TECHNISCHE UNIVERSITÄT DARMSTADT · UBIQUITOUS KNOWLEDGE PROCESSING · INGENUITY LABS RESEARCH INSTITUTE at Queen's University · UNIVERSITY OF BATH

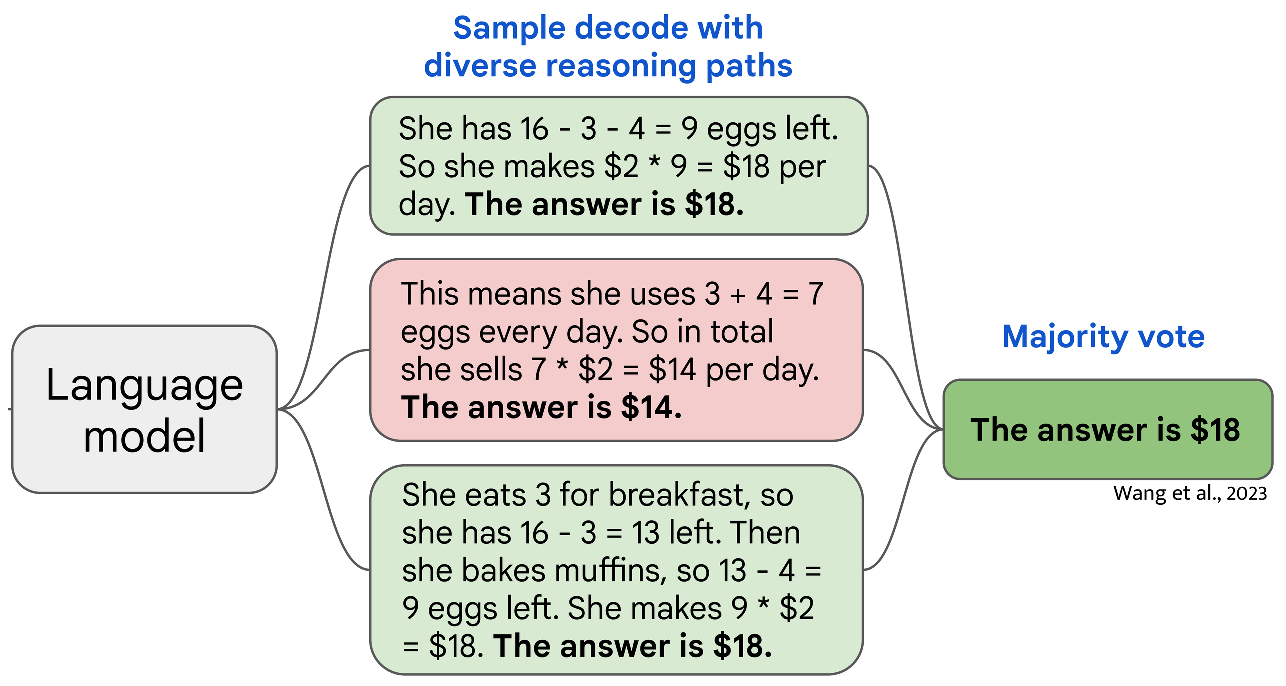
Fine-Tuning on Diverse Reasoning Chains Drives Within-Inference CoT Refinement in LLMs

Haritz Puerto, Tilek Chubakov, Xiaodan Zhu, Harish Tayyar Madabushi, Iryna Gurevych
haritz.puerto@tu-darmstadt.de

Paper Website

Generating multiple CoTs in a single inference step allows LLMs to refine their answers

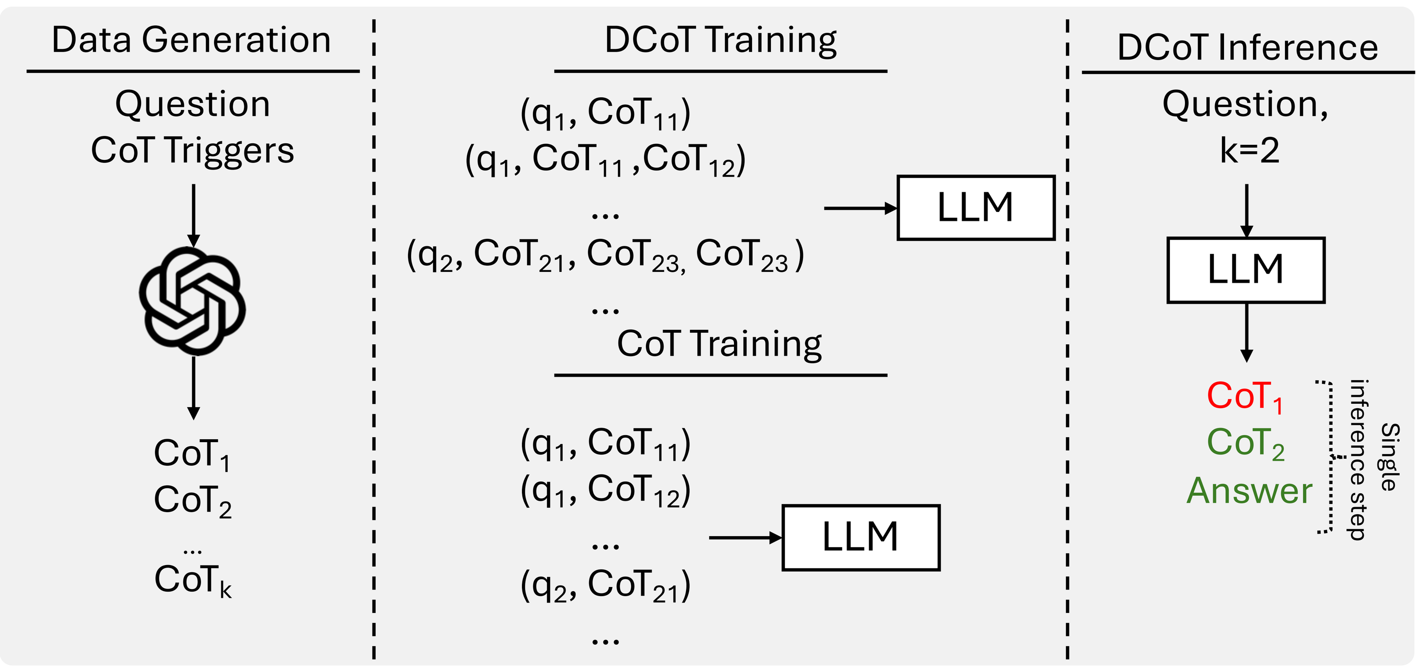
Motivation



CoTs are independent → cannot refine reasoning

Method

- How to train models to generate multiple CoTs in a *single inference step*?
- DCoT must have same #CoTs as CoT baseline



Training Data

- CoTs generated by GPT4o-mini
  For each question
    For each CoT trigger
      Generate CoT
- Training size: 32032 CoTs
- Same CoTs for DCoT and CoT

Evaluation Setup

- 2 model families from 1.3 to 70B
- Numeric: GMS8K
- Binary: StrategyQA
- Span-Extraction: CommonsenseQA, HotpotQA
- Multiple-Choice: ARC, BoardgameQA, Quartz

DCoT Outperforms CoT

LLM	Phi 1.5 (1.3B)	Phi 2 (2.7B)	LLaMA2 7B	LL. 13B	LL. 70B
CoT	47.2	60.85	58.97	64.39	66.96
DCoT	**49.39**	**62.6**	**60.8**	**66.18**	68.63
CoT + SC	46.48	61.5	**62.9**	66.82	-
DCoT + SC	**49.01**	**65.12**	61.09	**68.12**	-

DCoT is compatible with CoT extensions

One Refinement Achieve Gains

Method	Phi 1.5	Phi 2	LL. 7B	LL. 13B
CoT	47.51±1.77	63.51±.71	59.30±.54	65.41±.91
DCoT@1	47.87±1.71	63.91±2.58	61.28±.50	65.80±.44
DCoT@2	48.63±.67↑	65.33±2.80↑	62.46±.45↑	67.30±.49↑
DCoT@3	48.96±.66	65.30±1.72	62.37±.23	66.92±.59
DCoT@4	48.76±.33	64.89±2.39	62.42±.59	66.70±.55

- CoT == DCoT@1
- DCoT@2 > DCoT@1
- Converges after 2

Robust where CoT is Detrimental

Method	Phi 1.5	Phi 2	LL. 7B	LL. 13B
CoT	28.37	46.7	31.08	36.38
DCoT@1	28.31	44.56	31.23	34.59
DCoT@2	28.07	45.81	31.11	35.94
DCoT@3	28.35	45.92	31.00	36.90
DCoT@4	28.21	46.71	31.13	36.45

Results on BBH
CoT not beneficial in BBH
DCoT doesn't do worse

No Degradation in OOD

LLM	Method	CSQA	AQuA	ObjCnt	SVAMP
Phi 1.5	CoT	33.88	20.27	35.60	40.00
	DCoT@1	32.26	21.51	25.20	40.50
	DCoT@2	34.23	17.31	27.60	30.00
	DCoT@3	33.81	22.38	30.80	30.00
	DCoT@4	34.73	22.06	30.00	31.50
Phi 2	CoT	44.29	29.52	54.00	55.00
	DCoT@1	44.15	34.86	58.40	60.50
	DCoT@2	44.13	34.09	56.40	60.50
	DCoT@3	45.99	31.83	57.60	60.00
	DCoT@4	45.43	34.73	56.40	59.50

LLM	Method	CSQA	AQuA	ObjCnt	SVAMP
LLaMA2 7B	CoT	38.41	19.41	34.80	39.50
	DCoT@1	36.94	17.70	40.00	41.50
	DCoT@2	40.79	17.27	39.60	43.00
	DCoT@3	40.67	16.90	36.80	43.00
	DCoT@4	40.43	17.21	37.20	39.00
LLaMA2 13B	CoT	46.55	24.85	45.2	62.50
	DCoT@1	44.62	23.98	46.00	55.00
	DCoT@2	45.48	22.42	47.60	53.50
	DCoT@3	47.42	20.72	52.40	56.50
	DCoT@4	46.45	23.13	54.00	53.50

Takeaways

- How to present the CoTs in training has a major influence in performance
- Generating multiple CoTs in a single inference step allows models to refine their answers

Icons from iconixar, iconsmind - Flaticon
Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation
ATHENE National Research Center for Applied Cybersecurity