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

Yu Xia[1,2] **Rui Wang**[3] **Xu Liu**[1] **Mingyan Li**[1] **Tong Yu**[4]
Xiang Chen[4] **Julian McAuley**[2] **Shuai Li**[1*]
[1]Shanghai Jiao Tong University [2]UC San Diego [3]Duke University [4]Adobe Research
{yux078, jmcauley}@ucsd.edu {tyu, xiangche}@adobe.com
rui.wang16@duke.edu {liu_skywalker, QYLJM1217, shuaili8}@sjtu.edu.cn

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 various challenges across diverse domains and tasks involving LLMs. 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.

1 Introduction

Large Language Models (LLMs) have shown strong reasoning capabilities when prompted with the Chain-of-Thought (CoT) method (Wei et al., 2022; Yao et al., 2024; Besta et al., 2024a). The essence of CoT is to decompose complex problems into sequences of intermediate subtasks (Chu et al., 2023; Zhou et al., 2023). By handling these subtasks step by step, LLMs are able to focus on important details and assumptions, which substantially improves their performance across a wide range of reasoning tasks (Huang and Chang, 2023; Chu et al., 2023). Additionally, CoT's intermediate steps offer a more transparent reasoning process, facilitating easier interpretation and evaluation of LLMs (Yu et al., 2023b).

With the success of CoT, a number of Chain-of-X (CoX) methods have subsequently been developed (Yu et al., 2023a). Extending beyond reasoning thoughts, recent CoX methods have constructed the chain with various components, such as Chain-of-Feedback (Lei et al., 2023; Dhuliawala et al., 2023), Chain-of-Instructions (Zhang et al., 2023d; Hayati et al., 2024), Chain-of-Histories (Luo et al., 2024; Xia et al., 2024d), etc. These methods have been applied to tackle challenges in diverse tasks involving LLMs beyond reasoning, including multi-modal interaction (Xi et al., 2023a; Zhang et al., 2024a), hallucination reduction (Lei et al., 2023; Dhuliawala et al., 2023), planning with LLM-based agents (Zhan and Zhang, 2023; Zhang et al., 2024c), etc.

*Corresponding author.

Motivation Despite their growing prevalence, these CoX methods have not yet been collectively examined or categorized, leaving a gap in our understanding of their potential and nuances. To this end, this survey aims to offer a structured overview that captures the essence and diversity of CoX methods for further exploration and innovation.

Distinguishing Focus While several surveys have explored CoT (Chu et al., 2023; Yu et al., 2023b; Besta et al., 2024b), they focus primarily on the reasoning thoughts of different structures, e.g., Chain-of-Thought as illustrated in Figure 1(a). In contrast, this paper focuses on the multifaceted component designs of Chain-of-X beyond reasoning thoughts as shown in Figure 1, offering insights of the CoT concept in broader domains. We present a comprehensive review by taxonomies of the X in CoX and tasks to which these methods are applied.

Overview of the Survey We first provide background information on Chain-of-Thought and define Chain-of-X as its generalization (§2). Next, we categorize CoX methods by the types of components used to construct the chains (§3). Furthermore, based on the application areas of these CoX methods, we categorize them by tasks (§4). Then, we discuss insights from existing CoX methods and explore potential future directions (§5). A detailed structure of the survey is presented in Figure 2.

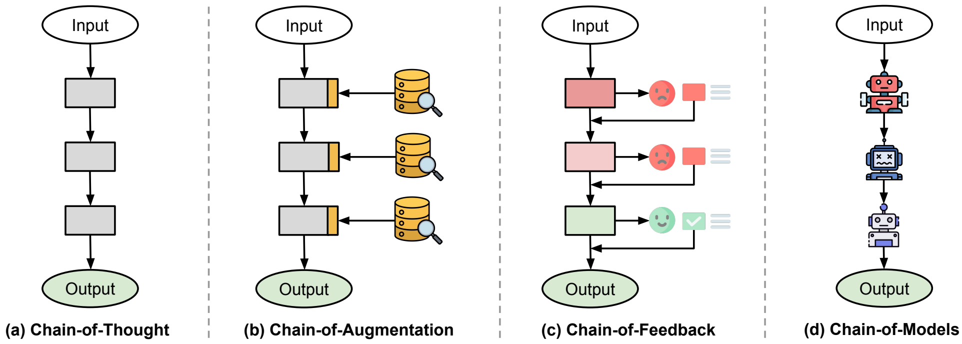

Figure 1: Illustrations of Chain-of-X Paradigms with Four Types of Nodes: (a) Intermediates, e.g., Thought (§3.1), (b) Augmentation (§3.2), (c) Feedback (§3.3), and (d) Models (§3.4).

2 What is Chain-of-X?

In this section, we introduce some background information about Chain-of-Thought prompting and then define a generalized concept of Chain-of-X.

Chain-of-Thought CoT prompting is a methodology that substantially enhances the reasoning capabilities of LLMs. Introduced by Wei et al. (2022), CoT involves prompting LLMs with a structured format of `<input, thoughts, output>`, where 'thoughts' encompass coherent and intermediate natural language reasoning steps leading to the final answer. CoT's effectiveness is most pronounced in tasks that require complex reasoning. Traditional few-shot learning methods often falter in such scenarios, as they tend to provide direct answers without the necessary intermediate steps. Rae et al. (2021) highlighted this limitation, noting the inadequacy of these methods with increased model size. In contrast, CoT prompting excels by incorporating intermediate reasoning steps. These steps guide the model through a logical progression, enhancing its capability to tackle complex problems, such as those involving arithmetic, commonsense, and symbolic reasoning (Wang et al., 2023d; Lyu et al., 2023). The essence of CoT lies in its strategy to tackle complex problems by breaking them down into manageable intermediate steps (Zhou et al., 2023). Kojima et al. (2022) have also demonstrated strong performance of zero-shot CoT by prompting "*Let's think step by step.*". The explicit reasoning steps also provide a transparent pathway for the model's thought process, allowing for further evaluations and corrections (Yu et al., 2023b).

Chain-of-X Inspired by the nature of the sequential breakdown of CoT, a substantial number of CoX methods have been developed recently (Yu et al., 2023a). Here, we define CoX as a generalization of the CoT method for diverse tasks beyond LLM reasoning. We refer to the X in CoX as the 'node' of the chain structure. Beyond the thoughts in CoT prompts, the X in CoX can take various forms tailored to specific tasks, including intermediates (§3.1), augmentation (§3.2), feedback (§3.3), and even models (§3.4), as illustrated in Figure 1. We summarize the types of nodes in existing CoX methods in Figure 2. The idea of CoX is to construct a sequence of problem-related components that either compositionally contribute to the solution or iteratively refine the outputs for a complex task. Similarly, we define a structured format for CoX as <input, $\mathtt{X}_1$, ..., $\mathtt{X}_n$, output> where n is the length of the chain. Note that this format extends beyond prompting strategies like CoT and can be adapted to a variety of algorithmic frameworks or structures for diverse tasks involving LLMs. For instance, Chain-of-Verification (Dhuliawala et al., 2023) is a hallucination reduction framework that employs an LLM to generate initial responses, composes a sequence of verification questions, and revises its previous responses based on these questions. In addition to hallucination reduction, CoX methods have been applied to a variety of tasks, as shown in Figure 2, including multi-modal interaction (§4.1), factuality & safety (§4.2), multi-step reasoning (§4.3), instruction following (§4.4), LLMs as Agents (§4.5), and evaluation tools (§4.6).

3 Chain-of-X Nodes

In this section, we survey existing CoX methods by taxonomy of nodes, categorizing them as shown in Figure 2 based on the distinct nature of the nodes.

3.1 Chain-of-Intermediates

Building on the concept of utilizing explicit intermediate steps, a natural evolution of CoT in-

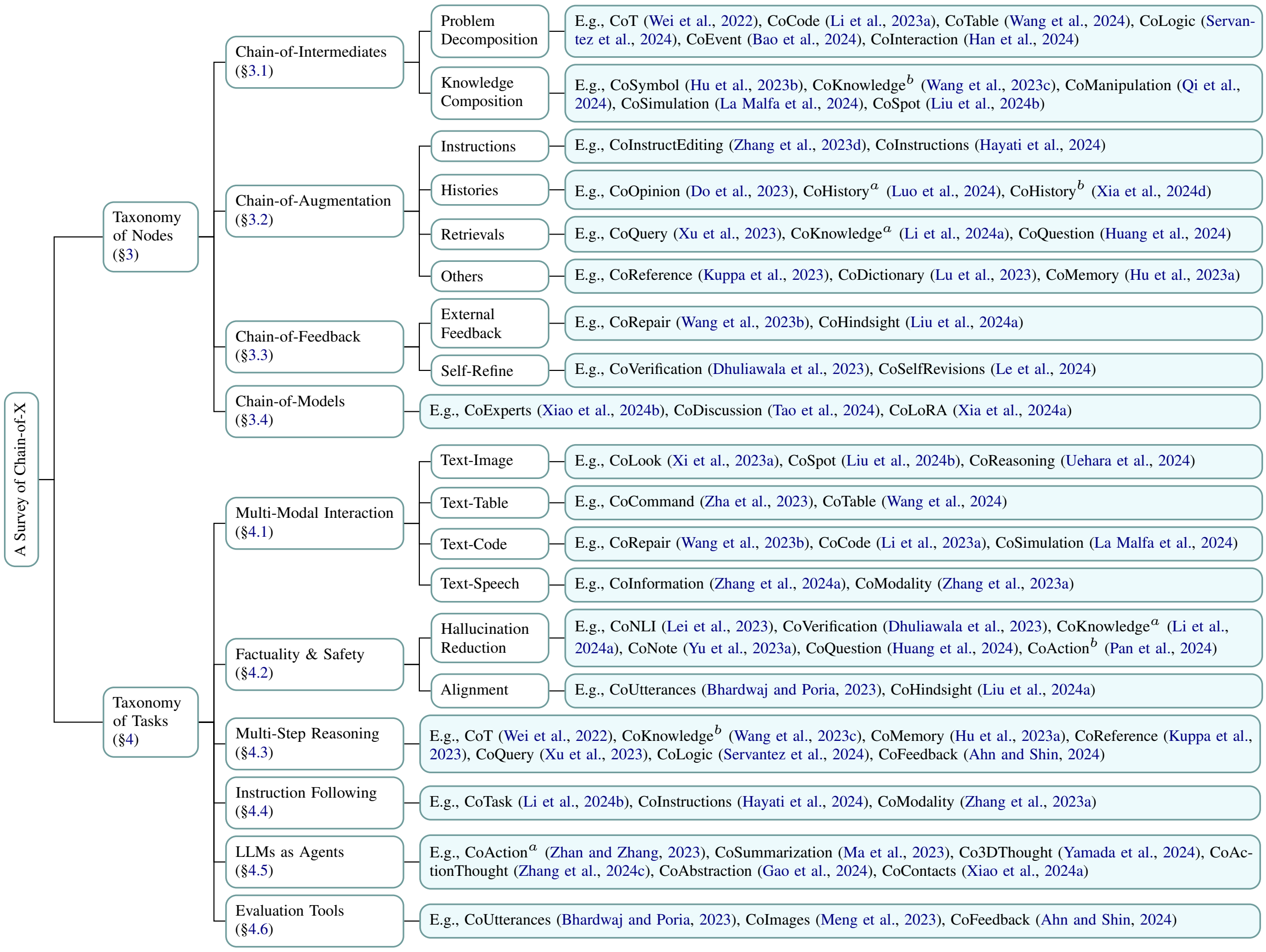

Figure 2: A Survey of Chain-of-X by Taxonomies of Nodes and Tasks (only representative methods are listed due to space limitation and a more complete version can be found in Appendix A).

volves replacing reasoning thoughts with other types of intermediate components, i.e., Chain-of-Intermediates. Based on the primary focuses, we further divide them into the following subtypes.

Problem Decomposition In problem decomposition, the intermediate steps consist of manageable subtasks derived from an original complex problem. This approach is exemplified by the classic Chain-of-Thought prompting (Wei et al., 2022). Extending this further, Li et al. (2023a) introduce Chain-of-Code, which segments a task into programmatic subtasks, enhancing the reasoning process through simulated code outputs. Similarly, Wang et al. (2024) have developed the Chain-of-Table framework. This framework restructures complex tables into question-specific formats via a sequence of strategic operations, making the data more accessible and tailored to the inquiry. Moreover, the Chain-of-Logic, introduced by Servantez et al. (2024), applies a logical decomposition to rule-based reasoning tasks, transforming them into a series of logical expressions. The methodological breakdown facilitates clearer reasoning pathways. These decomposition methods are also echoed in Chain-of-Event (Han et al., 2024), which simplify multi-document summarization into discrete and manageable event extraction tasks, significantly enhancing quality and reducing potential errors.

Knowledge Composition In knowledge composition, the primary goal of the intermediate steps is not on simplification but on the accumulation of relevant information and evidence. This approach aims to enrich the solution with a depth of understanding and details. For instance, Hu et al. (2023b) propose Chain-of-Symbol method that meticulously collects spatial relations during spatial planning tasks, enhancing the model's precision and effectiveness. Likewise, La Malfa et al. (2024) adopt Chain-of-Simulation prompting to ensures that each step in code execution is informed by program traces, thereby avoiding memorization pitfalls. Wang et al. (2023c) take a similar approach with Chain-of-Knowledge, extracting crucial pieces of evidence at each step to support more grounded and reliable question-answering sessions. This technique is particularly effective in fostering a deeper understanding of the queried material. In visual tasks, methods like Chain-of-Spot (Liu et al., 2024b) and Chain-of-Reasoning (Uehara et al., 2024) assist vision-language models in focusing on specific image details, which is crucial for tasks requiring detailed visual evidences. Through these evidence-rich methods, LLMs achieve a comprehensive and nuanced understanding of complex scenarios, leading to higher quality outputs.

3.2 Chain-of-Augmentation

A popular variant of CoX methods is Chain-of-Augmentation, where the chain is augmented with additional knowledge. Based on the types of augmented data, we categorize them as follows.

Instructions Instructions serve as an important augmentation, guiding LLMs through complex reasoning or task execution processes where determining the next step can be nontrivial (Zha et al., 2023). For instance, the Chain-of-InstructEditing framework (Zhang et al., 2023d) harnesses this concept by generating sequential instructions to guide image editing tasks, illustrating how specific editions can refine the output by focusing on relevant areas. Moreover, Zha et al. (2023) introduce the Chain-of-Command to tackle ambiguities in user instructions for table manipulations. Inferring from user instructions, it enables LLMs to employ a series of precise pre-defined commands for more accurate table execution. In the realm of e-commerce, Li et al. (2024b) implement a similar structured approach with their Chain-of-Task, which breaks down customer interactions into manageable atomic tasks, significantly simplifying complex operations. Similarly, the Chain-of-Instructions framework proposed by Hayati et al. (2024) iteratively solves decomposed subtasks using outputs of previous steps as instructions for the next step. The results show that stepwise guidance can substantially improve both the process and the outcomes of complex problem-solving tasks.

Histories Utilizing historical data for informed predictive modeling is another facet of Chain-of-Augmentation, drawing contextual insights from past interactions or events. This approach is exemplified by Do et al. (2023)'s Chain-of-Opinion, which analyzes historical user opinions to predict future reactions, offering valuable foresight into user sentiment. In user-interface exploration, Zhan and Zhang (2023) apply a Chain-of-Action^{a} framework, leveraging past actions to guide future interactions, thereby optimizing user experience through learned behaviors. Ma et al. (2023) take a similar approach in gaming environments like StarCraft II, where a Chain-of-Summarization provides strategic recommendations based on a synthesis of past gameplay observations. The development of taxonomy structures also benefits from historical data, as seen in the Chain-of-Layer by Zeng et al. (2024), which builds upon previously identified categories to enhance classification tasks. Temporal knowledge graphs receive a forward-looking treatment as well with methods like Chain-of-History from Luo et al. (2024) and Xia et al. (2024d), where historical graph structures inform predictions about future linkages and interactions.

Retrieval Chain-of-Retrievals methods are designed to intersperse the generation process with sequences of explicit retrievals, enhancing the quality of the generated content (Zhao et al., 2023). For instance, Xu et al. (2023) introduce the Chain-of-Query framework, which improves the search capabilities of LLMs through a systematic arrangement of query-answer pairs, each aimed at enhancing information retrieval. Similarly, Chain-of-Question proposed by Huang et al. (2024) focuses on refining query mechanisms, where each sub-question decomposed from the original question helps in retrieving more accurate knowledge from external knowledge base. Further refining this concept, Li et al. (2024a) have crafted Chain-of-Knowledge, which dynamically pulls relevant information from a knowledge base to correct and align inconsistent rationales within CoT frameworks. These methods illustrate how strategic retrieval integration improve LLMs' problem-solving accuracy, leading to enhanced output fidelity.

Others Beyond the conventional types of augmentation, various domain-specific augmentations have also been applied to CoX methods for LLMs. In the realm of emotional intelligence, Lee et al. (2023b) introduce the Chain-of-Empathy, infusing psychotherapy insights to cultivate empathetic responses from LLMs. Meanwhile, Kuppa et al. (2023) propose Chain-of-Reference method that integrates legal frameworks to meticulously deconstruct and address complex legal inquiries, showcasing the versatility of CoX in specialized fields.

On a similar note, Gao et al. (2024) develop a Chain-of-Abstraction framework that uses domain-specific tools to fill in abstract placeholders intentionally left in LLMs' reasoning chains. The enhancement of linguistic tools is also evident in Chain-of-Dictionary (Lu et al., 2023), which enhances machine translation with a multilingual dictionary tailored to each sentence. These diverse augmentations not only broaden the operational scope of LLMs but also underscore the potential of tailored, domain-specific enhancements.

3.3 Chain-of-Feedback

Chain-of-Feedback represents another variant of CoX. Unlike augmentation which typically precedes generation, feedback is interlaced throughout the generation process to enhance and fine-tune responses. Based on the feedback source, we categorize them as external and self-refinement feedback.

External Feedback Feedback from external sources provides valuable external perspectives that can guide the refinement process in LLMs. For instance, Yamada et al. (2024) introduce Chain-of-3DThought using external critiques help iteratively hone an LLM's understanding of 3D spaces. Similarly, Wang et al. (2023b) employ a teacher-student framework in their Chain-of-Repair, where feedback from the compiler is first interpreted by a teacher LLM and then used to guide a student LLM in code generation. This approach not only corrects errors but also facilitates a learning process whereby the student model gains proficiency over time. Additionally, Liu et al. (2024a) have developed Chain-of-Hindsight, transforming direct human preference into natural language feedback that better aligns with how LLMs process information. These feedback allows for more precise refinement to model's outputs, ensuring that responses are both accurate and contextually appropriate.

Self-Refine The potential costs and unavailability of external feedback have led to a growing interest in self-refinement abilities within LLMs (Lee et al., 2023a). Highlighted by Lei et al. (2023), Chain-of-NLI guides an LLM to evaluate and refine its outputs through a series of natural language inference tasks constructed based on its initial responses. Echoing this approach, Dhuliawala et al. (2023) introduce Chain-of-Verification empowers LLMs to self-assess through a sequence of self-generated verification questions, leading to progressively refined answers. Both methods identify and correct ungrounded outputs autonomously, enhancing the reliability of responses. Adams et al. (2023) further this concept with Chain-of-Density, which allows LLMs to iteratively incorporate self-detected missing information into their previous outputs. Together with Chain-of-SelfRevisions (Le et al., 2024) and Chain-of-Feedback (Ahn and Shin, 2024), these frameworks exemplify how LLMs can utilize their own outputs for continuous self-improvement.

3.4 Chain-of-Models

Previous CoX methods have mostly been designed for a single LLM. Recognizing that different LLMs may have distinct specialties (Xiao et al., 2024b; Xia et al., 2024b), another line of work proposes constructing a chain of models to leverage distinct strengths of each model. The Chain-of-Experts (Xiao et al., 2024b) exemplifies this collaborative strategy. It involves a consortium of expert LLMs that work in sequence, each contributing its specialized knowledge to build upon the reasoning developed by its predecessors. This method is particularly effective in addressing intricate problems in operation research, where the complexity often exceeds the processing capabilities of a single LLM. Similarly, Qiu et al. (2024) deploy a chain of specialized LoRA (Low Rank Adaptation (Hu et al., 2022)) networks, each fine-tuned to effectively handle different domains of a broader problem. This tailored approach ensures that specific tasks benefit from the most relevant and effective expertise, enhancing overall efficiency and outcome accuracy. In parallel, Tao et al. (2024) have developed the Chain-of-Discussion, where multiple LLMs engage in a structured dialogue, critiquing and refining each other's contributions before reaching a consensus in the final response. This process ensures that the synthesized output is not only comprehensive but also critically evaluated from multiple perspectives.

4 Chain-of-X Tasks

As presented in the previous section, the nodes of CoX can be of various forms, enabling their applications to extend beyond LLM reasoning. This section surveys existing CoX methods categorized by tasks, as shown in Figure 2.

4.1 Multi-Modal Interaction

Although CoT was originally proposed for text generation, various CoX methods have been developed to tackle challenges in multi-modality.

Text-Image In the realm of vision-language models, the synergy between textual and visual data is critical (Zhang et al., 2024b). CoX methods have been instrumental in enhancing this interplay. For instance, Chain-of-InstructEditing (Zhang et al., 2023d) utilizes text-based instructions to guide the nuanced task of image editing, specifically for facial manipulations. This method ensures that image alterations adhere closely to textual descriptions, enhancing the accuracy and relevance of the edits. Similarly, Chain-of-Look (Xi et al., 2023a) introduces a structured approach to visual entity recognition by constructing a visual semantic reasoning chain that mirrors the logical progression of CoT. This method facilitates deeper understanding and identification of visual elements through descriptive textual cues. Furthermore, Chain-of-QA (Kim et al., 2024) expands this approach into a dynamic dialogue between an LLM and a visual question-answering model, tackling complex queries with a combination of textual and visual analyses. Additionally, Chain-of-Reasoning (Uehara et al., 2024) and Chain-of-Manipulation (Qi et al., 2024) focus on refining the process of identifying and interpreting critical details within images. These methods systematically guide the model to focus on specific regions of an image, thereby improving the model's visual reasoning ability for more precise responses.

Text-Table The challenges of complex tabular data manipulation have also been studied using CoX methods. Chain-of-Command (Zha et al., 2023), for example, provides LLMs with a sequence of pre-defined commands, guiding them through the accurate manipulation of tables. This structured guidance helps prevent errors that might arise from ambiguous or incorrect interpretations of the task requirements. On a related note, Chain-of-Table (Wang et al., 2024) leverages tabular data as a part of the reasoning chain. Here, tables are not just data sources but act as evolving entities within the reasoning process, dynamically updating and refining themselves in response to the LLM's queries and tasks. This iterative process allows the model to engage with the table more naturally and effectively, leading to a more nuanced understanding and manipulation of the contained information.

Text-Code Code generation is another task that has benefited from the introduction of CoX methods (Zan et al., 2023). Chain-of-Code (Li et al., 2023a), for example, tackles code generation by breaking down problems into a sequence of programs and then simulates code execution to solve the task to address the overarching task effectively. Expanding on this idea, Chain-of-Simulation (La Malfa et al., 2024) takes a granular approach by executing code line by line. In contrast, Chain-of-Repair (Wang et al., 2023b) draws inspiration from traditional debugging processes, where feedback from compilers is used not just to identify but also to explain bugs, facilitating a deeper learning process for the LLM as it generates fixes. Meanwhile, Chain-of-SelfRevisions (Le et al., 2024) explores a creative reuse strategy, where snippets of code from previous tasks are recycled into new projects, enhancing efficiency and promoting a modular approach to code generation. Together, these methods underscore the versatility of CoX techniques in refining code generation tasks, highlighting their ability to adapt and respond to the intricacies of programming.

Text-Speech Similarly, the field of speech generation has seen innovative applications of CoX methods. Chain-of-Information (Zhang et al., 2024a), for instance, enhances speech synthesis by methodically separating and then reassembling semantic and perceptual components, which allows for more nuanced and accurate speech output. Another approach is the Chain-of-Modality (Zhang et al., 2023a), which merges textual and vocal instructions to guide the speech generation process. This method not only enhances the quality of speech generation but also equips LLMs with the ability to handle conversational nuances, effectively bridging the gap between textual and speech data.

4.2 Factuality & Safety

Ensuring factual consistency and safety in LLM outputs has been critical (Wang et al., 2023e; Zhang et al., 2023c; Dong et al., 2024). To make LLMs generate more factual and safer outputs, recent studies have explored the use of CoX methods in both hallucination reduction and alignment.

Hallucination Reduction LLMs have shown a propensity for generating hallucinations (Akhtar et al., 2023; Agrawal et al., 2023; Xia et al., 2024c). Studies have explored the use of CoX methods to reduce hallucinations. For example, Chain-of-NLI (Lei et al., 2023) utilizes a sequence of natural language inference problems derived from initial model outputs to guide systematic revisions, enhancing the factual accuracy of subsequent responses. Similarly, Chain-of-Verification (Dhuliawala et al., 2023) prompts an LLM to generate and answer its own verification questions, enabling it to critically assess and refine its responses. Furthermore, recognizing the effectiveness of retrieval-augmented approaches in grounding responses with accurate information (Gao et al., 2023), several CoX methods, e.g., Chain-of-Note (Yu et al., 2023a), Chain-of-Knowledge[a] (Li et al., 2024a), Chain-of-Action[b] (Pan et al., 2024), have been implemented to retrieve and integrate domain-specific knowledge at each step, effectively reducing the occurrence of incorrect or misleading information.

Alignment Aligning LLMs with human preferences is another critical area where CoX methods have shown promising results (Wang et al., 2023e). To enhance LLMs' understanding of human preferences, Chain-of-Hindsight (Liu et al., 2024a) transforms them into a sequence of natural language feedback for fine-tuning. Leveraging the language comprehension capabilities of LLMs, Chain-of-Hindsight achieves superior alignment performance compared to previous methods like RLHF (Ouyang et al., 2022). Meanwhile, Chain-of-Utterance prompting (Bhardwaj and Poria, 2023) has been proposed for LLM red-teaming, establishing a jailbreaking conversation between a harmful LLM and a helpful but unsafe LLM. The harmful questions gathered through Chain-of-Utterances are utilized to create the HarmfulQA dataset, which serves as a basis for further safety alignment efforts (Bhardwaj and Poria, 2023). By integrating these methods, CoX frameworks not only enhance the immediate utility of LLMs but also contribute to broader efforts in developing AI systems that are both effective and ethically responsible.

4.3 Multi-Step Reasoning

Reasoning has been a widely studied topic, particularly multi-step reasoning tasks that demand a robust understanding of context and logic (Wei et al., 2022). The sequential nature of CoX methods makes them ideally suited for this task. For instance, Chain-of-Knowledge[b] (Wang et al., 2023c) elicits explicit knowledge evidence at each step, thereby improving LLMs' performance in various reasoning tasks. Meanwhile, Chain-of-Feedback (Ahn and Shin, 2024) revises initial incorrect reasoning steps by breaking them down into smaller, individual tasks for more grounded reasoning. Other specialized reasoning tasks include rule-based reasoning (Servantez et al., 2024), database reasoning (Hu et al., 2023a), legal reasoning (Kuppa et al., 2023), user behavioral reasoning (Do et al., 2023; Han et al., 2024), structure and graph reasoning (Zeng et al., 2024; Luo et al., 2024; Xia et al., 2024d), as well as reasoning for text summarization (Adams et al., 2023; Bao et al., 2024) and machine translation (Lu et al., 2023). Through these varied applications, CoX methods demonstrate their ability to decompose complex tasks into manageable steps, enhancing LLMs' ability to process and analyze information effectively.

4.4 Instruction Following

Instruction following, as a celebrated abilities of LLMs, enables humans to provide explicit guidance for various tasks (Zhang et al., 2023b). The evolution of CoX methods has also led to various approaches for enhancing this feature. Chain-of-Task (Li et al., 2024b), for instance, offers a structured method where each instruction consists of intermediate atomic tasks, specifically curated to fine-tune an e-commerce LLM's responses to better meet customer needs. Extending this concept, Chain-of-Instructions (Hayati et al., 2024) introduces a compositional approach where each output feeds directly into the next, creating a continuous loop of task-specific tuning that refines the LLM's task handling progressively. For applications in speech generation, Chain-of-Modality (Zhang et al., 2023a) constructs fine-tuning sequences with concatenated textual and speech instructions. Furthermore, Chain-of-LoRA (Qiu et al., 2024) uses LoRA networks to specialize instruction handling, optimizing performances across varied tasks by tailoring instruction tuning processes to each LoRA. These developments underscore how CoX methods can improve the instruction-following capabilities of LLMs, enabling them to understand and execute tasks with more clarity.

4.5 LLMs as Agents

With the strong planning abilities, LLMs have been utilized as agents across a wide range of tasks (Xi et al., 2023b). CoX methods have been explored to further boost the planning abilities of LLM-based agents. In this vein, Chain-of-Action[a] (Zhan and Zhang, 2023) and Chain-of-ActionThought (Zhang et al., 2024c) utilize a series of planned actions to guide agent decision-making, ensuring each step is informed by the previous one. While in games like StarCraft II, Chain-of-Summarization (Ma et al., 2023) employs LLMs to summarize past observations to suggest future strategies. Chain-of-3DThought (Yamada et al., 2024) further utilizes LLM agents to compose objects in images through trial and error within a 3D simulation environment. LLMs also serve as planners in human-scene interaction tasks with Chain-of-Contacts (Xiao et al., 2024a), and in tool using tasks with Chain-of-Abstraction (Gao et al., 2024). CoX methods have also been applied to multi-agent settings, such as Chain-of-Experts (Xiao et al., 2024b) and Chain-of-Discussion (Tao et al., 2024). These methods highlights the integration of CoX methods in enhancing the multi-dimensional abilities of LLMs as autonomous and collaborative agents.

4.6 Evaluation Tools

Evaluating LLMs has become increasingly challenging as they grow more sophisticated (Chang et al., 2023), making CoX methods a valuable asset for evaluation purposes. Chain-of-Utterances prompting (Bhardwaj and Poria, 2023) is a prime example of how CoX methods can illuminate specific areas of concern, such as safety issues in scenarios where LLMs interact with potentially harmful models. This method uncovers vulnerable conversations that could lead to LLM jailbreaking, providing essential insights into the robustness of LLM safety. Besides, Chain-of-Feedback (Ahn and Shin, 2024) method has demonstrated the impact of prompts on LLM performance. By repeatedly providing LLMs with non-informative prompts like "make another attempt", researchers have observed a gradual decrease in the quality of responses. In visual reasoning, the Chain-of-Images (Meng et al., 2023) introduces a benchmark involving a sequence of images designed to progressively assess the LLM's reasoning ability. It provides a robust tool for measuring the model's capabilities in interpreting visual data. These CoX methods underscore the importance of nuanced evaluation for more thorough assessments of LLMs.

5 Future Directions

While LLMs have demonstrated remarkable abilities in step-by-step problem-solving for various tasks, several challenges remain to be addressed.

Causal Analysis on Intermediates Existing works generally focus on improving task-specific generative results. However, understanding and explaining the underlying mechanisms of LLM reasoning is also essential in realistic scenarios. For example, Wang et al. (2023d) show that LLMs may skip rational steps when generating final results. Wang et al. (2023a) observe a performance gain from CoT even with invalid rationales. These observations indicate the value of a causal analysis on how intermediate steps truly affect the final results.

Reducing Inference Cost A chain leading to the final node of generation often requires multiple sequential inference steps, which are computationally heavy and time-consuming, especially with LLMs. It would be interesting if future research could reduce the length of CoX chains while maintaining the quality of generation. For example, it would be worth studying whether the intermediate nodes of CoX could be executed in parallel or jointly within a single inference step.

Knowledge Distillation The knowledge elicited by the intermediate nodes of CoX contains fine-grained task instructions, which can benefit the training of smaller student models when using a teacher LLM for knowledge distillation. Li et al. (2023b) and Hsieh et al. (2023) have shown that the student model can effectively learn from the rationales of CoT generated by an LLM. Nonetheless, it remains an open question whether the intermediate nodes from broader CoX methods are equally informative in inspiring student learning.

End-to-End Fine-tuning One drawback of CoX is that it does not follow an end-to-end paradigm; i.e., generation errors may accumulate along the chain when self-correction (Le et al., 2024; Dhuliawala et al., 2023) is not enforced. Future research can explore fine-tuning LLMs with CoX prompting and penalizing errors from the final output. By reducing the generation errors end-to-end, we expect this will improve the quality of both the intermediate and final nodes in CoX.

6 Conclusion

This survey explored Chain-of-X methods, building upon the concept of Chain-of-Thought. By categorizing them based on nodes and tasks, we provide a comprehensive overview that highlights the potential of CoX in enhancing LLM capabilities and opens new avenues for future research. Through this survey, we aim to inspire further exploration in a deeper understanding and more creative use of CoX paradigms for LLMs.

References

Griffin Adams, Alex Fabbri, Faisal Ladhak, Eric Lehman, and Noémie Elhadad. 2023. From sparse to dense: GPT-4 summarization with chain of density prompting. In *Proceedings of the 4th New Frontiers in Summarization Workshop*, pages 68–74, Singapore. Association for Computational Linguistics.

Garima Agrawal, Tharindu Kumarage, Zeyad Alghami, and Huan Liu. 2023. Can knowledge graphs reduce hallucinations in llms?: A survey. *arXiv preprint arXiv:2311.07914*.

Jinwoo Ahn and Kyuseung Shin. 2024. Recursive chain-of-feedback prevents performance degradation from redundant prompting. *arXiv preprint arXiv:2402.02648v2*.

Mubashara Akhtar, Michael Schlichtkrull, Zhijiang Guo, Oana Cocarascu, Elena Simperl, and Andreas Vlachos. 2023. Multimodal automated fact-checking: A survey. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 5430–5448, Singapore. Association for Computational Linguistics.

Songlin Bao, Tiantian Li, and Bin Cao. 2024. Chain-of-event prompting for multi-document summarization by large language models. *International Journal of Web Information Systems*.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. 2024a. Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17682–17690.

Maciej Besta, Florim Memedi, Zhenyu Zhang, Robert Gerstenberger, Nils Blach, Piotr Nyczyk, Marcin Copik, Grzegorz Kwaśniewski, Jürgen Müller, Lukas Gianinazzi, et al. 2024b. Topologies of reasoning: Demystifying chains, trees, and graphs of thoughts. *arXiv preprint arXiv:2401.14295*.

Rishabh Bhardwaj and Soujanya Poria. 2023. Red-teaming large language models using chain of utterances for safety-alignment. *arXiv preprint arXiv:2308.09662*.

Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, et al. 2023. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*.

Zheng Chu, Jingchang Chen, Qianglong Chen, Weijiang Yu, Tao He, Haotian Wang, Weihua Peng, Ming Liu, Bing Qin, and Ting Liu. 2023. A survey of chain of thought reasoning: Advances, frontiers and future. *arXiv preprint arXiv:2309.15402*.

Shehzaad Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, and Jason Weston. 2023. Chain-of-verification reduces hallucination in large language models. *arXiv preprint arXiv:2309.11495*.

Xuan Long Do, Kenji Kawaguchi, Min Yen Kan, and Nancy F Chen. 2023. Choire: Characterizing and predicting human opinions with chain of opinion reasoning. *arXiv preprint arXiv:2311.08385*.

Zhichen Dong, Zhanhui Zhou, Chao Yang, Jing Shao, and Yu Qiao. 2024. Attacks, defenses and evaluations for llm conversation safety: A survey. *arXiv preprint arXiv:2402.09283*.

Silin Gao, Jane Dwivedi-Yu, Ping Yu, Xiaoqing Ellen Tan, Ramakanth Pasunuru, Olga Golovneva, Koustuv Sinha, Asli Celikyilmaz, Antoine Bosselut, and Tianlu Wang. 2024. Efficient tool use with chain-of-abstraction reasoning. *arXiv preprint arXiv:2401.17464*.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Peiyuan Gong and Jiaxin Mao. 2023. Coascore: Chain-of-aspects prompting for nlg evaluation. *arXiv preprint arXiv:2312.10355*.

Guangzeng Han, Weisi Liu, Xiaolei Huang, and Brian Borsari. 2024. Chain-of-interaction: Enhancing large language models for psychiatric behavior understanding by dyadic contexts. *arXiv preprint arXiv:2403.13786*.

Shirley Anugrah Hayati, Taehee Jung, Tristan Bodding-Long, Sudipta Kar, Abhinav Sethy, Joo-Kyung Kim, and Dongyeop Kang. 2024. Chain-of-instructions: Compositional instruction tuning on large language models. *arXiv preprint arXiv:2402.11532*.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alex Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8003–8017, Toronto, Canada. Association for Computational Linguistics.

Chenxu Hu, Jie Fu, Chenzhuang Du, Simian Luo, Junbo Zhao, and Hang Zhao. 2023a. Chatdb: Augmenting llms with databases as their symbolic memory. *arXiv preprint arXiv:2306.03901*.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

Hanxu Hu, Hongyuan Lu, Huajian Zhang, Wai Lam, and Yue Zhang. 2023b. Chain-of-symbol prompting elicits planning in large langauge models. *arXiv preprint arXiv:2305.10276*.

Fan Huang, Haewoon Kwak, and Jisun An. 2023. Chain of explanation: New prompting method to generate quality natural language explanation for implicit hate speech. In *Companion Proceedings of the ACM Web Conference 2023*, pages 90–93.

Jie Huang and Kevin Chen-Chuan Chang. 2023. Towards reasoning in large language models: A survey. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1049–1065, Toronto, Canada. Association for Computational Linguistics.

Qiang Huang, Feng Huang, DeHao Tao, YueTong Zhao, BingKun Wang, and YongFeng Huang. 2024. Coq: An empirical framework for multi-hop question answering empowered by large language models. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11566–11570. IEEE.

Taehee Kim, Yeongjae Cho, Heejun Shin, Yohan Jo, and Dongmyung Shin. 2024. Generalizing visual question answering from synthetic to human-written questions via a chain of qa with a large language model. *arXiv preprint arXiv:2401.06400*.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.

Aditya Kuppa, Nikon Rasumov-Rahe, and Marc Voses. 2023. Chain of reference prompting helps llm to think like a lawyer. In *Generative AI+ Law Workshop*.

Emanuele La Malfa, Christoph Weinhuber, Orazio Torre, Fangru Lin, Anthony Cohn, Nigel Shadbolt, and Michael Wooldridge. 2024. Code simulation challenges for large language models. *arXiv preprint arXiv:2401.09074*.

Hung Le, Hailin Chen, Amrita Saha, Akash Gokul, Doyen Sahoo, and Shafiq Joty. 2024. Codechain: Towards modular code generation through chain of self-revisions with representative sub-modules. In *The Twelfth International Conference on Learning Representations*.

Harrison Lee, Samrat Phatale, Hassan Mansoor, Kellie Lu, Thomas Mesnard, Colton Bishop, Victor Carbune, and Abhinav Rastogi. 2023a. Rlaif: Scaling reinforcement learning from human feedback with ai feedback. *arXiv preprint arXiv:2309.00267*.

Yoon Kyung Lee, Inju Lee, Minjung Shin, Seoyeon Bae, and Sowon Hahn. 2023b. Chain of empathy: Enhancing empathetic response of large language models based on psychotherapy models. *arXiv preprint arXiv:2311.04915*.

Deren Lei, Yaxi Li, Mingyu Wang, Vincent Yun, Emily Ching, Eslam Kamal, et al. 2023. Chain of natural language inference for reducing large language model ungrounded hallucinations. *arXiv preprint arXiv:2310.03951*.

Chengshu Li, Jacky Liang, Andy Zeng, Xinyun Chen, Karol Hausman, Dorsa Sadigh, Sergey Levine, Li Fei-Fei, Fei Xia, and Brian Ichter. 2023a. Chain of code: Reasoning with a language model-augmented code emulator. *arXiv preprint arXiv:2312.04474*.

Liunian Harold Li, Jack Hessel, Youngjae Yu, Xiang Ren, Kai-Wei Chang, and Yejin Choi. 2023b. Symbolic chain-of-thought distillation: Small models can also "think" step-by-step. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2665–2679, Toronto, Canada. Association for Computational Linguistics.

Xingxuan Li, Ruochen Zhao, Yew Ken Chia, Bosheng Ding, Shafiq Joty, Soujanya Poria, and Lidong Bing. 2024a. Chain-of-knowledge: Grounding large language models via dynamic knowledge adapting over heterogeneous sources. In *The Twelfth International Conference on Learning Representations*.

Yangning Li, Shirong Ma, Xiaobin Wang, Shen Huang, Chengyue Jiang, Hai-Tao Zheng, Pengjun Xie, Fei Huang, and Yong Jiang. 2024b. Ecomgpt: Instruction-tuning large language models with chain-of-task tasks for e-commerce. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18582–18590.

Hao Liu, Carmelo Sferrazza, and Pieter Abbeel. 2024a. Chain of hindsight aligns language models with feedback. In *The Twelfth International Conference on Learning Representations*.

Zuyan Liu, Yuhao Dong, Yongming Rao, Jie Zhou, and Jiwen Lu. 2024b. Chain-of-spot: Interactive reasoning improves large vision-language models. *arXiv preprint arXiv:2403.12966*.

Hongyuan Lu, Haoyang Huang, Dongdong Zhang, Haoran Yang, Wai Lam, and Furu Wei. 2023. Chain-of-dictionary prompting elicits translation in large language models. *arXiv preprint arXiv:2305.06575*.

Ruilin Luo, Tianle Gu, Haoling Li, Junzhe Li, Zicheng Lin, Jiayi Li, and Yujiu Yang. 2024. Chain of history: Learning and forecasting with llms for temporal knowledge graph completion. *arXiv preprint arXiv:2401.06072*.

Qing Lyu, Shreya Havaldar, Adam Stein, Li Zhang, Delip Rao, Eric Wong, Marianna Apidianaki, and Chris Callison-Burch. 2023. Faithful chain-of-thought reasoning. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 305–329, Nusa Dua, Bali. Association for Computational Linguistics.

Weiyu Ma, Qirui Mi, Xue Yan, Yuqiao Wu, Runji Lin, Haifeng Zhang, and Jun Wang. 2023. Large language models play starcraft ii: Benchmarks and a chain of summarization approach. *arXiv preprint arXiv:2312.11865*.

Fanxu Meng, Haotong Yang, Yiding Wang, and Muhan Zhang. 2023. Chain of images for intuitively reasoning. *arXiv preprint arXiv:2311.09241*.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Zhenyu Pan, Haozheng Luo, Manling Li, and Han Liu. 2024. Chain-of-action: Faithful and multimodal question answering through large language models. *arXiv preprint arXiv:2403.17359*.

Ji Qi, Ming Ding, Weihan Wang, Yushi Bai, Qingsong Lv, Wenyi Hong, Bin Xu, Lei Hou, Juanzi Li, Yuxiao Dong, et al. 2024. Cogcom: Train large vision-language models diving into details through chain of manipulations. *arXiv preprint arXiv:2402.04236*.

Xihe Qiu, Teqi Hao, Shaojie Shi, Xiaoyu Tan, and Yu-Jie Xiong. 2024. Chain-of-lora: Enhancing the instruction fine-tuning performance of low-rank adaptation on diverse instruction set. *IEEE Signal Processing Letters*.

Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. 2021. Scaling language models: Methods, analysis & insights from training gopher. *arXiv preprint arXiv:2112.11446*.

Sergio Servantez, Joe Barrow, Kristian Hammond, and Rajiv Jain. 2024. Chain of logic: Rule-based reasoning with large language models. *arXiv preprint arXiv:2402.10400*.

Mingxu Tao, Dongyan Zhao, and Yansong Feng. 2024. Chain-of-discussion: A multi-model framework for complex evidence-based question answering. *arXiv preprint arXiv:2402.16313*.

Kohei Uehara, Nabarun Goswami, Hanqin Wang, Toshiaki Baba, Kohtaro Tanaka, Tomohiro Hashimoto, Kai Wang, Rei Ito, Takagi Naoya, Ryo Umagami, et al. 2024. Advancing large multi-modal models with explicit chain-of-reasoning and visual question generation. *arXiv preprint arXiv:2401.10005*.

Boshi Wang, Sewon Min, Xiang Deng, Jiaming Shen, You Wu, Luke Zettlemoyer, and Huan Sun. 2023a. Towards understanding chain-of-thought prompting: An empirical study of what matters. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2717–2739, Toronto, Canada. Association for Computational Linguistics.

Hanbin Wang, Zhenghao Liu, Shuo Wang, Ganqu Cui, Ning Ding, Zhiyuan Liu, and Ge Yu. 2023b. Intervenor: Prompt the coding ability of large language models with the interactive chain of repairing. *arXiv preprint arXiv:2311.09868*.

Jianing Wang, Qiushi Sun, Nuo Chen, Xiang Li, and Ming Gao. 2023c. Boosting language models reasoning with chain-of-knowledge prompting. *arXiv preprint arXiv:2306.06427*.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023d. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations*.

Yufei Wang, Wanjun Zhong, Liangyou Li, Fei Mi, Xingshan Zeng, Wenyong Huang, Lifeng Shang, Xin Jiang, and Qun Liu. 2023e. Aligning large language models with human: A survey. *arXiv preprint arXiv:2307.12966*.

Zilong Wang, Hao Zhang, Chun-Liang Li, Julian Martin Eisenschlos, Vincent Perot, Zifeng Wang, Lesly Miculicich, Yasuhisa Fujii, Jingbo Shang, Chen-Yu Lee, and Tomas Pfister. 2024. Chain-of-table: Evolving tables in the reasoning chain for table understanding. In *The Twelfth International Conference on Learning Representations*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed H. Chi, Quoc V Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*.

Nan Xi, Jingjing Meng, and Junsong Yuan. 2023a. Chain-of-look prompting for verb-centric surgical triplet recognition in endoscopic videos. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 5007–5016.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. 2023b. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*.

Wenhan Xia, Chengwei Qin, and Elad Hazan. 2024a. Chain of lora: Efficient fine-tuning of language models via residual learning. *arXiv preprint arXiv:2401.04151*.

Yu Xia, Fang Kong, Tong Yu, Liya Guo, Ryan A Rossi, Sungchul Kim, and Shuai Li. 2024b. Which llm to play? convergence-aware online model selection with time-increasing bandits. *arXiv preprint arXiv:2403.07213*.

Yu Xia, Xu Liu, Tong Yu, Sungchul Kim, Ryan A Rossi, Anup Rao, Tung Mai, and Shuai Li. 2024c. Hallucination diversity-aware active learning for text summarization. *arXiv preprint arXiv:2404.01588*.

Yuwei Xia, Ding Wang, Qiang Liu, Liang Wang, Shu Wu, and Xiaoyu Zhang. 2024d. Enhancing temporal knowledge graph forecasting with large language models via chain-of-history reasoning. *arXiv preprint arXiv:2402.14382*.

Zeqi Xiao, Tai Wang, Jingbo Wang, Jinkun Cao, Wenwei Zhang, Bo Dai, Dahua Lin, and Jiangmiao Pang. 2024a. Unified human-scene interaction via prompted chain-of-contacts. In *The Twelfth International Conference on Learning Representations*.

Ziyang Xiao, Dongxiang Zhang, Yangjun Wu, Lilin Xu, Yuan Jessica Wang, Xiongwei Han, Xiaojin Fu, Tao Zhong, Jia Zeng, Mingli Song, and Gang Chen. 2024b. Chain-of-experts: When LLMs meet complex operations research problems. In *The Twelfth International Conference on Learning Representations*.

Shicheng Xu, Liang Pang, Huawei Shen, Xueqi Cheng, and Tat-seng Chua. 2023. Search-in-the-chain: Towards the accurate, credible and traceable content generation for complex knowledge-intensive tasks. *arXiv preprint arXiv:2304.14732*.

Yutaro Yamada, Khyathi Chandu, Yuchen Lin, Jack Hessel, Ilker Yildirim, and Yejin Choi. 2024. L3go: Language agents with chain-of-3d-thoughts for generating unconventional objects. *arXiv preprint arXiv:2402.09052*.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2024. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36.

Wenhao Yu, Hongming Zhang, Xiaoman Pan, Kaixin Ma, Hongwei Wang, and Dong Yu. 2023a. Chain-of-note: Enhancing robustness in retrieval-augmented language models. *arXiv preprint arXiv:2311.09210*.

Zihan Yu, Liang He, Zhen Wu, Xinyu Dai, and Jiajun Chen. 2023b. Towards better chain-of-thought prompting strategies: A survey. *arXiv preprint arXiv:2310.04959*.

Daoguang Zan, Bei Chen, Fengji Zhang, Dianjie Lu, Bingchao Wu, Bei Guan, Wang Yongji, and Jian-Guang Lou. 2023. Large language models meet NL2Code: A survey. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7443–7464, Toronto, Canada. Association for Computational Linguistics.

Qingkai Zeng, Yuyang Bai, Zhaoxuan Tan, Shangbin Feng, Zhenwen Liang, Zhihan Zhang, and Meng Jiang. 2024. Chain-of-layer: Iteratively prompting large language models for taxonomy induction from limited examples. *arXiv preprint arXiv:2402.07386*.

Liangyu Zha, Junlin Zhou, Liyao Li, Rui Wang, Qingyi Huang, Saisai Yang, Jing Yuan, Changbao Su, Xiang Li, Aofeng Su, et al. 2023. Tablegpt: Towards unifying tables, nature language and commands into one gpt. *arXiv preprint arXiv:2307.08674*.

Zhuosheng Zhan and Aston Zhang. 2023. You only look at screens: Multimodal chain-of-action agents. *arXiv preprint arXiv:2309.11436*.

Dong Zhang, Shimin Li, Xin Zhang, Jun Zhan, Pengyu Wang, Yaqian Zhou, and Xipeng Qiu. 2023a. SpeechGPT: Empowering large language models with intrinsic cross-modal conversational abilities. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 15757–15773, Singapore. Association for Computational Linguistics.

Dong Zhang, Xin Zhang, Jun Zhan, Shimin Li, Yaqian Zhou, and Xipeng Qiu. 2024a. Speechgpt-gen: Scaling chain-of-information speech generation. *arXiv preprint arXiv:2401.13527*.

Jingyi Zhang, Jiaxing Huang, Sheng Jin, and Shijian Lu. 2024b. Vision-language models for vision tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Jiwen Zhang, Jihao Wu, Yihua Teng, Minghui Liao, Nuo Xu, Xiao Xiao, Zhongyu Wei, and Duyu Tang. 2024c. Android in the zoo: Chain-of-action-thought for gui agents. *arXiv preprint arXiv:2403.02713*.

Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, et al. 2023b. Instruction tuning for large language models: A survey. *arXiv preprint arXiv:2308.10792*.

Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, et al. 2023c. Siren's song in the ai ocean: a survey on hallucination in large language models. *arXiv preprint arXiv:2309.01219*.

Zhenduo Zhang, Bowen Zhang, and Guang Liu. 2023d. Coie: Chain-of-instruct editing for multi-attribute face manipulation. *arXiv preprint arXiv:2312.07879*.

Ruochen Zhao, Hailin Chen, Weishi Wang, Fangkai Jiao, Xuan Long Do, Chengwei Qin, Bosheng Ding, Xiaobao Guo, Minzhi Li, Xingxuan Li, and Shafiq Joty. 2023. Retrieving multimodal information for augmented generation: A survey. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 4736–4756, Singapore. Association for Computational Linguistics.

Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc V Le, and Ed H. Chi. 2023. Least-to-most prompting enables complex reasoning in large language models. In *The Eleventh International Conference on Learning Representations*.

Enshen Zhou, Yiran Qin, Zhenfei Yin, Yuzhou Huang, Ruimao Zhang, Lu Sheng, Yu Qiao, and Jing Shao. 2024. Minedreamer: Learning to follow instructions via chain-of-imagination for simulated-world control. *arXiv preprint arXiv:2403.12037*.

A Taxonomies of Nodes and Tasks

We present in Figure 3 the complete version of Figure 2 on Chain-of-X taxonomies categorized by nodes and tasks.

Figure 3: A Survey of Chain-of-X by Taxonomies of Nodes and Tasks.