \right] \leq \varepsilon$$

consider the realizable setting for now

Basic intuition: standard end-to-end learning

Suppose we want to distinguish two hypotheses $\mathcal{H} = \{h_1, h_2\}$.

We can distinguish them once we observe x such that $h_1(x) \neq h_2(x)$.

If $\underset{x \sim \mathcal{D}}{\mathbb{P}} [h_1(x) \neq h_2(x)] = \varepsilon$, this will take

$$\mathcal{O}\left(\frac{1}{\varepsilon}\right)$$

samples.

Sample Complexity in Classical Statistical Learning

Extend to larger classes (> 2 hypotheses) via union bound

Gives sample complexity $\mathcal{O}(\log |\mathcal{H}| / \varepsilon)$, scaling with class size/complexity $|\mathcal{H}|$ and error parameter ε.

(Or: VC dimension, Rademacher, Covering Numbers, etc.)

We will be mainly interested in the ε-dependence in the denominator, interpreting it as "*the amount of information per observed sample*"

---

e.g., under noise (agnostic setting), with less "information" per sample, the rate is $1/\varepsilon^2$; under low-noise conditions, we interpolate between $1/\varepsilon$ and $1/\varepsilon^2$

Challenges of Classical (end-to-end) Learning for Long-Form Reasoning Problems

- What if we want to learn very complicated functions $h_\star$?
- E.g., w/ modern LLMs, want to learn long-form multi-step reasoning: mathematical reasoning, coding, etc...
- Statistically very difficult to learn these complex multi-step functions
- Observing input-output examples (x_i, y_i) alone reveals relatively little information about the function $h_\star$

Idea: Train with a stronger signal

- Provide additional information to the learner
- Not only input-output pairs, but also expose step-by-step computational trace
- Reveals more information that allows identifying function to be learned faster

Intuition (Preview): Why this enables more rapid learning?

Suppose instead that h_1, h_2 emit two observable outputs:

$$y = h^{\text{e2e}}(x) \quad \textit{AND} \quad z = h^{\text{CoT}}(x)$$

Then, h_1, h_2 can be distinguished if we observe x where *either* $h_1^{\text{e2e}}(x) \neq h_2^{\text{e2e}}(x)$ *OR* $h_1^{\text{CoT}}(x) \neq h_2^{\text{CoT}}(x)$.

It's now easier to distinguish the two hypotheses (possibly *much* easier).

Let $I(\varepsilon) := \mathbb{P}_{x \sim \mathcal{D}} \left[h_1^{\text{e2e}}(x) \neq h_2^{\text{e2e}}(x) \text{ OR } h_1^{\text{CoT}}(x) \neq h_2^{\text{CoT}}(x)\right] \geq \varepsilon$.

Now only need $\mathcal{O}\left(1/I(\varepsilon)\right)$ samples.

CoT Hypothesis Class

We adopt an abstract definition of CoT hypothesis classes.

CoT Hypothesis Class

A family $\mathcal{H} \subset (\mathcal{Y} \times \mathcal{Z})^{\mathcal{X}}$ of functions $h : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$, where $y \in \mathcal{Y}$ is the final output and $z \in \mathcal{Z}$ is the CoT.

$h^{\mathrm{e2e}} : \mathcal{X} \to \mathcal{Y}$: end-to-end restriction of h

$h^{\mathrm{CoT}} : \mathcal{X} \to \mathcal{Z}$: CoT restriction of h

For example, $\mathcal{H}$ can be a class of sequence models (e.g., Transformers) that output a CoT trace $\boldsymbol{z} = (z_1, \ldots, z_{t_h(x)})$, followed by the final output y.

PAC Learning under CoT Supervision

CoT learning algorithm

$$\mathcal{A} : (\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})^* \to \mathcal{Y}^{\mathcal{X}}, \quad S = \{(x_i, y_i, z_i)\}_{i=1}^m \mapsto \hat{h}$$

We say $\mathcal{A}$ has sample complexity $m_{\mathcal{H},\mathcal{D}}(\varepsilon, \delta)$ if

$$m \geq m_{\mathcal{H},\mathcal{D}}(\varepsilon, \delta) \implies \mathop{\mathbb{P}}_{S \sim \mathcal{D}^m} \left[\mathop{\mathbb{P}}_{x,y \sim \mathcal{D}_{x,y}} [\mathcal{A}(S)(x) \neq y] \leq \varepsilon \right] \geq 1 - \delta$$

CoT learning algorithm observes (x, y, z): input, output, *& CoT*

Evaluation metric is *end-to-end error*

Statistical Guarantees for Learning under CoT Supervision

Goal: obtain statistical guarantees that characterize the advantage of training with CoT supervision

Central Challenge

The central challenge is the asymmetry between the training objective and the testing objective.

Training Objective: CoT Risk

$$\mathcal{R}_{\mathcal{D}}^{\mathrm{CoT}}(h) := \underset{x,y,z\sim\mathcal{D}}{\mathbb{P}}\left[(h^{\mathrm{e2e}}(x), h^{\mathrm{CoT}}(x)) \neq (y, z)\right]$$

Test Objective: End-to-End Risk

$$\mathcal{R}_{\mathcal{D}}^{\mathrm{e2e}}(h) := \underset{x,y\sim\mathcal{D}_{x,y}}{\mathbb{P}}\left[h^{\mathrm{e2e}}(x) \neq y\right]$$

Asymmetry prevents direct application of standard learning theory results.

Possible approach: bound CoT risk instead

Recent related work by Joshi et al. (2025) `[arXiv:2503.07932]` sidestep asymmetry by noting that

$$\begin{aligned}\mathcal{R}_{\mathcal{D}}^{\mathrm{CoT}}(h) &:= \mathop{\mathbb{P}}_{x,y,z\sim\mathcal{D}}\left[(h^{\mathrm{e2e}}(x), h^{\mathrm{CoT}}(x)) \neq (y,z)\right] \\ &\leq \mathop{\mathbb{P}}_{x,y\sim\mathcal{D}_{x,y}}\left[h^{\mathrm{e2e}}(x) \neq y\right] =: \mathcal{R}_{\mathcal{D}}^{\mathrm{e2e}}(h), \forall h \in \mathcal{H}\end{aligned}$$

Approach: establish a guarantee on the *CoT risk* instead of the end-to-end risk

This problem is now symmetric because both training and testing objective are CoT error; can apply standard learning theory results

Joshi et al.'s Sample Complexity

This implies a sample complexity of

$$\mathcal{O}\left(\frac{\mathrm{VC}(\mathcal{L}^{\mathrm{CoT}}(\mathcal{H}))}{\varepsilon}\right)$$

for learning with an error $\leq \varepsilon$.

$\mathcal{L}^{\mathrm{CoT}}(\mathcal{H})$ is the CoT loss class

$$\mathcal{L}^{\mathrm{CoT}}(\mathcal{H}) := \left\{\ell_h^{\mathrm{CoT}} : (x, y, z) \mapsto \mathbf{1}\{h(x) \neq (y, z)\} : h \in \mathcal{H}\right\}$$

The main technical innovation of Joshi et al. (2025) is bounding $\mathrm{VC}(\mathcal{L}^{\mathrm{CoT}}(\mathcal{H}))$ for "autoregressive" or "iterated" function classes.

Limitations of this analysis

- This gives the same $\mathcal{O}(1/\varepsilon)$ rate as in standard end-to-end supervision
- Intuitively, observing the CoT z provides more *information* and should enable *faster statistical rates* .
- Need a more refined analysis that explicitly *links* the two different risks and *quantifies* the “information content” per observed CoT sample

We need to explicitly capture the *amount of information* encoded in the CoT for *distinguishing hypotheses in the hypothesis class*.

New Concept: Chain-of-Thought Information

For a CoT hypothesis class $\mathcal{H} \subset (\mathcal{Y} \times \mathcal{Z})^{\mathcal{X}}$ and distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ realizable by $\mathcal{H}$, the CoT information is the function

$$\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H}) := \inf_{h \in \Delta_{\mathcal{D}}^{\mathrm{e2e}}(\varepsilon; \mathcal{H})} \left\{ -\log \mathop{\mathbb{P}}_{x,y,z \sim \mathcal{D}} \left[(h^{\mathrm{e2e}}(x), h^{\mathrm{CoT}}(x)) = (y, z) \right] \right\},$$

where

$$\Delta_{\mathcal{D}}^{\mathrm{e2e}}(\varepsilon; \mathcal{H}) := \left\{ h \in \mathcal{H} : \mathop{\mathbb{P}}_{x,y} \left[h^{\mathrm{e2e}}(x) \neq y \right] > \varepsilon \right\}$$

The CoT Information captures the statistical rate of CoT learning

Properties of the CoT Information

- $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H}) \geq \varepsilon$
- $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})$ is monotonically increasing in ε
- $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})$ is monotonically decreasing in $\mathcal{H}$

CoT Information: Intuition

- When the CoT information is large, observing the CoT reveals additional information about the target hypothesis
- Alternative hypotheses can't "fake" the reasoning steps in the target's CoT
- A CoT sample is more valuable than an end-to-end sample

Key result: CoT information captures the statistical rate of CoT Learning

Learning under CoT supervision

The CoT consistency rule

$\texttt{CoT-Cons}(S;\mathcal{H}) := \{h \in \mathcal{H} : h^{\mathrm{e2e}}(x_i) = y_i, h^{\mathrm{CoT}}(x_i) = z_i, \forall i\}$

has end-to-end error sample complexity

$$m(\varepsilon, \delta) = \frac{\log|\mathcal{H}| + \log(1/\delta)}{\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})}.$$

I.e., $m \geq m(\varepsilon, \delta)$ implies that w.p. $\geq 1 - \delta$ over $S \sim \mathcal{D}^m$

$$\forall h \in \texttt{CoT-Cons}(S;\mathcal{H}),\ \mathcal{R}_{\mathcal{D}}^{\mathrm{e2e}}(h) = \mathbb{P}_{x,y}\left[h^{\mathrm{e2e}} \neq y\right] \leq \varepsilon$$

Take away

$1/\varepsilon$ rate improved to potentially much faster $1/\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon;\mathcal{H})$

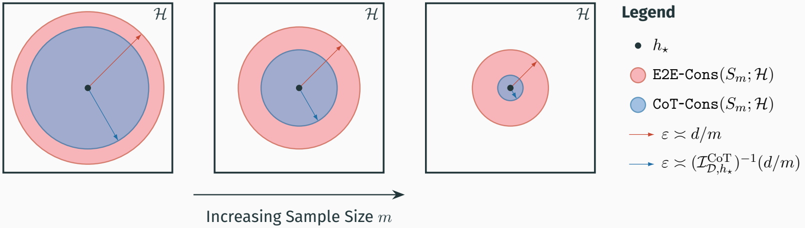
The ratio

$$\frac{\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon;\mathcal{H})}{\varepsilon}$$

can be interpreted as

how many end-to-end samples is one CoT sample worth?

Geometry of CoT Information



Infinite Classes

The result can be extended to infinite classes

Learning Infinite Classes with CoT Supervision

For infinite classes, the CoT consistency rule

$\texttt{CoT-Cons}(S;\mathcal{H}) := \{h \in \mathcal{H} : h^{\mathrm{e2e}}(x_i) = y_i, h^{\mathrm{CoT}}(x_i) = z_i, \forall i\}$

has end-to-end error sample complexity

$$m(\varepsilon,\delta) = \mathcal{O}\left(\frac{\mathrm{VC}(\mathcal{L}^{\mathrm{CoT}}(\mathcal{H}))\cdot\log\left(\frac{1}{\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon;\mathcal{H})}+1\right)+\log(1/\delta)}{\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon;\mathcal{H})}\right).$$

Recall, $\mathcal{L}^{\mathrm{CoT}}(\mathcal{H})$ is the *CoT loss class*.

Preview: the agnostic setting

So far in the talk, we've addressed the *realizable* setting (i.e., distribution $\mathcal{D}$ over (x, y, z) is realizable by $\mathcal{H}$)

What if $\mathcal{D}$ is not realizable by $\mathcal{H}$?

Preview: the agnostic setting

In general, CoT supervision can be detrimental in the agnostic setting (unlike the realizable).

Example

Let $\mathcal{D}$ be a distribution over $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ such that the output component is realizable by $\mathcal{H}$ but the CoT component is not realizable. In particular, let $\mathcal{D}$ be such that $\inf_{h \in \mathcal{H}} \mathcal{R}^{\mathrm{e2e}}_{\mathcal{D}}(h) = 0$ but $\inf_{h \in \mathcal{H}} \mathcal{R}^{\mathrm{CoT}}_{\mathcal{D}}(h) = 1$. Then, `CoT-ERM` provides no guarantees since $\texttt{CoT-ERM}(S; \mathcal{H}) = \mathcal{H}, \forall S$. By contrast, end-to-end learning is possible with sample complexity $\mathcal{O}(1/\varepsilon \cdot \mathrm{VC}(\mathcal{L}^{\mathrm{e2e}}(\mathcal{H})))$.

Preview: Extension of the CoT Information to the Agnostic Setting

To extend our results to the agnostic setting, we define a variant of the CoT information that links the *excess* CoT risk with the *excess* end-to-end risk.

This variant of the CoT information measures how *aligned* the observed CoT in the data distribution is with the hypothesis class.

For more on this, please see paper

Information-Theoretic Lower Bounds

So far: statistical *upper* bounds — “CoT supervision enables learning *at least* `[this fast]`”

Answer expressed in terms of *CoT Information*

To really determine if the CoT information measure is fundamentally the correct characterization, we also need corresponding *lower* bounds.

I.e., “Learning with CoT supervision requires *at least* `[this many]` samples”

Lower Bound: 2-Point Method

Fix a distribution $\mathcal{D}$ over $\mathcal{X}$. Let $x_1, \ldots, x_m \overset{\text{i.i.d.}}{\sim} \mathcal{D}$. For any $h_\star \in \mathcal{H}$, if the sample size satisfies

$$m < \frac{\log(1/\delta)}{\mathcal{I}^{\mathrm{CoT}}_{\mathcal{D},h_\star}(\varepsilon; \mathcal{H})},$$

then with probability at least δ there exists $h \in \mathcal{H}$ with end-to-end error at least ε which is indistinguishable from $h_\star$ on the sample. Moreover, for any algorithm $\mathcal{A}$,

$$\sup_{h_\star \in \mathcal{H}} \mathop{\mathbb{E}}_{S \sim P^{\otimes}_{h_\star}} \left[\mathcal{R}^{\mathrm{e2e}}_{\mathcal{D}}(\mathcal{A}(S))\right] \geq \frac{1}{2} \sup_{\substack{h_\star \in \mathcal{H} \\ \varepsilon > 0}} \varepsilon \cdot \exp(-m \cdot \mathcal{I}^{\mathrm{CoT}}_{\mathcal{D},h_\star}(\varepsilon; \mathcal{H})).$$

Idea: reduce to binary hypotheses testing (LeCam, 1973).

This lower bound exhibits the expected scaling wrt the error parameter ε: CoT information $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})$ appears

But... doesn't scale with $\mathrm{size}(\mathcal{H})$: limitation of 2-point method

Next: a lower bound that scales with $\mathrm{size}(\mathcal{H})$ via Fano's method

Lower Bound: Fano's Method

Fix a distribution $\mathcal{D}$ over $\mathcal{X}$. Let $x_1, \ldots, x_m \overset{\text{i.i.d.}}{\sim} \mathcal{D}$. For any algorithm $\mathcal{A}$, if the sample size satisfies

$$m \leq \frac{\log M(\mathcal{H}; d_{\mathcal{D}}^{\mathrm{e2e}}, \varepsilon)}{2 \cdot \left(C_Q \cdot \sup_\pi \underset{h_1, h_2 \sim \pi}{\mathbb{E}} \left[\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(h_1, h_2) \right] + \log 2 \right)}$$

then the end-to-end error must be large for some $h_\star$, i.e.,

$$\sup_{h_\star \in \mathcal{H}} \underset{S \sim P_{h_\star}^{\otimes}}{\mathbb{P}} \left[\mathcal{R}_{\mathcal{D}, h_\star}^{\mathrm{e2e}}(\mathcal{A}(S)) \geq \varepsilon/2 \right] \geq \frac{1}{2}$$

Idea: pack $\mathcal{H}$ w.r.t. end-to-end metric; use Fano's inequality and relate MI to CoT information.

Take away

Lower bounds validate *CoT Information* as a fundamental measure of statistical complexity in *learning with CoT supervision*.

Simulations

How well does the CoT Information theory capture the statistical advantage of training with CoT supervision in practice?

Basic Experimental Set Up

- Consider some CoT hypothesis class $\mathcal{H} \subset (\mathcal{Y} \times \mathcal{Z})^{\mathcal{X}}$
- Fix a data distribution $\mathcal{D}$ and a reference hypothesis $h_\star$
- Compute the CoT information function $\mathcal{I}^{\mathrm{CoT}}_{\mathcal{D},h_\star}(\varepsilon; \mathcal{H})$
- Run simulations to evaluate the statistical rates of learning with `E2E-Cons` (end-to-end supervision) *vs.* `CoT-Cons` (CoT consistency)

Does $\mathcal{I}^{\mathrm{CoT}}_{\mathcal{D},h_\star}(\varepsilon; \mathcal{H})$ accurately capture the statistical advantage?

Here, we consider finite $\mathcal{H}$ and compute everything exactly.

Example 1: Linear Autoregressive Model

In practice, CoT supervision implemented via sequence model class (e.g., Transformers) trained to generate CoT as a sequence token-by-token, before returning final output.

Consider a simple form of such autoregressive generation: generate tokens autoregressively for T steps by linear function of fixed-size window of history

Example 1: Linear Autoregressive Model

$\mathcal{X} = \Sigma^*, \mathcal{Z} = \Sigma^T, \mathcal{Y} = \Sigma$. Take $\Sigma = \{0, 1\}$.

Linear next-token generation:

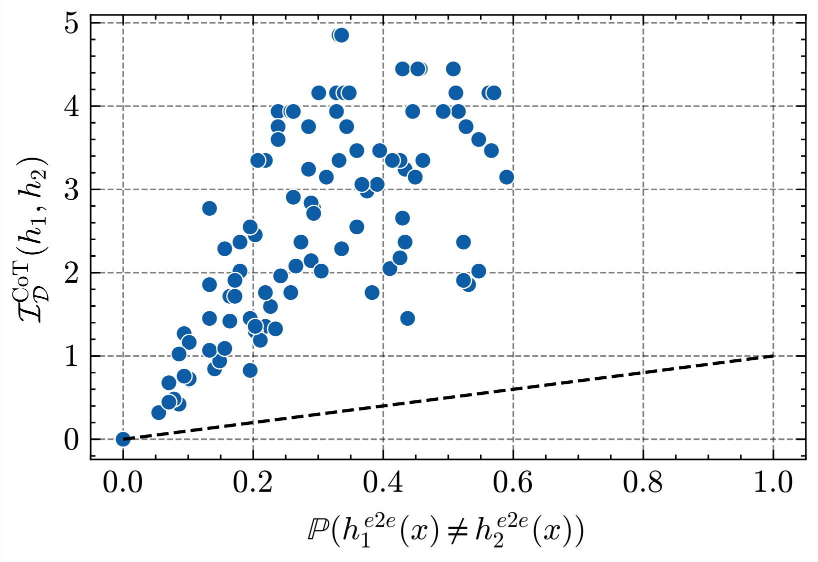
$$f_w : \boldsymbol{x} \mapsto (\boldsymbol{x}, z) \in \{0,1\}^{|\boldsymbol{x}|+1},\ z = \mathbf{1}\left\{\sum_{i=0}^{d-1} w_i \cdot x_{n-i} \geq 0\right\}.$$

Hypothesis class $\mathcal{H} = \{h_w = (h_w^{\mathrm{CoT}}, h_w^{\mathrm{e2e}}) : w \in \mathcal{W}\}$, where

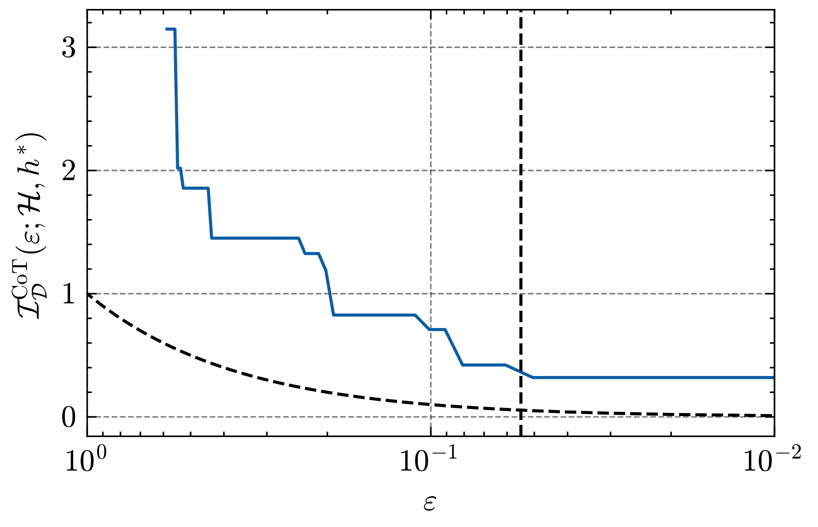
$$h_w^{\mathrm{CoT}} : \boldsymbol{x} \mapsto (z_1, ..., z_T),\ h_w^{\mathrm{e2e}} : \boldsymbol{x} \mapsto z_T$$

$$(\boldsymbol{x}, (z_1, \ldots, z_T)) = \underbrace{(f_w \circ \cdots \circ f_w)}_{T \text{ times}}(\boldsymbol{x}).$$

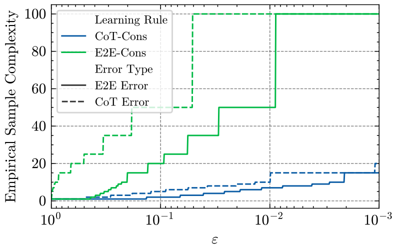
Window size $d = 8$, # iterations $T = 16$, weight class $\mathcal{W} = \{-1, 0, +1\}^d$.



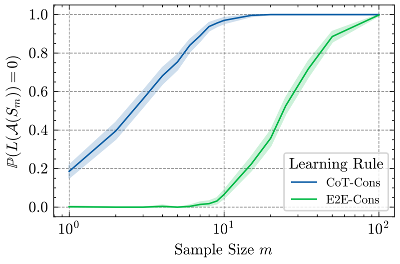
Relative CoT information between pairs of hypotheses vs. end-to-end disagreement.



CoT information $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})$ as a function of ε. Dashed line is ε baseline. Vertical dashed line is $\varepsilon^* := \min\{\mathcal{R}_{\mathcal{D}}^{\mathrm{e2e}}(h) : \mathcal{R}_{\mathcal{D}}^{\mathrm{e2e}}(h) > 0\}$.



(a) Empirical sample complexity.



(b) Empirical probability of each learning rule returning a predictor with zero error.

Take away

$\lim_{\varepsilon \to 0} \mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})/\varepsilon^{+} \approx 6$, suggesting a $6\times$ gain in sample-efficiency.

Empirical sample complexity indicates a $\approx 5\times$ advantage for CoT supervision.

Example 2: Learning Regular Languages

Let $\mathcal{L}$ be a regular language.

End-to-end function is membership in $\mathcal{L}$

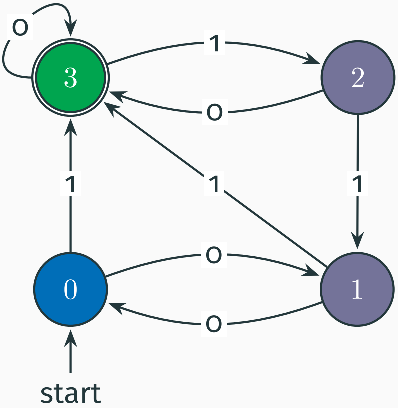
$$h^{\mathrm{e2e}}(x) = \begin{cases} 1, & \text{if } x \in \mathcal{L} \\ 0, & \text{otherwise} \end{cases}$$

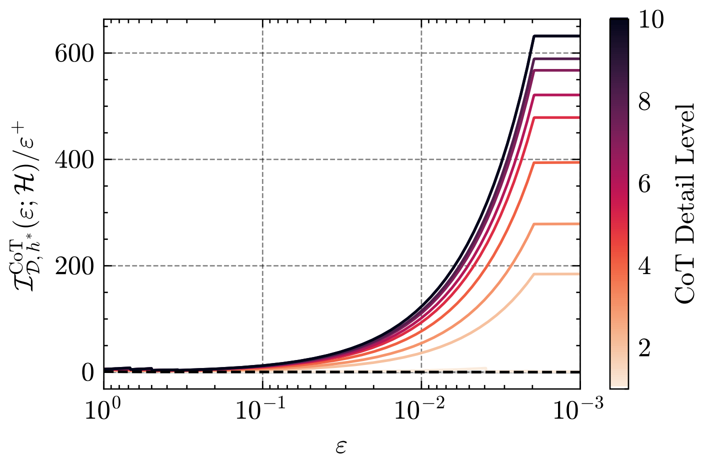
Chain-of-thought is sequence of states for DFA recognizing $\mathcal{L}$

$$h^{\mathrm{CoT}}(x) = (z_1, \ldots, z_n),$$

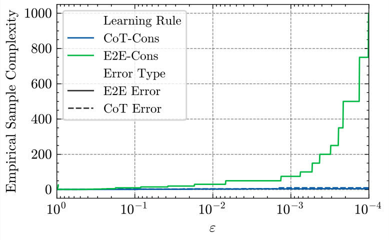
where DFA has state transitions $z_t \xrightarrow{x_t} z_{t+1}$.

Example 2: Learning regular languages

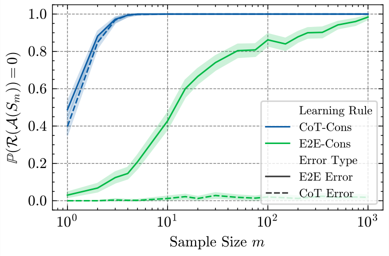




CoT information ratio $\mathcal{I}^{\mathrm{CoT}}_{\mathcal{D},h_\star}(\varepsilon;\mathcal{H})/\varepsilon^+$, varying level of CoT detail (how much of the state trajectory is observed).



(a) Empirical Sample Complexity.



(b) Empirical probability of each learning rule returning a predictor with zero error.

Take away

$\lim_{\varepsilon \to 0} \mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})/\varepsilon^{+} \approx 600$, suggesting a $600\times$ gain in sample-efficiency.

Empirical sample complexity gains of $10^2 - 10^3\times$ for CoT compared to E2E.

Extensions: Applying CoT Information to Complex Infinite Classes

Informal

The CoT information $\mathcal{I}_{\mathcal{D}}^{\mathrm{CoT}}(\varepsilon; \mathcal{H})$ can be estimated consistently from random samples uniformly in $\varepsilon \in (0, 1)$.

Implication: We can estimate the CoT information for complex classes (e.g., neural networks) by setting up an appropriate optimization problem on random CoT samples.

This provides a means of *quantifying* the value of CoT supervision in different model classes.

For example, compare different methods of generating CoT traces to train on.

Concluding Remarks

Summary & Concluding Remarks

- Extension of statistical learning theory to CoT settings
- Information-theoretic characterization of the statistical advantage of CoT supervision
- Many open questions and directions for future work: role of CoT in OOD generalization, use of RL to learn CoT reasoning, scaling, optimal CoT representation, etc.

Thank You

Discussion Time...

- Joint work with *Omar Montasser* & *John Lafferty*
- Supported by: ARNI NSF AI Institute
- Paper: `arXiv:2505.15927` / NeurIPS ’25
- Personal webpage: `https://awni.xyz`