2023 16th International Symposium on Computational Intelligence and Design (ISCID) | 979-8-3503-1745-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISCID59865.2023.00029

# Unleashing the Second Brain: Enhancing Large Language Models through Chain of Thought with Human Feedback

Jia Yu[1,2], Minghui Luo[3], Honggu Zhou[4], *Zhenzhong Lan[1,2]
1. Zhejiang University, China
2.School of Engineering, Westlake University, China
3.Westlake Xinchen (Hangzhou) Technology Co., Ltd., China
4.School of Media and Desgin, Hangzhou Dianzi University, China
Email: yujia@westlake.edu.cn, luominghui@xinchen-inc.com,hongguzhou@hdu.edu.cn,lanzhenzhong@westlake.edu.cn

***Abstract*—The expansion of large language models has led to improved performance and efficiency, with prompt engineering emerging as a key strategy for various LLM tasks. However, while these methods have proven beneficial, they are not without their constraints, particularly when it comes to sustaining a comprehensive, logical flow of ideas throughout the entire reasoning process. Despite the effectiveness of the Chain-of-Thought (CoT) and Tree-of-Thought (ToT) methods, they have limitations due to their end-to-end reasoning process. To address this, we introduce the chain-of-thought with human feedback. This new method incorporates human feedback into the model's reasoning process, allowing for real-time adjustments and optimization. This approach fosters a more interactive and dynamic model operation, enabling the model to learn from human intuition and expertise, and improve its reasoning process over time. We have validated the effectiveness of our method through experiments in GSM8K and MMLU. Our main contributions include proposing the first human feedback to the chain of thought and the development of an intuitive interface for individuals to utilize large-scale models for problem-solving.**

***Keywords*-Chain-of-Thought, Prompting, Large Language Model, Human-AI-interaction**

## I. Introduction

The field of Natural Language Processing (NLP) has undergone a significant metamorphosis, primarily driven by the advent and subsequent influence of Large Language Models (LLMs). A wealth of empirical evidence underscores the assertion that the augmentation of these models precipitates a myriad of advantages, most notably, an enhancement in performance alongside a more nuanced utilisation of sample data. Recent years have borne witness to substantial advancements, propelled by the advent of Transformer-based variants such as GPT [1–4], PaLM[5], and LLaMA [6][7].

As these models continue to evolve, prompt engineering has emerged as an indispensable tool, demonstrating its efficacy in managing a diverse array of LLM tasks. Assuming the task description is meticulously formulated, the LLM, through its autoregressive token-based text generation mechanism, is able to execute the task. The prompt may either encompass example tasks with their corresponding solutions—termed as few-shot prompts or In-Context Learning (ICL)—or it may completely exclude example tasks, a scenario referred to as zero-shot prompts. Recent studies attest to this approach's proficiency across various tasks, from mathematics to symbolic reasoning.

In this context, CoT [8] represents a trailblazing advancement in the realm of prompt engineering, enhancing LLM problem-solving by simulating human cognitive processes without model updates. However, CoT's application is constrained to a singular intermediate process. ToT [9–11], in an endeavor to address this limitation, abstracts the intermediate process into a tree a multi-branch data structure thereby more accurately simulating human cognitive processes.

Despite these significant advancements, a persistent challenge with these methods is the end-to-end nature of the model's reasoning process. This implies that any misjudgment during the intermediate process could result in an inflated computation time and, more critically, yield inaccurate results. To circumvent this issue, we propose a novel approach—chain-of-thought with human feedback. This innovative strategy weaves human feedback into the model's reasoning process, serving as a corrective mechanism during the intermediate steps of the model's computation. This allows for real-time adjustments and optimization.

Our research commenced with the presentation of a well-defined problem to the Large Language Model (LLM). In response, the LLM proposed a set of potential initial steps towards resolving the problem. Human intervention was then integrated into the process, with the responsibility of selecting the most appropriate step from the proposed alternatives. This chosen step was then reintroduced into the model as an input, which in turn prompted the model to generate three additional subsequent steps. This iterative process persisted until a solution was deduced and computed. For a comprehensive understanding of this process, please refer to Figure1.

The chain-of-thought with human feedback allows for a more interactive and dynamic model operation. Instead of the model working in isolation, it can now engage in a back-and-forth exchange with humans, gaining insights and corrections as it progresses through the problem-solving process. This interaction can lead to more accurate and efficient results, as

2473-3547/23/$31.00 ©2023 IEEE
DOI 10.1109/ISCID59865.2023.00029

the model can leverage human intuition and expertise in areas where it might struggle[12]. In essence, the chain-of-thought with human feedback not only addresses the limitations of the end-to-end reasoning process in large language models but also introduces a more collaborative and adaptive approach to problem-solving in the field of Natural Language Processing.

We have conducted to validate the effectiveness of our method in GSM8K [13] and MMLU [14]. Furthermore, a sequence of ablation experiments have validated the impact of the proposed human feedback on our method. The main contributions of our approach are as follows:

- We are the first to propose adding human feedback to the chain of thought.
- We have engineered an intuitive, user-friendly interface designed to empower individuals to leverage large language models for problem-solving, without necessitating any technical proficiency.

## II. Related Work

### A. Large Language Models

There's been a surge in interest in large-scale language models, utilizing deep learning algorithms like the Transformer [15] to train vast text datasets and generate human-like text. Models like OpenAI's GPT-3 [1] and GPT-4 [16], Google's Palm[5] and T5[17], Anthropic's Claude [18, 19], and Facebook's BART [20] and Llama [6] have shown impressive performance across various tasks, highlighting the effectiveness of these large language models (LLMs) in natural language processing tasks.

### B. prompting approaches

The Input-Output (IO) method is a direct approach where a Large Language Models (LLMs) transforms an input sequence into an output. The Chain-of-Thought (CoT) [8] introduces intermediate thoughts for improved reasoning, but its extension, Self-Consistency with CoT (CoT-SC) [21], lacks local path exploration. The Tree of Thought (ToT) [9] models reasoning as a tree of partial solutions, guided by a search algorithm. However, all these methods lack control over thought quality, risking incorrect answers or inefficiencies.

### C. Human-LLM Interaction

Navigating the complex landscape of human-large language model (LLM) interaction interfaces, one encounters a rich tapestry of research threads[22]. The evolution of LLMs saw a shift towards context-aware systems, aiming for a more personalized user interaction[23]. This progress, however, unveiled a new set of challenges related to privacy and data security, reflecting the intricate balance needed in the field.The increasing complexity of LLMs has catalyzed the emergence of explainable AI, aiming to foster user trust by enhancing transparency in decision-making processes[24][25]. Yet, achieving a harmonious balance between model complexity and explainability remains a contemporary conundrum. The current state of AI interaction interfaces explores various paradigms, including conversational AI interfaces, mixed reality interfaces, and brain-computer interfaces, each with their unique strengths and weaknesses in terms of efficiency, effectiveness, learnability, and user satisfaction. Another critical area of research delves into the cognitive aspects of human-LLM interaction, examining how concepts like mental models, cognitive load, and attention influence the design and effectiveness of interaction interfaces[26][27]. Evaluation of these interfaces typically involves metrics such as task completion time, error rates, user satisfaction scores, and learnability. The field of human-LLM interaction interfaces[28] remains vibrant and evolving, with numerous uncharted territories that beckon further exploration and empirical investigation.

## III. Methods

Studies into human problem-solving strategies indicate that individuals navigate a complex combinatorial problem space. This space can be visualized as a chain, tree or graph structure, where nodes symbolize partial solutions, and the branches represent the operations which may also be solutions [29–31]. The choice of which branch to follow is guided by heuristics for humans.This raises a question: humans cannot provide intuitive choices for difficult abstractions. Take the process of solving complex algebraic equations as an example. Although the ultimate goal is clear (finding the value of the variable), the steps to achieve this goal may involve a series of abstract operations. Without prompts, it is difficult for individuals at an average cognitive level to come up with these abstract processes.

In pursuit of understanding whether the concretization of cognitive processes can enhance human decision-making, we have devised a thought-chain model with human feedback. The crux of this model utilizes LLMs to generate prospective intermediate steps, offering humans a selection of paths for problem-solving. During this process, human users can exercise subjective control over the problem-solving journey by selecting from the array of intermediate steps generated by the LLMs. This approach allows users to choose the problem-solving path that best aligns with their unique perspectives and understanding. Subsequently, the selected intermediate steps are used as new inputs for the LLMs to infer the next round of reasoning. This process is iteratively repeated until a final solution is reached. This iterative method facilitates a deeper understanding of the problem for the user and incrementally constructs a complete path to the solution[28]. Through this approach, we can concretize the abstract cognitive processes, enabling users to make more informed decisions throughout the problem-solving process. This method not only provides a new lens to understand human decision-making but also offers a novel tool to enhance human interaction with Large Language Models.

## IV. Experience

### A. Experimental Setup

We explore chain-of-thought with human feedback for various language models on multiple benchmarks.

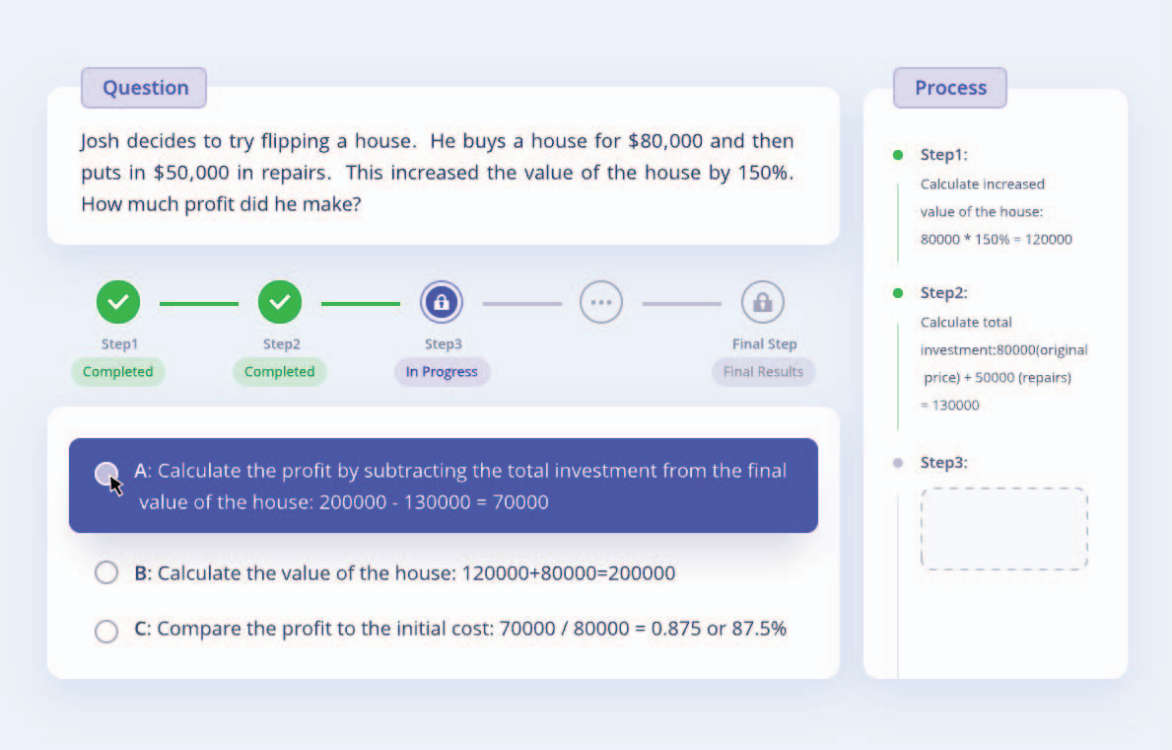


Fig. 1: User Interface for Chain-of-Thought with Human Feedback

TABLE I: The results on GSM8K and MMLU. † represents the average accuracy of four subsets.

| Method | Params | GSM8K | | | MMLU† | | |
|---|---|---|---|---|---|---|---|
| | | ST | CoT | Cot w/ HF | ST | CoT | Cot w/ HF |
| GPT-4 | ? | - | 92.0 | **96.2** | 85.3 | 90.8 | **91.7** |
| GPT-3.5-Turbo | ? | 57.1 | 74.9 | **81.5** | 70.4 | 79.8 | **83.2** |
| Llama2 | 70B | 56.8 | 63.5 | **70.8** | 53.7 | 62.1 | **60.2** |
| Llama | 65B | 50.9 | 65.1 | **66.6** | 47.4 | 56.9 | **52.4** |

**Benchmarks.** Two types of datasets were considered, which include math and general knowledge questions: (1) **GSM8K**, represents a meticulously curated dataset encompassing 8.5K linguistically diverse, high-quality grade school math word problems, crafted by expert human problem writers. The complexity of these problems varies, requiring between 2 and 8 steps to reach a solution. (2) **MMLU**, encompasses 57 subjects spanning STEM, humanities, social sciences and others. The benchmark assesses a range of difficulty levels, from elementary to advanced professional, evaluating both knowledge of the world and problem-solving skills. Given the vast volume of tasks in the MMLU and the associated cost of human feedback, we have selected a subset of tasks from each of the four categories within the MMLU to explore the impact of human feedback on large-scale multi-task language understanding beyond GSM8K. Specifically, we chose Abstract Algebra from the STEM category, Formal Logic from the Humanities, High School Microeconomics from the Social Sciences, and Global Facts from the Others category to evaluate the effectiveness of human feedback, a total of 400 problems approximately were involved.

**Language models.** We conducted an evaluation of two language models. The GPT series was included in our evaluation, specifically GPT4 and GPT3.5. Llama, specifically the 65B model, as well as the latest Releases Llama2 75B was also evaluated. To ensure a fair comparison, we followed the protocol of CoT and employed greedy decoding for sampling from the models.

**Baselines.** Our methodology is based on two baselines: (1) The Standard Prompt, an extension of the few-shot prompt paradigm, where the language model predicts outputs based

on given input-output pairs within context samples [1]. (2) The Chain of Thought (CoT) prompting, which enhances LLMs' reasoning capabilities by generating a series of intermediary reasoning steps, forming a cognitive chain. With examples of this chain provided as prompts, the model is guided to generate both the cognitive process and the solution.

**Setup and participants.** 103 undergraduates and postgraduates from the Zhejiang University were invited to participate in this experiment (54 males and 49 females). The age of all subjects ranged from 22 to 26 years old, with an average age of 23.8 years (SD = 1.1536), of which the average age of males was 23.6 (SD =0.9445) and the average age of females was 24.0 (SD = 1.2415). They received a fixed amount of remuneration after the end of the experiment, Due to the extensive size of the dataset, all participants engaged in an experiment spanning two weeks, they completed 100 tasks each trial at a day, Every participant was mandated to accomplish 14 trials in total, 1k questions for GSM8K and about 400 questions for MMLU respectively. Dell SE2419H_HX LCD monitors were used in this experiment. The monitor size was 24 and the resolution was 1920 x 1080 pixels. The monitor was placed 65cm in front of the subjects' eyes, and the experiment was carried out in a quiet environment with sufficient lighting.

The task was architectured to facilitate an interaction with LLMs. Initially, each participant is presented with a problem statement accompanied by a triad of options on their screen. The participant's role is to identify and select the option that best aligns with the coherent flow of chain of thought, it continues until the LLM infers an answer. It is incumbent upon participants to finalize their selection within a 30-second window, unsuccessful attempts fall outside this time frame. Chosen reasoning steps appear in the interface's upper-right quadrant. Upon successful inference of an answer, the participant is transitioned to the subsequent question with a 30-second interval between each. The task culminates with each participant answering a comprehensive set of 100 questions at one trial.

### B. Result

Table I delineates the results of chain-of-thought with human feedback. Our findings align with the concept of chain-of-thought prompting as an emergent capability of large models, as proposed by [32]. Typically, chain-of-thought prompting improves performance on a range of arithmetic, commonsense, and symbolic reasoning tasks.

Secondly, the chain-of-thought with human feedback achieved greater performance improvements compared to CoT . On GSM8K, the chain-of-thought with human feedback demonstrated a 4.7% improvement over CoT in GPT4, 8.8% improvement over CoT in GPT3.5. as well as 11.4% and 2.3% improvements over CoT in Llama2 and Llama respectively. On MMLU, the chain-of-thought with human feedback showcased a 0.9% improvement over CoT in GPT4, 3.4% improvement over CoT in GPT3.5, however, we found and 1.9% decrease over CoT in Llama2 and 4.5% decrease over CoT in Llama.

### C. Ablation Study

The utilization of Chain-of-thought with human feedback naturally precipitates the question of whether comparable performance enhancements can be conferred through other types of prompts. Table II presents an ablation study, incorporating variations of two thought processes described below.

**Comparison with Humans.** In addition to engaging 103 individuals to conduct our human-feedback-based experiments, we also invited a diverse group of ten participants across different age groups to perform tasks on GSM8K and MMLU, with the comparative results depicted in Table II. The findings indicate that, despite the complex logical reasoning required by GSM8K and the domain-specific knowledge questions posed by MMLU, which present considerable challenges to the participants, the use of a human-feedback-based chain of thought can still guide individuals to make the correct choices, even in scenarios where they initially lack the necessary knowledge.

| Methods | Params | GSM8K | MMLU† |
|---|---|---|---|
| LLama w/ HF | 65B | **66.6** | **52.4** |
| Humans | 0 | 65.5 | 36.3 |
| Llama w/ random | 65B | 7.1 | 13.6 |

TABLE II: Ablation Study on Humans and Random Intermediate Thought. We are presenting the results for Llama, as conducting additional queries for GPT are constrained due to their high cost and limited availability. And † represents the average accuracy of four subsets.

**Random Intermediate Thought.** Our study aimed to elucidate the role of human feedback, focusing on a comparison between structured and randomized feedback. We introduced modifications to the prompts, facilitating the generation of randomized chain of thought by Large Language Models (LLMs) for subsequent human evaluation. This process, founded on human selection, was iteratively repeated to produce a final result. Experimental findings demonstrated that the group exposed to the random feedback realized a 7.1% accuracy rate on the GSM8K dataset. This suggests that even with the integration of human feedback, maintaining an accurate context is essential for LLMs to make correct inferences.

## V. Conclusion

This study introduces a novel method for Large Language Models (LLMs) application in Natural Language Processing (NLP)—a chain-of-thought with human feedback. This method surpasses traditional CoT and ToT methods by effectively addressing reasoning errors and allowing real-time optimization. Human feedback enhances the model's problem-solving abilities by providing human intuition and expertise. The effectiveness of this approach is confirmed by significant improvements in GSM8K and MMLU benchmarks and validated through ablation experiments. Our research offers

a fresh perspective in NLP, demonstrating that human feedback integration leads to more effective problem-solving. This method overcomes the limitations of end-to-end reasoning in LLMs and promotes a more cooperative and adaptive problem-solving approach, potentially paving the way for future research to improve LLM performance in NLP tasks.

## REFERENCES


[1] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[2] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever *et al.*, "Improving language understanding by generative pre-training," 2018.

[3] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg *et al.*, "Sparks of artificial general intelligence: Early experiments with gpt-4," *arXiv preprint arXiv:2303.12712*, 2023.

[4] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[5] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann *et al.*, "Palm: Scaling language modeling with pathways," *arXiv preprint arXiv:2204.02311*, 2022.

[6] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[7] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale *et al.*, "Llama 2: Open foundation and fine-tuned chat models," *arXiv preprint arXiv:2307.09288*, 2023.

[8] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 824–24 837, 2022.

[9] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, and K. Narasimhan, "Tree of thoughts: Deliberate problem solving with large language models," *arXiv preprint arXiv:2305.10601*, 2023.

[10] J. Long, "Large language model guided tree-of-thought," *arXiv preprint arXiv:2305.08291*, 2023.

[11] Y. Xie, K. Kawaguchi, Y. Zhao, X. Zhao, M.-Y. Kan, J. He, and Q. Xie, "Decomposition enhances reasoning via self-evaluation guided decoding," *arXiv preprint arXiv:2305.00633*, 2023.

[12] Q. Yang, J. Cranshaw, S. Amershi, S. T. Iqbal, and J. Teevan, "Sketching nlp: A case study of exploring the right things to design with language intelligence," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[13] K. Cobbe, V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano *et al.*, "Training verifiers to solve math word problems," *arXiv preprint arXiv:2110.14168*, 2021.

[14] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt, "Measuring massive multitask language understanding," *arXiv preprint arXiv:2009.03300*, 2020.

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[16] OpenAI, "Gpt-4 technical report," 2023.

[17] H. W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, E. Li, X. Wang, M. Dehghani, S. Brahma *et al.*, "Scaling instruction-finetuned language models," *arXiv preprint arXiv:2210.11416*, 2022.

[18] A. Askell, Y. Bai, A. Chen, D. Drain, D. Ganguli, T. Henighan, A. Jones, N. Joseph, B. Mann, N. DasSarma *et al.*, "A general language assistant as a laboratory for alignment," *arXiv preprint arXiv:2112.00861*, 2021.

[19] Y. Bai, A. Jones, K. Ndousse, A. Askell, A. Chen, N. DasSarma, D. Drain, S. Fort, D. Ganguli, T. Henighan *et al.*, "Training a helpful and harmless assistant with reinforcement learning from human feedback," *arXiv preprint arXiv:2204.05862*, 2022.

[20] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[21] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," *arXiv preprint arXiv:2203.11171*, 2022.

[22] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen *et al.*, "Guidelines for human-ai interaction," in *Proceedings of the 2019 chi conference on human factors in computing systems*, 2019, pp. 1–13.

[23] Q. Yang, A. Steinfeld, C. Rosé, and J. Zimmerman, "Re-examining whether, why, and how human-ai interaction is uniquely difficult to design," in *Proceedings of the 2020 chi conference on human factors in computing systems*, 2020, pp. 1–13.

[24] S. S. Sundar, "Rise of machine agency: A framework for studying the psychology of human–ai interaction (haii)," *Journal of Computer-Mediated Communication*, vol. 25, no. 1, pp. 74–88, 2020.

[25] G. Dove, K. Halskov, J. Forlizzi, and J. Zimmerman, "Ux design innovation: Challenges for working with machine learning as a design material," in *Proceedings of the 2017 chi conference on human factors in computing systems*, 2017, pp. 278–288.

[26] Y. Zhang, Q. V. Liao, and R. K. Bellamy, "Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 295–305.

[27] G. Bansal, B. Nushi, E. Kamar, W. S. Lasecki, D. S. Weld, and E. Horvitz, "Beyond accuracy: The role of mental models in human-ai team performance," in *Proceedings of the AAAI conference on human computation and crowdsourcing*, vol. 7, no. 1, 2019, pp. 2–11.

[28] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the ai: informing design practices for explainable ai user experiences," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–15.

[29] A. Newell, H. A. Simon *et al.*, *Human problem solving*. Prentice-hall Englewood Cliffs, NJ, 1972, vol. 104, no. 9.

[30] A. Newell, J. C. Shaw, and H. A. Simon, "Report on a general problem solving program," in *IFIP congress*, vol. 256. Pittsburgh, PA, 1959, p. 64.

[31] K. Friston, "Hierarchical models in the brain," *PLoS computational biology*, vol. 4, no. 11, p. e1000211, 2008.

[32] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler *et al.*, "Emergent abilities of large language models," *arXiv preprint arXiv:2206.07682*, 2022.

## APPENDIX

We have concurrently conducted a visual comparison of standard prompting, the Chain-of-Thought (CoT), and our method, as depicted in Figure 2.


Fig. 2: An illustration of our Chain of thought with human feedback compared with Standard Prompting and Chain-of-thought prompting