Beyond Chain-of-Thought: A Survey of Chain-of-X Paradigms for LLMs

Yu Xia (/profile?id=~Yu_Xia9), Rui Wang (/profile?id=~Rui_Wang25), Xu Liu (/profile?id=~Xu_Liu12), Mingyan Li (/profile?id=~Mingyan_Li1), Tong Yu (/profile?id=~Tong_Yu3), Xiang Chen (/profile?id=~Xiang_Chen9), Julian McAuley (/profile?id=~Julian_McAuley1), Shuai Li (/profile?id=~Shuai_Li3)

12 Jun 2024 (modified: 23 Aug 2024) ACL ARR 2024 June Submission June, Senior Area Chairs, Area Chairs, Reviewers, Authors, Commitment Readers Revisions (/revisions?id=SntkuLtWTB)

Abstract:
Chain-of-Thought (CoT) has been a widely adopted prompting method, eliciting impressive reasoning abilities of Large Language Models (LLMs). Inspired by the sequential thought structure of CoT, a number of Chain-of-X (CoX) methods have been developed to address challenges across diverse domains and tasks. In this paper, we provide a comprehensive survey of Chain-of-X methods for LLMs in different contexts. Specifically, we categorize them by taxonomies of nodes, i.e., the X in CoX, and application tasks. We also discuss the findings and implications of existing CoX methods, as well as potential future directions. Our survey aims to serve as a detailed and up-to-date resource for researchers seeking to apply the idea of CoT to broader scenarios.

Paper Type: Long
Research Area: NLP Applications
Research Area Keywords: NLP Applications
Contribution Types: Surveys
Languages Studied: English
Previous URL: /forum?id=8chEm6PLRn (/forum?id=8chEm6PLRn)
Response PDF: pdf (/attachment?id=SntkuLtWTB&name=response_PDF)
Reassignment Request Action Editor: No, I want the same action editor from our previous submission and understand that a new action editor may be assigned if the previous one is unavailable
Reassignment Request Reviewers: No, I want the same set of reviewers from our previous submission and understand that new reviewers may be assigned if any of the previous ones are unavailable
A1 Limitations Section: This paper has a limitations section.
A2 Potential Risks: N/A
A3 Abstract And Introduction Summarize Claims: Yes
A3 Elaboration: Section 1
B Use Or Create Scientific Artifacts: No
B1 Cite Creators Of Artifacts: N/A
B2 Discuss The License For Artifacts: N/A
B3 Artifact Use Consistent With Intended Use: N/A
B4 Data Contains Personally Identifying Info Or Offensive Content: N/A
B5 Documentation Of Artifacts: N/A
B6 Statistics For Data: N/A
C Computational Experiments: No
C1 Model Size And Budget: N/A
C2 Experimental Setup And Hyperparameters: N/A
C3 Descriptive Statistics: N/A

C4 Parameters For Packages: N/A
D Human Subjects Including Annotators: No
D1 Instructions Given To Participants: N/A
D2 Recruitment And Payment: N/A
D3 Data Consent: N/A
D4 Ethics Review Board Approval: N/A
D5 Characteristics Of Annotators: N/A
E Ai Assistants In Research Or Writing: No
E1 Information About Use Of Ai Assistants: N/A
Reviewing Volunteers: Tong Yu (/profile?id=~Tong_Yu3)
Reviewing Volunteers For Emergency Reviewing: The volunteers listed above are only willing to serve as regular reviewers.
Reviewing No Volunteers Reason: N/A - An author was provided in the previous question.
Preprint: no
Preprint Status: There is a non-anonymous preprint (URL specified in the next question).
Existing Preprints: https://arxiv.org/abs/2404.15676 (https://arxiv.org/abs/2404.15676)
Preferred Venue: EMNLP
Consent To Share Data: yes
Consent To Share Submission Details: On behalf of all authors, we agree to the terms above to share our submission details.
Author Submission Checklist: I confirm that the paper is anonymous and that all links to data/code repositories in the paper are anonymous., I confirm that the paper has proper length (Short papers: 4 content pages maximum, Long papers: 8 content pages maximum, Ethical considerations and Limitations do not count toward this limit), I confirm that the paper is properly formatted (Templates for *ACL conferences can be found here: https://github.com/acl-org/acl-style-files (https://github.com/acl-org/acl-style-files).)
Association For Computational Linguistics - Blind Submission License Agreement: On behalf of all authors, I agree
Submission Number: 556

Discussion (/forum?id=SntkuLtWTB#discussion)

Filter by reply type... | Filter by author... | Search keywords... | Sort: Newest First

Everyone | Submission556... | Submission556 Area... | Submission556 Authors | Submission556... | Program Chairs | Submission556... | Submission556... | Submission556... | Submission556... | Submission556...

9 / 9 replies shown

Add: **Author-Editor Confidential Comment** **Withdrawal**

Meta Review of Submission556 by Area Chair Fsrs

Meta Review by Area Chair Fsrs 05 Aug 2024, 01:52 (modified: 23 Aug 2024, 06:35)
Senior Area Chairs, Area Chairs, Authors, Reviewers Submitted, Program Chairs, Commitment Readers
Revisions (/revisions?id=KmEKy1DRj6)

Metareview:
This paper proposes a comprehensive survey of Chain-of-X methods (a series of methodologies defining Chain-of-Thought (CoT)). The paper organizes the role and describes the contributions made by multiple cutting-edge works in different paradigms.

The paper is well structured (see discussion of strengths). However, substantial shortcomings emerge (see weaknesses and related descriptions).

Summary Of Reasons To Publish:
The paper has its merits indeed:

- The writing is simple because the structure is understandable and easy for the reader;
- The organization is well structured;
- The overview is comprehensive (although it is very difficult to comply with this constraint these days, so a positive point).

Summary Of Suggested Revisions:
Although the contribution has its own merits, there are points that emerged during the rebuttal process that should be improved.

- Limited consideration of the benefits and impacts of CoX methods on downstream tasks.
- Missing work, as pointed out by one reviewer (however, this is very common in this type of survey).

In conclusion, although there were some minor points of disagreement between the reviewers and authors, I believe that these can be considered and resolved during the camera-ready process. I recommend this paper for a CL conference as findings.

Overall Assessment: 3 = There are major points that may be revised
Suggested Venues: EMNLP 2024 (findings)
Best Paper Ae: No
Ethical Concerns:
There are no concerns with this submission

Needs Ethics Review: No
Author Identity Guess: 1 = I do not have even an educated guess about author identity.

Add: **Author-Editor Confidential Comment**

Official Review of Submission556 by Reviewer xgrs

Official Review by Reviewer xgrs 24 Jul 2024, 19:33 (modified: 23 Aug 2024, 06:35)
Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer xgrs, Commitment Readers
Revisions (/revisions?id=ryeps2bC7x)

Paper Summary:
This survey presents a comprehensive overview of Chain-of-X methods. Chain-of-Thought (CoT) has been a widely adopted prompting method, eliciting impressive reasoning abilities of LLMs. A number of Chain of-X (CoX) methods have been developed to address challenges across diverse domains and tasks. Specifically, the author categorized them by taxonomies of nodes, i.e., the X in CoX, and application tasks; discuss the findings and implications of existing CoX methods, as well as potential future directions.

Summary Of Strengths:
The authors defined a generalized concept of Chain-of-X. And they surveyed existing COX method by taxonomy of nodes and taxonomy of tasks. The representative methods are classified clearly in figure.They also pointed out future directions.

Summary Of Weaknesses:
Although the authors show us a clear classification tree of the Chain-of-X, from the writing part, I still have the following concerns:

1. For some applications and tasks, the accuracy of the results is very important. Although the authors pointed out evaluation tools, is it appropriate to make evaluation tools as section 4.6 which is part of taxonomy of tasks?
2. As a survey, of course, we would like to know the classification of different models. But we also want to know the merits and demerits of these models. Then, we can use the guidelines to adopt them.

Comments Suggestions And Typos:
As shown in "Summary Of Weaknesses", we would like to know the answers.

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.
Soundness: 3 = Acceptable: This study provides sufficient support for its major claims/arguments. Some minor points may need extra support or details.
Overall Assessment: 3 = Good: This paper makes a reasonable contribution, and might be of interest for some (broad or narrow) sub-communities, possibly with minor revisions.
Best Paper: No
Needs Ethics Review: No
Reproducibility: 2 = They would be hard pressed to reproduce the results: The contribution depends on data that are simply not available outside the author's institution or consortium and/or not enough details are provided.
Datasets: 1 = No usable datasets submitted.
Software: 1 = No usable software released.
Knowledge Of Or Educated Guess At Author Identity: No
Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources
Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources
Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**

Reply to Reviewer xgrs

Official Comment

by Authors (Xu Liu (/profile?id=~Xu_Liu12), Shuai Li (/profile?id=~Shuai_Li3), Yu Xia (/profile?id=~Yu_Xia9), Xiang Chen (/profile?id=~Xiang_Chen9), +4 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission556/Authors))

30 Jul 2024, 15:48 (modified: 23 Aug 2024, 06:35)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer xgrs, Commitment Readers

Revisions (/revisions?id=Yz08aPPj4x)

Comment:
Thank you for your time and effort in reviewing our paper. We provide responses to your questions as follows.

Response to Weakness 1:

Thank you for your question. In Section 4.6 (Evaluation Tools), we discuss methods that apply Chain-of-X as tools for more fine-grained evaluation of LLMs on various text generation tasks, revealing problems that traditional evaluations might not capture. On the other hand, as you suggested, the evaluation of Chain-of-X methods themselves is indeed very important to validate their effectiveness. As previous surveys, e.g., [1], have already collected comprehensive evaluation benchmarks, we will include some additional discussion on evaluations in our updated version.

Response to Weakness 2:

Thank you for your insightful suggestion. We have discussed the merits and demerits of different methods within each Chain-of-X paradigm, e.g., Section 3.2. The differences between various Chain-of-X paradigms are also shown in Figure 1. As you suggested, we will include further discussion on their merits and demerits. For example, Chain-of-Models (Section 3.4), although collecting multi-aspect information from different LLMs, tends to be more costly than other paradigms.

[1] Chu, Zheng, et al. A Survey of Chain of Thought Reasoning: Advances, Frontiers and Future. ACL 2024.

Add: **Author-Editor Confidential Comment**

Official Review of Submission556 by Reviewer JsNR

Official Review by Reviewer JsNR 23 Jul 2024, 01:08 (modified: 23 Aug 2024, 06:35)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer JsNR, Commitment Readers

Revisions (/revisions?id=cZ9zt1zS0s)

Paper Summary:
The other propose a survey paper on Chain-of-X (CoX) methods, described here as the generalization of Chain-of-Thought. They categorize CoX methods by the types of components used to construct the chains. Furthermore, based on the application areas of these CoX methods, they classify them by tasks. lastly, they discuss insights from existing CoX methods and explore potential future directions.

Summary Of Strengths:
Such a paper was cruelly lacking in the literature, and I am glad the authors went through the effort of compiling so much information in their manuscript. The paper is very well written and the taxonomy based both on the nodes nature and the tasks these methods are applied in provides a very useful and informative structure. In particular, Figure 2 provides an extensive bird's eye view on these methods. Furthermore, the descriptions of each nodes and tasks, although compact, are perfectly explained and will be very useful for the community, both experts and people that wish to gain superficial information on the matter.

Summary Of Weaknesses:
This is a survey paper, as such, I based my opinion on the quality of the structure and how comprehensive it was when providing information on CoX methods. Given that his paper is very well structure and includes, to my knowledge, most of the papers on this topic, I do not see any weaknesses in this work.

Comments Suggestions And Typos:
I have no more concerns and could not find any typos or editing issues.

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.
Soundness: 4 = Strong: This study provides sufficient support for all of its claims/arguments. Some extra experiments could be nice, but not essential.
Overall Assessment: 4 = This paper represents solid work, and is of significant interest for the (broad or narrow) sub-communities that might build on it.
Best Paper: No
Limitations And Societal Impact:
N/A

Ethical Concerns:
N/A

Needs Ethics Review: No
Reproducibility: 5 = They could easily reproduce the results.
Datasets: 1 = No usable datasets submitted.
Software: 1 = No usable software released.
Knowledge Of Or Educated Guess At Author Identity: No
Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources
Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources
Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**

Reply to Reviewer JsNR

Official Comment

by Authors (Xu Liu (/profile?id=~Xu_Liu12), Shuai Li (/profile?id=~Shuai_Li3), Yu Xia (/profile?id=~Yu_Xia9), Xiang Chen (/profile?id=~Xiang_Chen9), +4 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission556/Authors))

30 Jul 2024, 15:49 (modified: 23 Aug 2024, 06:35)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer JsNR, Commitment Readers

Revisions (/revisions?id=PzTHhyi1Eo)

Comment:
Thank you for your time and effort in reviewing our paper. We sincerely appreciate your recognition of our work and believe our survey will be useful to the community.

Add: **Author-Editor Confidential Comment**

Official Review of Submission556 by Reviewer CuFr

Official Review by Reviewer CuFr 21 Jul 2024, 15:25 (modified: 23 Aug 2024, 06:35)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer CuFr, Commitment Readers

Revisions (/revisions?id=tS5uAGICP7)

Paper Summary:
This paper provides a survey of variants of chain-of-thought (CoT) methods. CoT methods are known to effectively improve the performance of LLMs in reasoning tasks by eliciting the reasoning steps leading to the answer. Given the success, many variants (chain-of-X; CoX) have been proposed. This paper categorizes existing CoX methods from two points of view: what constitutes each reasoning step (i.e., X) and the applications.

Summary Of Strengths:
- This paper is well-written and easy to follow.
- This paper provides a comprehensive list of existing CoX methods and their review. Unlike the previous version, the current manuscript refers to methods that are not literally named CoX but essentially extend the idea of CoT, leading to broader coverage.
- The taxonomy focusing on the applications is beneficial for practitioners.

Summary Of Weaknesses:
- This paper should again carefully check the contributions of the surveyed papers. For example, Chain-of-Hindsight (line 360) is a technique for constructing instruction data and is not related to CoT methods that elicit reasoning steps during runtime.

Comments Suggestions And Typos:
- Line035: Yu et al. (2023a) could be a mistake for Yu et al. (2023b).

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.
Soundness: 3 = Acceptable: This study provides sufficient support for its major claims/arguments. Some minor points may need extra support or details.
Overall Assessment: 3 = Good: This paper makes a reasonable contribution, and might be of interest for some (broad or narrow) sub-communities, possibly with minor revisions.
Best Paper: No
Ethical Concerns:
None

Needs Ethics Review: No
Reproducibility: 5 = They could easily reproduce the results.
Datasets: 1 = No usable datasets submitted.
Software: 1 = No usable software released.
Knowledge Of Or Educated Guess At Author Identity: No
Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources
Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources
Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**

Reply to Reviewer CuFr

Official Comment

by Authors (Xu Liu (/profile?id=~Xu_Liu12), Shuai Li (/profile?id=~Shuai_Li3), Yu Xia (/profile?id=~Yu_Xia9), Xiang Chen (/profile?id=~Xiang_Chen9), +4 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission556/Authors))

30 Jul 2024, 15:51 (modified: 23 Aug 2024, 06:35)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer CuFr, Commitment Readers

Revisions (/revisions?id=NOoL867A6E)

Comment:

Thank you for your time and effort in reviewing our paper. We provide responses to your suggestions as follows.

Response to Weakness:

Thank you for your valuable suggestion. As discussed in Section 5 of the Chain-of-Hindsight paper [1], the authors view CoT prompts as consisting of human-written stepwise instructions, while their proposed Chain-of-Hindsight consists of human-written hindsight feedback for instruction data construction. We believe this concept naturally falls into one of the CoX applications in the taxonomy of tasks beyond reasoning, as shown in Figure 2. As you suggested, we will include additional discussion on the connections between the surveyed CoX methods and CoT to improve clarity.

Response to Comment:

Thank you for your careful reading. We have fixed the typo.

[1] Liu, Hao, et al. Chain of Hindsight Aligns Language Models with Feedback. ICLR 2024

Add: **Author-Editor Confidential Comment**

Official Review of Submission556 by Reviewer MBG7

Official Review by Reviewer MBG7 20 Jul 2024, 06:08 (modified: 23 Aug 2024, 06:35)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer MBG7, Commitment Readers

Revisions (/revisions?id=Zz1DlF9DOE)

Paper Summary:

This paper introduces the concept of Chain-of-X (CoX), a generalization of the popular chain-of-thought prompting framework, to indicate methods that involve a multi-step output generation to obtain a final model prediction. In particular, the authors define four macro categories of CoX methods depending on the intermediate steps leading to the final prediction: Chain of x, for x \in {Thought, Augmentation, Feedback, Models}. For each of these categories, the paper lists several methods organized into sub-categories. Additionally, the paper provides an alternative classification of the CoX methods based on the task to which they are or can be applied.

Summary Of Strengths:

- The survey is presented clearly and structured well. Figures 2 and 4 represent a good visualization of the categorization presented by the authors.
- The motivation behind the survey is reasonable.

Summary Of Weaknesses:

- My main concern with this work is about the actual utility of this survey for the community.
- Lines 29-31: the transparency of CoT is debatable (see, e.g., Turpin et al., 2024). This seems like a topic that a survey like the one presented should at least mention.

Turpin, Miles, et al. "Language models don't always say what they think: unfaithful explanations in chain-of-thought prompting." Advances in Neural Information Processing Systems 36 (2024).

Comments Suggestions And Typos:

- Since the two dimensions along which the authors categorize CoX tasks are orthogonal (nature of the intermediate nodes and application task), a visualization of the tasks presented as a matrix in which the axes are these two

dimensions would be appreciated.

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Soundness: 3 = Acceptable: This study provides sufficient support for its major claims/arguments. Some minor points may need extra support or details.

Overall Assessment: 3 = Good: This paper makes a reasonable contribution, and might be of interest for some (broad or narrow) sub-communities, possibly with minor revisions.

Best Paper: No

Needs Ethics Review: No

Reproducibility: 5 = They could easily reproduce the results.

Datasets: 1 = No usable datasets submitted.

Software: 1 = No usable software released.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**

Reply to Reviewer MBG7

Official Comment

by Authors (Xu Liu (/profile?id=~Xu_Liu12), Shuai Li (/profile?id=~Shuai_Li3), Yu Xia (/profile?id=~Yu_Xia9), Xiang Chen (/profile?id=~Xiang_Chen9), +4 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission556/Authors))

30 Jul 2024, 15:56 (modified: 23 Aug 2024, 06:35)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer MBG7, Commitment Readers

Revisions (/revisions?id=nCURAcDkSU)

Comment:

Thank you for your time and effort in reviewing our paper. We provide responses to your suggestions as follows.

Response to Weakness 1:

Thank you for your question. With the rapidly growing number of Chain-of-X methods, our survey aims to serve as an up-to-date resource for: 1) beginners who wish to gain a broad overview of the rapidly evolving CoT concept (e.g., Figure 1 and Figure 2), 2) researchers who are seeking ideas for more creative design or analysis of CoX methodology (e.g., Section 3 and Section 5), and 3) practitioners who are looking for guidance in applying the CoX concept to various tasks beyond reasoning (e.g., Section 4).

Response to Weakness 2:

Thank you for your valuable suggestion. We will include some discussion on this topic and adjust our claims about the transparency of CoT accordingly.

Response to Comment:

Thank you for your valuable suggestion. A matrix visualization of CoX methods along two dimensions, i.e., nodes and tasks, is indeed a very useful way of presentation. We will include it in our updated version.

Add: **Author-Editor Confidential Comment**