LEARNING TO REASON WITH AUTOREGRESSIVE IN-CONTEXT DISTILLATION

Yuxuan Liu
Peking University
yx.liu@stu.pku.edu.cn

ABSTRACT

We investigate the joint distillation of in-context learning and reasoning from advanced large language models (LLMs) to their smaller counterparts. We introduce Autoregressive In-Context Distillation (AICD), a simple yet effective paradigm for this purpose. AICD employs meta-teacher forcing on chain-of-thought (CoT) examples and leverages the autoregressive nature of LLMs to jointly optimize the likelihood of all rationales in-context. Experiments on both mathematical and commonsense reasoning tasks demonstrate the efficacy of AICD. Furthermore, AICD enhances the capability of student LLMs in generating meaningful CoTs.

1 INTRODUCTION

The rise of LLMs (OpenAI, 2023) gives birth to a plethora of promising emergent capabilities, like in-context learning (Brown et al., 2020) and complex reasoning (Kojima et al., 2022). Limited by the computation budget, it remains empirically heavy to host and tune LLMs. To overcome this limitation, an emerging line of work study distilling these desirable capabilities to its smaller LMs with knowledge distillation (Hsieh et al., 2023; Wang et al., 2023c; Mukherjee et al., 2023). On this front, a plethora of work delved into distilling and specializing smaller models on a desired task or capability, including instruction following (Taori et al., 2023; Xu et al., 2023; Zheng et al., 2023)), math reasoning (Wang et al., 2023a; Hsieh et al., 2023; Luo et al., 2023a; Yu et al., 2024), coding Luo et al. (2023b) etc. However, these current works remain on distilling a *single* capability, i.e. task performance on a skill set or domains.

As a step beyond, this work presents a novel preliminary study on *joint* distillation of in-context learning and reasoning capability. Our primary focus lies in seeking a *unified learning objective* that connects the two distilling goals. Inspired by in-context tuning approaches (Min et al., 2022; Gu et al., 2023)[1], we adapt in-context exemplars when distilling task-related examples from LLM to smaller models. We present AICD, namely Autoregressive In-Context Distillation, an objective to align these two goals of distillation. Specifically, AICD performs meta teacher-forcing on each in-context example, and jointly trains the likelihood of all generated reasoning chains in-context. Leveraging the autoregressive nature of LMs, we achieve a one-pass optimization, better utilizing the long context of modern language models. Furthermore, the structured few-shot format also fosters the capability of in-context learning of student language models. Comprehensive experiments on both math and commonsense reasoning demonstrate the effectiveness of AICD.

2 AUTOREGRESSIVE IN-CONTEXT DISTILLATION

To overcome the limitations of in-context tuning approaches above, we propose **A**utoregressive **I**n-**C**ontext **D**istillation (AICD), a training objective that aligns the distillation of in-context learning and reasoning. Denote $\{x_i, y_i\}_{i=1}^n$ a n-shot training instance sampled from dataset $\mathcal{D}$. AICD is formulated as:

$$\mathcal{L}_{\text{AICD}} = \sum_{i=1}^{n} \omega_i \log p(y_i | x_1, y_1, ..., x_i), \tag{1}$$

[1]We discuss the limitations of these existing methods in Appendix A.

Models	GSM8K	SVAMP	MultiArith	StrategyQA
Vanilla Performance				
GPT-J 6B	2.7	20.7	9.0	47.2
In-Context Tuning Methods				
Finetune w/o Chain-of-Thoughts	8.4	32.3	18.6	61.9
STaR Zelikman et al. (2022) ‡	10.7	26.7	53.9	60.0
Meta-ICL (*distill w/ Cond. Gen. Loss*)	21.2	47.0	76.4	62.2
P-ICL (*distill w/ LM Loss*)	22.2	48.3	75.8	60.8
AICD (Ours)	**23.6**	**51.7**	**80.9**	**63.3**

Table 1: Performance of AICD objective on distilling math and commonsense reasoning capability in-context. ‡ denote results cited from (Wang et al., 2023c).

where ω_i denote weights for a pair of in-context sample (x_i, y_i). Note that y_i represents a rationale from LLM. We highlight the key improvements of AICD over existing objectives as follows:

1) **No Semantic Shifts.** P-ICL implements in-context tuning as a pre-training task through utilizing the language modeling loss (i.e., $\log p\left(x_1, y_1, \cdots, x_n, y_n\right)$). However, since the in-context examples are manually assembled, the oracle probability of transition from y_i to x_{i+1} is low, which leads to an undesired shift from pre-training semantics. Worse, Xie et al. (2021) proves that such a low likelihood between transitions contributes to in-context learning, which should be kept.
2) **Training Efficiency.** Compared with the vanilla conditional generation learning objective utilized by Meta-ICL and the most others - $\log p\left(y_i | x_1, y_1, \cdots, x_{i-1}, y_{i-1}, x_i\right)$, AICD simultaneously trains a superposition of n in-context samples, with one back-propagation call, thanks to the autoregressive nature of LMs. We figuratively denote this feature as *meta-teacher forcing*. This enables a better utilization of long context of modern LMs.
3) **Weight Assignment.** As in Eq.1, each in-context exemplar is assigned with a distinct weight ω. This enables calibration on the focus of in-context learning, which helps to improve the generalization capability of student models.

3 Experiments and Discussion

We select GPT-J 6B (Wang & Komatsuzaki, 2021) as student LM, and `GPT-3.5-turbo` as teacher model for obtaining CoTs. We experiment on GSM-8K (Cobbe et al., 2021), SVAMP (Patel et al., 2021) and MultiArith (Roy & Roth, 2016) for math, and StrategyQA (Geva et al., 2021) for commonsense reasoning. We set ω_i to 0.1 for $i{<}n$ and $\omega_n = 1$, in-context size $n = 4$ for math and 6 for commonsense, learning rate of $7e-6$, and a maximum of 10 training epochs. All models are trained on 8 NVIDIA V100 GPUs in FP16, with a global batch size of 16. Please refer to Appendix B for detailed experimental setup.

As illustrated in Table 1, AICD significantly outperforms existing approaches like STaR and tuning w/o CoT. Besides, AICD also outperforms models trained with conditional generation loss or LM loss, demonstrating its superiority over existing baselines. Similarly, AICD also improves the performance of student model on commonsense reasoning, by achieving superior results on StrategyQA. These results highlight the generalization capability of the proposed AICD learning objective. We conduct further experiments on the improvement on in-context learning, interpretability of generated CoTs, and sensitivity to hyperparameters, which we defer to Appendix C doe to space limitations.

4 Conclusion

We present Autoregressive In-context Distillation (AICD), a novel learning objective for jointly distilling reasoning capability and in-context learning. AICD perform meta-teacher forcing on chain-of-thoughts of in-context examples, and leverages the autoregressive nature of LLMs to jointly optimize the likelihood of all rationales simultaneously in-context. Experiments on both math reasoning and commonsense reasoning tasks demonstrate the effectiveness of proposed AICD. We believe AICD would further contribute to the realm of LLM specialization and distillation.

URM STATEMENT

The authors acknowledge that the sole author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

REFERENCES

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Yanda Chen, Ruiqi Zhong, Sheng Zha, George Karypis, and He He. Meta-learning via language model in-context tuning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 719–730, 2022.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Mor Geva, Daniel Khashabi, Elad Segal, Tushar Khot, Dan Roth, and Jonathan Berant. Did aristotle use a laptop? a question answering benchmark with implicit reasoning strategies. *Transactions of the Association for Computational Linguistics*, 9:346–361, 2021.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Pre-training to learn in context. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 4849–4870, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.267. URL https://aclanthology.org/2023.acl-long.267.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. *arXiv preprint arXiv:2305.02301*, 2023.

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

Haipeng Luo, Qingfeng Sun, Can Xu, Pu Zhao, Jianguang Lou, Chongyang Tao, Xiubo Geng, Qingwei Lin, Shifeng Chen, and Dongmei Zhang. Wizardmath: Empowering mathematical reasoning for large language models via reinforced evol-instruct. *arXiv preprint arXiv:2308.09583*, 2023a.

Ziyang Luo, Can Xu, Pu Zhao, Qingfeng Sun, Xiubo Geng, Wenxiang Hu, Chongyang Tao, Jing Ma, Qingwei Lin, and Daxin Jiang. Wizardcoder: Empowering code large language models with evol-instruct. *arXiv preprint arXiv:2306.08568*, 2023b.

Sewon Min, Mike Lewis, Luke Zettlemoyer, and Hannaneh Hajishirzi. Metaicl: Learning to learn in context. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2791–2809, 2022.

Subhabrata Mukherjee, Arindam Mitra, Ganesh Jawahar, Sahaj Agarwal, Hamid Palangi, and Ahmed Awadallah. Orca: Progressive learning from complex explanation traces of gpt-4. *arXiv preprint arXiv:2306.02707*, 2023.

OpenAI. Gpt-4 technical report. *ArXiv*, abs/2303.08774, 2023.

Arkil Patel, Satwik Bhattamishra, and Navin Goyal. Are NLP models really able to solve simple math word problems? In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2080–2094, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.168. URL https://aclanthology.org/2021.naacl-main.168.

Subhro Roy and Dan Roth. Solving general arithmetic word problems. *arXiv preprint arXiv:1608.01413*, 2016.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. Stanford alpaca: An instruction-following llama model. https://github.com/tatsu-lab/stanford_alpaca, 2023.

Megh Thakkar, Tolga Bolukbasi, Sriram Ganapathy, Shikhar Vashishth, Sarath Chandar, and Partha Talukdar. Self-influence guided data reweighting for language model pre-training. *arXiv preprint arXiv:2311.00913*, 2023.

Ben Wang and Aran Komatsuzaki. Gpt-j-6b: A 6 billion parameter autoregressive language model, 2021.

Peifeng Wang, Zhengyang Wang, Zheng Li, Yifan Gao, Bing Yin, and Xiang Ren. Scott: Self-consistent chain-of-thought distillation. *arXiv preprint arXiv:2305.01879*, 2023a.

Xinyi Wang, Wanrong Zhu, and William Yang Wang. Large language models are implicitly topic models: Explaining and finding good demonstrations for in-context learning. *arXiv preprint arXiv:2301.11916*, 2023b.

Zhaoyang Wang, Shaohan Huang, Yuxuan Liu, Jiahai Wang, Minghui Song, Zihan Zhang, Haizhen Huang, Furu Wei, Weiwei Deng, Feng Sun, et al. Democratizing reasoning ability: Tailored learning from large language model. *arXiv preprint arXiv:2310.13332*, 2023c.

Sang Michael Xie, Aditi Raghunathan, Percy Liang, and Tengyu Ma. An explanation of in-context learning as implicit bayesian inference. In *International Conference on Learning Representations*, 2021.

Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, and Daxin Jiang. Wizardlm: Empowering large language models to follow complex instructions. *arXiv preprint arXiv:2304.12244*, 2023.

Longhui Yu, Weisen Jiang, Han Shi, Jincheng YU, Zhengying Liu, Yu Zhang, James Kwok, Zhenguo Li, Adrian Weller, and Weiyang Liu. Metamath: Bootstrap your own mathematical questions for large language models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=N8N0hgNDRt.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, 35:15476–15488, 2022.

Yiming Zhang, Shi Feng, and Chenhao Tan. Active example selection for in-context learning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 9134–9148, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. URL https://aclanthology.org/2022.emnlp-main.622.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. Judging llm-as-a-judge with mt-bench and chatbot arena. *arXiv preprint arXiv:2306.05685*, 2023.

A PRELIMINARIES

First introduced in (Kojima et al., 2022), chain-of-thoughts (or reasoning paths) demonstrated strong empirical merits in guiding LLMs towards correct reasoning. Consequently, existing works on reasoning distillation primarily focus on the learning of chained reasoning capability from LLM (Wang et al., 2023a; Hsieh et al., 2023; Wang et al., 2023c; Yu et al., 2024; Luo et al., 2023a), which we consider a good starting point. Below, we briefly summarize and discuss related works on in-context fine-tuning and/or distillation.

In-Context Tuning As the earliest and most straightforward method, Meta-ICL (Min et al., 2022) and In-Context Tuning (Chen et al., 2022) treat a set of in-context demonstrations as a conditional prefix during training, and maximize the following log-likelihood maximization objective:

$$\log P\left(y_{k+1}|x_1, y_1, \cdots, x_k, y_k, x_{k+1}\right).$$

Essentially, this objective treats in-context demonstrations as a conditional prefix, and one may argue whether this objective is equal to 'learning to learn in context', since no explicit clues other than the conditional prefix are given.

Channel In-Context Tuning As an alternative to In-Context Tuning, Min et al. (2022) alters the order of x and y, and train to maximize the negative log-likelihood of x_{k+1} given y_{k+1} and precursor (y, x) pairs. The predicted class is selected w.r.t the largest probability of channel $\mathcal{C}$ during inference:

$$\text{argmax}_{c\in\mathcal{C}}\, P\left(x|y_1, x_1, \cdots, y_k, x_k, c\right).$$

Although this method demonstrates empirical improvements over In-Context Tuning, it increases the inference cost from $\mathcal{O}(1)$ to $\mathcal{O}(\mathcal{C})$, and could only be applied to niche scenarios where $|\mathcal{C}|$ are finite and small in scale (e.g. True/False classification). Therefore, it could not be applied to broader scenarios where we have sequential outputs.

In-Context Pre-Training / Language Modeling Instead of regarding demonstration pairs as conditions, P-ICL (Gu et al., 2023) treat in-context tuning as a language modeling task. Specifically, a language modeling loss over all demonstrations

$$\log P\left(x_1, y_1, \cdots, x_k, y_k, x_{k+1}, y_{k+1}\right)$$

is applied, under the motivation that the intrinsic tasks are already in the natural language format. However, this loss forms alter in-context tuning to a pertaining task (Gu et al., 2023), which demands significantly larger amounts of data and greater training steps. Such limitation is even crucial when training data is limited and costly to obtain (e.g. annotated reasoning chains).

B DETAILED EXPERIMENTAL SETUP

Models We select GPT-J 6B (Wang & Komatsuzaki, 2021) as student LM, and `GPT-3.5-turbo` as teacher model for obtaining CoTs. We apply the following prompt template:

"*Question:* {*question*} *Hint: The answer should be* {*answer*}*. Answer: Let's think step by step.*"

where the correct answer is prompted to improve the quality of teacher's annotations. 3 diverse CoTs are collected for an arbitrary sample, with temperature sampling at 1.0. All in-context training examples are wrapped within a simple prompt "*Question:*x *Answer:*y ..." during training.

Datasets We test AICD on two reasoning domains: math reasoning (solving math word problems) and commonsense reasoning (true/false classification based on commonsense facts). We select GSM-8K (Cobbe et al., 2021), SVAMP (Patel et al., 2021) and MultiArith (Roy & Roth, 2016) for math, and StrategyQA (Geva et al., 2021) for commonsense reasoning.

Implementation We set ω_i to 0.1 for $i<n$ and $\omega_n = 1$, learning rate of $7e-6$, and a maximum of 10 training epochs. We set in-context size $n = 4$ for math reasoning and 6 for commonsense reasoning. All models are trained on 8 NVIDIA V100 GPUs in FP16, with a global batch size of 16, cosine decay, and Adam optimizer. Code is available at `https://anonymous.4open.science/r/COT-1892`.

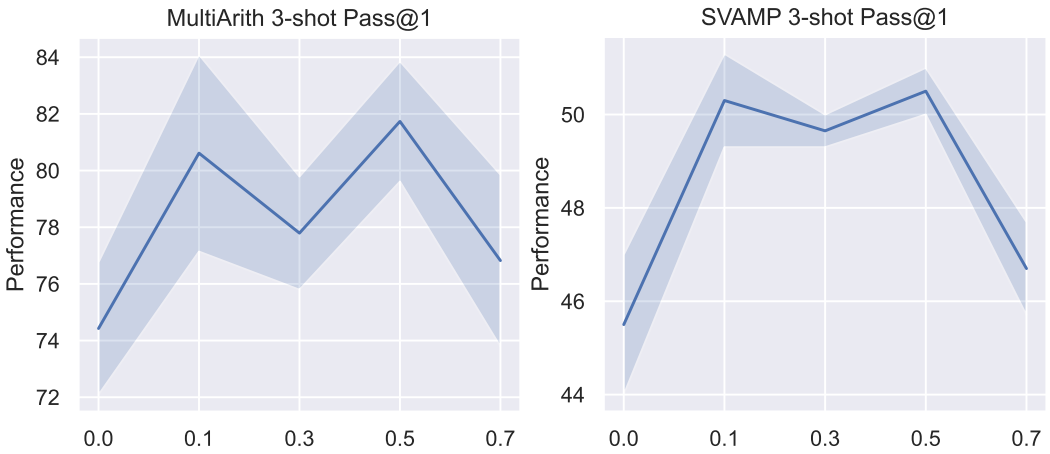

Figure 1: Performance of AICD on MultiArith and SVAMP, varying α.

C Additional Experimental Results

To test the effect of AICD in improving ICL capability, we compare AICD-tuned models against two baselines:

1) vanilla ICL, where we do not fine-tune the student model on any supervised data

2) Fine-tune the student model with in-context examples, with conditional generation objective.

As illustrated in Figure 2, AICD-tuned model significantly outperforms vanilla model with ICL, and also ones tuned with conditional generation objective. Specifically, the performance at 0,1,2,3 shots kept increasing, demonstrating the effectiveness of AICD training. Noteworthy, AICD's performance on 1 and 2 shots largely outperforms baselines, indicating that the AICD objective positively contributes to the learning of ICL capabilities, since it achieves better performance with fewer samples. These results demonstrate that AICD successfully aligns the learning of ICL and reasoning.

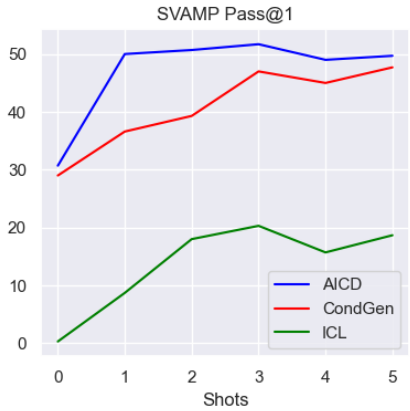

Figure 2: ICL performance varying in-context sizes after tuning w/ AICD, on SVAMP.

Finally, we study the sensitivity of AICD to the weight w of in-context samples. As illustrated in Figure 1, most AICD variants outperforms the baseline (i.e. assigning zero weights). Particularly, a smaller (0.1) or medium (0.5) weighted paradigm yields stronger results, while the latter has improved generalizability to different in-context samples (smaller variance).

We believe such configuration better balances the learning of demonstration samples as well as the final one (which is most crucial to the test results), leading to better distillation.

D Case Study

To further study the effect of AICD, we present a case study in Table 2. As shown in the table, models trained with AICD could generate more reasonable and factually correct reasoning chains, showing AICD's advantage in batter utilizing in-context samples and producing high-quality reasoning.

Variant	Generated Reasoning Chain and Answer
Case #1	*Is calling ABBA the Swedish Beatles a preposterous claim? (Yes)*
AICD	ABBA is a Swedish pop music group that was formed in the 1970s. The Beatles are a British rock band that was formed in 1960s. So, the answer is Yes.
CONDGEN	ABBA is a Swedish band that was very successful in the 1970s and 1980s. The Beatles are a British band that was very successful in the 1960s and 1970s. So, the answer is Yes.
LM LOSS	ABBA is a Swedish band that was very successful in the 1970s and 1980s. The band had multiple hits and has been recognized as one of the most successful musical groups in the world. So, the answer is No. *(incorrect answer)*
Case #2	*Can The Hobbit be read in its entirety in four minutes? (No)*
AICD	The Hobbit is a novel of more than 400 pages and is considered to be a long novel. The novel can't be read in its entirety in four minutes. So, the answer is No.
CONDGEN	The Hobbit is a children's storybook and can be read at a faster pace as it is shorter in length. The Hobbit can be read in its entirety in four minutes. So, the answer is yes. *(wrong reasoning and answer)*
LM LOSS	The Hobbit is a fantasy novel of more than 400 pages. It takes more than four minutes to read it completely. So, the answer is No.
Case #3	*Can Darth Vader hypothetically outdunk Bill Walton without using The Force? (No)*
AICD	Darth Vader is a fictional character and therefore has superhuman abilities. He is a Sith Lord, who is also known as the Emperor's Apprentice. Bill Walton is a real person, but he is an NBA player and has limited abilities. So, the answer is No.
CONDGEN	Darth Vader is a fictional character and has the ability to use The Force, which is a power that allows one to perform supernatural feats. Bill Walton is a real human and cannot use The Force, therefore Darth Vader cannot outdunk Bill Walton without using The Force. *(wrong reasoning)* So, the answer is No.
LM LOSS	Darth Vader is a Sith Lord, a member of the Jedi Sith, who uses The Force to increase his strength and agility. Bill Walton is a retired NBA player who is not known to have any superpowers. So, the answer is No.

Table 2: Case study on generated reasoning chains for student GPT-J-6B trained varying objectives (AICD, conditional generation and language modeling loss) on StrategyQA.

E LIMITATIONS

Better Sampling of ICL Examples Currently, in-context examples are randomly sampled from training datasets. While this straightforward implementation surpasses baselines, multiple recent works (Zhang et al., 2022; Wang et al., 2023b) suggest that this might be suboptimal, and propose strategies to mine optimal in-context examples. A promising future direction is to empower AICD with these strategies to facilitate training and mitigate potential bias in ICL examples.

In-Depth Explorations into Weight Schedules Another possible future direction is to propose an automatic strategy in assigning weights to each in-context example, which is currently designed according to experiments. We believe that the weight ω in AICD plays the following two key roles:

1) **Aligns with the nature of ICL.** Since one may obtain a better result given more ICL examples, an incremental weight schedule aligns with this inherent nature of ICL (as the confidence increases). Therefore, we believe a better design of these schedules fosters the learning of ICL capability for smaller LMs.

2) **Adjust the weight for an arbitrary sample.** Apart from learning the ICL capability, having such weight schedules also enables an adjustment to the significance of an arbitrary training sample. Therefore, it's possible to further explore schedules to mitigate bias towards a better generalization to ICL tasks and samples. Existing works on pre-training data-mixture (Thakkar et al., 2023) would serve as a viable baseline.

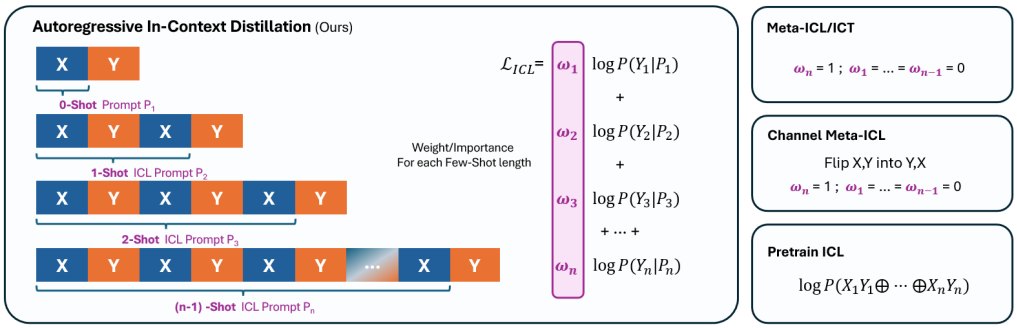

Figure 3: Comparison between learning objectives of AICD and existing in-context tuning methods.

F Illustration of AICD

For a comprehensive understanding of AICD, we provide an illustration of AICD as well as a comparison to other baselines in Figure 3. As elaborated in Chapter 2, AICD could be understood as a *meta-teacher forcing*, where we 'autoregressively' feed in-context samples into LM's context.