

MARKOVIAN TRANSFORMERS FOR INFORMATIVE LANGUAGE MODELING

Anonymous authors

Paper under double-blind review

ABSTRACT

Chain-of-Thought (CoT) reasoning often fails to faithfully reflect a language model’s underlying decision process. We address this by introducing a *Markovian* language model framework that can be understood as a reasoning autoencoder: it creates a text-based bottleneck where CoT serves as an intermediate representation, forcing the model to compress essential reasoning into interpretable text before making predictions. We train this system using a policy gradient algorithm inspired by Group Relative Policy Optimization (GRPO), with parallel sampling and actor reward gradients derived from the constraint that our reward model uses the same parameters θ as the policy model. This approach achieves a 33.2% absolute accuracy improvement on GSM8K with Llama 3.1 8B. Comprehensive perturbation analysis across 5,888 comparison points and four model pairs demonstrates that Markovian training produces systematically higher sensitivity to CoT perturbations (effect differences +0.0154 to +0.3276), with 52.9%–87.6% consistency in showing greater fragility compared to Non-Markovian approaches. Cross-model evaluation confirms that learned CoTs generalize across architectures, indicating they capture transferable reasoning patterns rather than model-specific artifacts.

1 INTRODUCTION

The rapid advancement of language models (LMs) has led to impressive performance on complex cognitive tasks (?). Yet it is often unclear *why* an LM arrives at a particular conclusion (???), causing issues in high-stakes applications (???). Traditional interpretability methods analyze hidden activations or attention patterns to extract “explanations” (???????). Modern LMs, however, already generate coherent text: we might hope *prompting* the model to articulate its reasoning (“Chain-of-Thought” or CoT) (??) would yield a faithful record of its thought process.

Unfortunately, CoT explanations can be *unfaithful*. For example, ? show that spurious in-context biases often remain hidden in the CoT, and ? find that altering CoT text may not affect the final answer. Such observations indicate that standard CoTs are not “load-bearing.”

In this work, we take a *pragmatic* approach to interpretability, focusing on *informativeness* over full faithfulness. Rather than insisting the CoT mirrors the model’s entire internal process, we require that *the CoT alone suffices to produce the final answer*. In other words, if we remove the original prompt and rely only on the CoT, the model should still reach the correct output. This makes the CoT *causally essential* and *fragile*: changing it necessarily alters the prediction.

What distinguishes our approach is the clear distinction between the model *relying on its CoT* versus generating *more informative CoTs*. While traditional approaches train models to generate better-quality CoTs, they don’t fundamentally change how the model uses them. Our Markovian framework, by contrast, forces the model to process information through the CoT bottleneck, making the CoT not just informative but *causally load-bearing* for prediction.

For instance, Mistral-7B’s CoT on arithmetic tasks changed dramatically after training. **Before training**, it simply listed all numbers and their (incorrect) sum (e.g., “Sum = 76 + 90 + 92 + ... = 2314”). **After training**, it performed correct step-by-step calculations (e.g., “calculate 6 + 89 = 95; Next, calculate 95 + 38 = 133...”), breaking the task into manageable steps that can be verified independently and enabling accurate answer prediction even when the original question is removed.

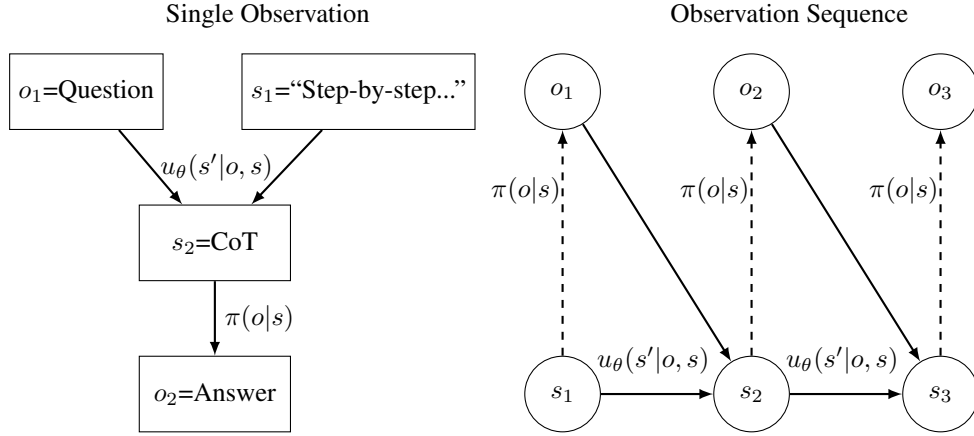


Figure 1: Markovian training as a reasoning autoencoder. Left: Single time-step process from Question to CoT to Answer, creating a text-based bottleneck where the CoT must capture all information needed for answer prediction. Right: Causal structure showing the generation of states from observations and previous states using the state update function $u_\theta(s'|o, s)$, and the prediction of observations from states using the policy $\pi_\theta(o|s)$. This architecture forces reasoning through an interpretable text bottleneck, but prevents direct backpropagation, necessitating RL-based gradient estimation. In experiments, both u_θ and π_θ are implemented using the same transformer (Mistral 7B or Llama 3.1 8B), with only u_θ 's weights updated during training.

Recipient-Specific Compression. A key insight is that an *informative* CoT can also serve as a *recipient-specific compression* of the model’s hidden knowledge: it distills the essential reasoning into text that another recipient (e.g. a different model or a human) can use to predict the same outcome. Our experiments confirm that the learned CoTs generalize across interpreters, suggesting that these textual explanations genuinely encode transferable problem-solving steps rather than model-specific quirks (Section ??).

Contributions.

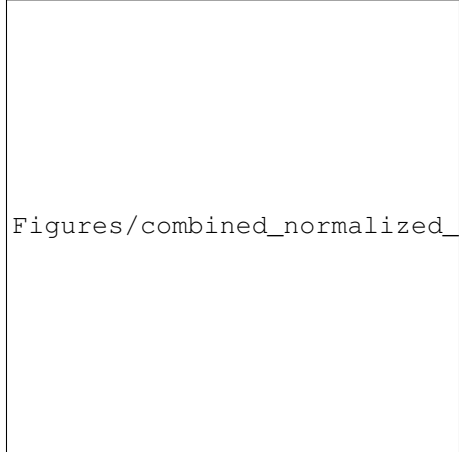
1. We introduce a Markovian language model framework that structurally enforces Chain-of-Thought (CoT) generation to be causally essential, ensuring reliance on the CoT for predictions.
2. We apply this framework to arithmetic problems (Mistral 7B) and the GSM8K dataset (?) (Llama 3.1 8B), observing a 33.2% absolute improvement on GSM8K.
3. We show through systematic perturbation analysis across four model pairs that Markovian training produces significantly higher sensitivity to CoT perturbations compared to Non-Markovian approaches, with effect differences ranging from +0.0154 to +0.3276 in log-probability sensitivity.
4. We demonstrate cross-model transfer: CoTs trained on one model remain informative for other models. This underscores the CoT’s *recipient-specific* interpretability and suggests it captures a shared reasoning strategy.

Section ?? reviews related work, Section ?? details our Markovian framework, and Section ?? describes the RL training. Section ?? presents empirical results, and Section ?? discusses limitations and future directions.

1.1 STYLE

Papers to be submitted to ICLR 2026 must be prepared according to the instructions presented here.

Authors are required to use the ICLR L^AT_EX style files obtainable at the ICLR website. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.



Figures/combined_normalized_reward_gp_smoothed.png

Figure 2: Normalized reward progression during Wikipedia continuation training across four model architectures. The normalized reward $\ln \pi_{\theta}(\text{ans} \mid \text{CoT}) - \ln \pi_{\theta}(\text{ans} \mid \text{CoT}')$ measures how much more informative the trained CoT becomes compared to baseline reasoning from the unmodified model. Each curve represents a different model architecture: Llama 3.1 8B (blue), Phi-3.5 Mini (orange), Qwen3 4B (green), and Mistral 7B (red). The plot uses Gaussian Process-style smoothing with confidence bands to highlight training trends. All models show consistent improvement in CoT informativeness, demonstrating the generalizability of the Markovian training approach across diverse architectures.

1.2 RETRIEVAL OF STYLE FILES

The style files for ICLR and other conference information are available online at:

<http://www.iclr.cc/>

The file `iclr2026_conference.pdf` contains these instructions and illustrates the various formatting requirements your ICLR paper must satisfy. Submissions must be made using \LaTeX and the style files `iclr2026_conference.sty` and `iclr2026_conference.bst` (to be used with $\text{\LaTeX}2\epsilon$). The file `iclr2026_conference.tex` may be used as a “shell” for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in sections ??, ??, and ?? below.

2 GENERAL FORMATTING INSTRUCTIONS

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing of 11 points. Times New Roman is the preferred typeface throughout. Paragraphs are separated by 1/2 line space, with no indentation.

Paper title is 17 point, in small caps and left-aligned. All pages should start at 1 inch (6 picas) from the top of the page.

Authors’ names are set in boldface, and each name is placed above its corresponding address. The lead author’s name is to be listed first, and the co-authors’ names are set to follow. Authors sharing the same address can be on the same line.

Please pay special attention to the instructions in section ?? regarding figures, tables, acknowledgments, and references.

There will be a strict upper limit of **9 pages** for the main text of the initial submission, with unlimited additional pages for citations. This limit will be expanded to **10 pages** for rebuttal/camera ready.

3 HEADINGS: FIRST LEVEL

First level headings are in small caps, flush left and in point size 12. One line space before the first level heading and 1/2 line space after the first level heading.

3.1 HEADINGS: SECOND LEVEL

Second level headings are in small caps, flush left and in point size 10. One line space before the second level heading and 1/2 line space after the second level heading.

3.1.1 HEADINGS: THIRD LEVEL

Third level headings are in small caps, flush left and in point size 10. One line space before the third level heading and 1/2 line space after the third level heading.

4 CITATIONS, FIGURES, TABLES, REFERENCES

These instructions apply to everyone, regardless of the formatter being used.

4.1 CITATIONS WITHIN THE TEXT

Citations within the text should be based on the `natbib` package and include the authors' last names and year (with the "et al." construct for more than two authors). When the authors or the publication are included in the sentence, the citation should not be in parenthesis using `\citet{}` (as in "See ? for more information."). Otherwise, the citation should be in parenthesis using `\citep{}` (as in "Deep learning shows promise to make progress towards AI (?).").

The corresponding references are to be listed in alphabetical order of authors, in the REFERENCES section. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

4.2 FOOTNOTES

Indicate footnotes with a number¹ in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).²

4.3 FIGURES

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

4.4 TABLES

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table ??.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

¹Sample of the first footnote

²Sample of the second footnote

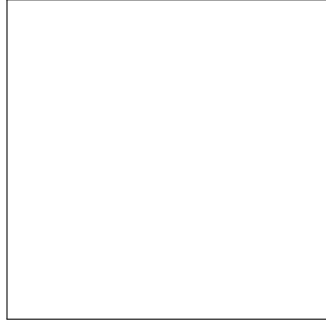


Figure 3: Sample figure caption.

Table 1: Sample table title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

5 DEFAULT NOTATION

In an attempt to encourage standardized notation, we have included the notation file from the textbook, *Deep Learning* ? available at https://github.com/goodfeli/dlbook_notation/. Use of this style is not required and can be disabled by commenting out `math_commands.tex`.

Numbers and Arrays

a	A scalar (integer or real)
\mathbf{a}	A vector
\mathbf{A}	A matrix
\mathbf{A}	A tensor
\mathbf{I}_n	Identity matrix with n rows and n columns
\mathbf{I}	Identity matrix with dimensionality implied by context
$\mathbf{e}^{(i)}$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position i
$\text{diag}(\mathbf{a})$	A square, diagonal matrix with diagonal entries given by \mathbf{a}
a	A scalar random variable
\mathbf{a}	A vector-valued random variable
\mathbf{A}	A matrix-valued random variable

Sets and Graphs

270	\mathbb{A}	A set
271	\mathbb{R}	The set of real numbers
272	$\{0, 1\}$	The set containing 0 and 1
273	$\{0, 1, \dots, n\}$	The set of all integers between 0 and n
274	$[a, b]$	The real interval including a and b
275	$(a, b]$	The real interval excluding a but including b
276	$\mathbb{A} \setminus \mathbb{B}$	Set subtraction, i.e., the set containing the elements of \mathbb{A} that are not in \mathbb{B}
277	\mathcal{G}	A graph
278	$Pa_{\mathcal{G}}(\mathbf{x}_i)$	The parents of \mathbf{x}_i in \mathcal{G}

Indexing

285	a_i	Element i of vector \mathbf{a} , with indexing starting at 1
286	\mathbf{a}_{-i}	All elements of vector \mathbf{a} except for element i
287	$A_{i,j}$	Element i, j of matrix \mathbf{A}
288	$\mathbf{A}_{i,:}$	Row i of matrix \mathbf{A}
289	$\mathbf{A}_{:,i}$	Column i of matrix \mathbf{A}
290	$\mathbf{A}_{i,j,k}$	Element (i, j, k) of a 3-D tensor \mathbf{A}
291	$\mathbf{A}_{:,:,i}$	2-D slice of a 3-D tensor
292	\mathbf{a}_i	Element i of the random vector \mathbf{a}

Calculus

295	$\frac{dy}{dx}$	Derivative of y with respect to x
296	$\frac{\partial y}{\partial x}$	Partial derivative of y with respect to x
297	$\nabla_{\mathbf{x}} y$	Gradient of y with respect to \mathbf{x}
298	$\nabla_{\mathbf{X}} y$	Matrix derivatives of y with respect to \mathbf{X}
299	$\nabla_{\mathbf{X}} y$	Tensor containing derivatives of y with respect to \mathbf{X}
300	$\frac{\partial f}{\partial \mathbf{x}}$	Jacobian matrix $\mathbf{J} \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$
301	$\nabla_{\mathbf{x}}^2 f(\mathbf{x})$ or $\mathbf{H}(f)(\mathbf{x})$	The Hessian matrix of f at input point \mathbf{x}
302	$\int f(\mathbf{x}) d\mathbf{x}$	Definite integral over the entire domain of \mathbf{x}
303	$\int_{\mathbb{S}} f(\mathbf{x}) d\mathbf{x}$	Definite integral with respect to \mathbf{x} over the set \mathbb{S}

Probability and Information Theory

324	$P(a)$	A probability distribution over a discrete variable
325	$p(a)$	A probability distribution over a continuous variable, or
326		over a variable whose type has not been specified
327		
328	$a \sim P$	Random variable a has distribution P
329	$\mathbb{E}_{x \sim P}[f(x)]$ or $\mathbb{E}f(x)$	Expectation of $f(x)$ with respect to $P(x)$
330	$\text{Var}(f(x))$	Variance of $f(x)$ under $P(x)$
331	$\text{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(x)$
332	$H(x)$	Shannon entropy of the random variable x
333	$D_{\text{KL}}(P Q)$	Kullback-Leibler divergence of P and Q
334	$\mathcal{N}(x; \mu, \Sigma)$	Gaussian distribution over x with mean μ and covariance
335		Σ
336		
337		
338		

Functions

340	$f : \mathbb{A} \rightarrow \mathbb{B}$	The function f with domain \mathbb{A} and range \mathbb{B}
341	$f \circ g$	Composition of the functions f and g
342	$f(x; \theta)$	A function of x parametrized by θ . (Sometimes we write
343		$f(x)$ and omit the argument θ to lighten notation)
344		
345	$\log x$	Natural logarithm of x
346	$\sigma(x)$	Logistic sigmoid, $\frac{1}{1 + \exp(-x)}$
347	$\zeta(x)$	Softplus, $\log(1 + \exp(x))$
348	$\ x\ _p$	L^p norm of x
349	$\ x\ $	L^2 norm of x
350	x^+	Positive part of x , i.e., $\max(0, x)$
351	$\mathbf{1}_{\text{condition}}$	is 1 if the condition is true, 0 otherwise
352		
353		
354		
355		
356		

6 FINAL INSTRUCTIONS

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the REFERENCES section; see below). Please note that pages should be numbered.

7 PREPARING POSTSCRIPT OR PDF FILES

Please prepare PostScript or PDF files with paper size “US Letter”, and not, for example, “A4”. The `-t letter` option on `dvips` will produce US Letter files.

Consider directly generating PDF files using `pdflatex` (especially if you are a MiKTeX user). PDF figures must be substituted for EPS figures, however.

Otherwise, please generate your PostScript and PDF files with the following commands:

```
dvips mypaper.dvi -t letter -Ppdf -G0 -o mypaper.ps
ps2pdf mypaper.ps mypaper.pdf
```

7.1 MARGINS IN LATEX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package.

Always specify the figure width as a multiple of the line width as in the example below using .eps graphics

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.eps}
```

or

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

for .pdf graphics. See section 4.4 in the graphics bundle documentation (<http://www.ctan.org/tex-archive/macros/latex/required/graphics/grfguide.ps>)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the \- command.

AUTHOR CONTRIBUTIONS

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

ACKNOWLEDGMENTS

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

REFERENCES

- Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards AI. In *Large Scale Kernel Machines*. MIT Press, 2007.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.
- Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

A APPENDIX

You may include other additional sections here.